Received: 19 May 2025 | Accepted: 26 January 2026

DOI: 10.1111/jopr.70117

ORIGINAL ARTICLE

Performance assessment of five artificial intelligence-based algorithms for automated tooth segmentation and labeling on intraoral scans

Germán del Cacho-Salvador B.Eng., MSc[1] | **Xavier Paolo Burgos-Artizzu B.Eng., MSc, PhD**[2] | **Joaquín García-Arranz RDT**[3] | **Óscar González-Martín DDS, MS, PhD**[4,5] | **Carlos Oteo-Morilla DDS, MS, PhD**[6] | **Mercedes Gallas-Torreira DDS, PhD**[7] | **Eva Cernadas B.Eng., PhD**[8] | **Wenceslao Piedra-Cascón DDS, MS**[2,9,10,11]

[1]Doctoral Programme in Research in Information Technologies, Artificial Intelligence Area, University of Santiago de Compostela, Madrid, Spain

[2]MovumTech, Madrid, Spain

[3]Director and Dental Laboratory Technician, Ortodentis, Madrid, Spain

[4]Faculty of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine, Boston, Massachusetts, USA

[5]Faculty of Periodontology, Complutense University of Madrid, Madrid, Spain

[6]Affiliate Faculty Graduate in Esthetic Dentistry Program, Complutense University of Madrid, Madrid, Spain

[7]Planning and Management in Dental Clinics. Stomatology Area, Department of Surgery and Medical-Surgery Specialities. Digital Dentistry Unit of the School of Dentistry, Faculty of Medicine and Dentistry, University of Santiago de Compostela, Santiago, Spain

[8]Computer Science and Artificial Intelligence, Researcher at the Research Centre in Intelligent Technologies (CiTIUS), University of Santiago de Compostela, Santiago de Compostela, Spain

[9]Doctoral Progamme in Dental Science, Stomatology Area, Department of Surgery and Medical-Surgery Specialitites, University of Santiago de Compostela, Santiago, Spain

[10]Affiliate Faculty Esthetic Dentistry Program, Complutense University of Madrid, Madrid, Spain

[11]Private Practitioner, Oviedo, Spain

Correspondence
Wenceslao Piedra-Cascón, Rua Entrerríos s/n, Santiago de Compostela, 15782, Spain.
Email: wpiedra@movumtech.com

Funding information
Ortodentis dental laboratory

Abstract

Purpose: This study aimed to evaluate and compare the performance of five AI algorithms for tooth segmentation and labeling on intraoral scans, as their performance remains unclear.

Materials and Methods: A total of 100 intraoral scans in the STL file format were classified into two main groups: complete dentition (C) and partial dentition (P, fewer than 12 teeth). Each group was further divided by arch into four subgroups (n = 25 each): complete maxillary (Mx-C), complete mandibular (Md-C), partial maxillary (Mx-P), and partial mandibular (Md-P). The algorithms tested were the Tooth Group Network (Team CGIP), Dentbird Studio (Dentbird), Medit Ortho Simulation (Medit), NemoSmile 3D (Nemotec), and MovumStudio (MovumTech). Manual segmentations by an expert operator served as the ground truth. Performance was assessed using Python and five metrics, with Intersection over Union (IOU) as the primary indicator. Statistical analysis included permutation tests with the Bonferroni–Holm correction (α = 0.05).

Results: Significant differences were observed between groups and algorithms ($P < 0.05$). IOU scores ranged from 0.72 to 0.92 in complete dentition and showed greater variability in partial dentition (0–0.928). The Tooth Group Network and

MovumStudio consistently outperformed the others, with MovumStudio achieving the highest performance across all metrics and groups. Its performance matched that of a human expert when compared against a subset of the data.
Conclusions: Tooth segmentation and labeling performance vary depending on dentition completeness and algorithm choice. MovumStudio demonstrated the most robust and consistent results, comparable to expert human annotation.

KEYWORDS
artificial intelligence, digital dentistry, prosthodontics, segmentation

Artificial intelligence (AI) refers to technology that enables machines to replicate human cognitive functions, including learning and problem-solving tasks,[1,2] and includes subfields, namely machine learning (ML) and deep learning (DL).[3–5] ML allows systems to learn patterns autonomously, while DL is an advanced subset of ML that uses artificial neural networks (ANNs) to model complex and high-dimensional data.[3–6] These networks, composed of multiple interconnected layers of artificial neurons, are designed to hierarchically extract features and capture intricate relationships within the data. AI is being increasingly integrated into digital dentistry, with different AI-driven software solutions incorporated into the digital workflow, including implant recognition,[7,8] the identification of diseases and oral lesions,[9–16] dental plaque detection,[17] cephalometric analysis,[18–20] occlusion optimization,[21,22] cone beam computed tomography (CBCT) segmentation,[23,24] and the automatic design of restorations or dental devices.[25]

The primary goal of AI-based algorithms is to deliver accurate results for the specific tasks they are trained to perform, including automating repetitive processes.[26] Manual segmentation and labeling of teeth in digital files is highly time-consuming, and in prosthodontic practice, this step directly affects the efficiency and accuracy of multiple digital procedures, including virtual diagnostic waxing, restorative space evaluation, occlusion-driven design, and the fabrication of tooth- and implant-supported prostheses. For this reason, the development of fully automated algorithms capable of reliably identifying and segmenting teeth on intraoral scans is clinically relevant, as it can streamline prosthodontic workflows, reduce operator-dependent variability, and improve the predictability of digital treatment planning.[27–41]

Tooth segmentation generally involves outlining the contours of individual teeth and assigning class labels based on specific nomenclatures.[41] This process ensures that all intraoral scan surfaces representing a single crown or gingiva are assigned to their corresponding class label. Tooth segmentation has a wide range of applications in digitally assisted dentistry, including virtual extractions, implant planning,[42] autotransplantation,[43] orthodontic diagnosis,[18,20] and the design of both tooth- and implant-supported prostheses.[25] The quality of tooth segmentation is crucial for achieving reliable and long-term treatment outcomes.[44–48]

Several AI algorithms are commercially available to extract the geometric surface information of teeth from intraoral scans and classify them with the corresponding labels. Although widely used in digitally assisted dentistry workflows, automatic tooth segmentation faces several limitations, including a lack of robustness because of the high variability in tooth morphology, missing teeth, abnormal dental conditions, tooth crowding, the accuracy and postprocessing procedures of intraoral scanners, and the conversion of point clouds into three-dimensional (3D) meshes, all of which impact the geometric properties of the resulting digital files, making the outcomes of automatic tooth segmentation uncertain.[40–58]

The aim of the present study was to evaluate the performance of five AI algorithms for automatic tooth segmentation and labeling on intraoral scanner (IOS) data, including patients with complete and partially edentulous dentitions. The null hypotheses were that no significant differences would be found among the tested algorithms and that no significant differences would be observed between the complete dentition and partially edentulous dentition groups.

MATERIALS AND METHODS

The present study was conducted according to the Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines (http://www.equator-network.org/reporting-guidelines/stard/). The protocol of this study was reviewed and approved by an ethical committee (C.I. 25/050-E) in accordance with the principles outlined in the Declaration of Helsinki.

The research was conducted at the Complutense University of Madrid by using data from the postgraduate Esthetic Dentistry program. Fifty participants were selected for intraoral scanning based on the number of teeth present. The inclusion criteria were participants over 18 years of age in good general health (American Society of Anesthesiologists [ASA] Type I), without known systemic diseases and with no previous maxillofacial or dental trauma, muscular disorders, or temporomandibular joint disorders.

Maxillary, mandibular, and maximum intercuspation position (MIP) digital scans were acquired from each participant using an intraoral scanner (Primescan, Dentsply Sirona), resulting in a dataset of 50 maxillary and 50 mandibular intraoral scans in the standard tessellation language (STL) file format. Two experimental groups were created based on the number of teeth present: the C group (complete dentition)

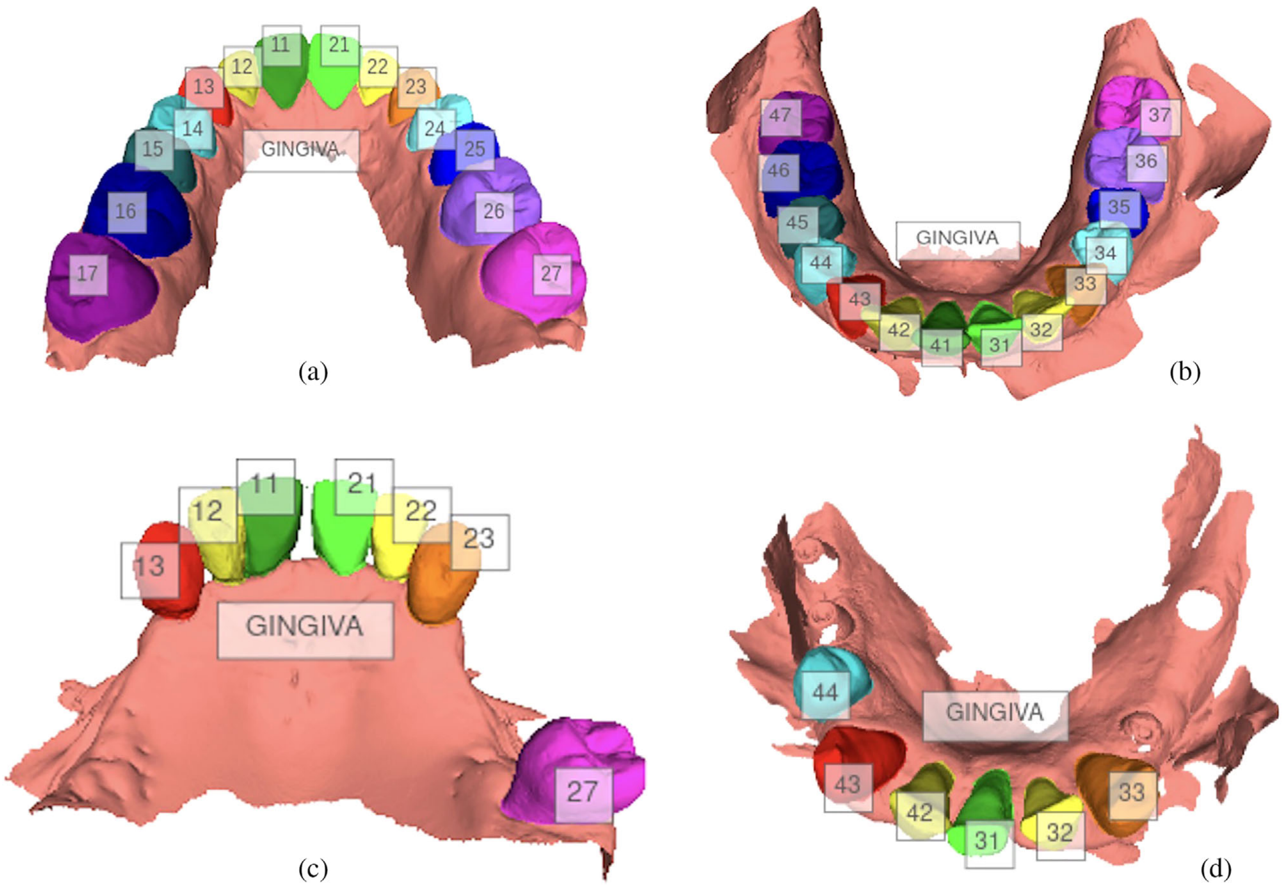


FIGURE 1 Representative ground truth tooth segmentations for each group. (a) Complete maxillary group (Mx-C). (b) Complete mandibular group (Md-C). (c) Partial maxillary group (Mx-P). (d) Partial mandibular group (Md-P).

and the P group (fewer than 12 teeth). Each primary group was further divided into two subgroups based on their anatomic location in the maxillary or mandibular regions, resulting in four classifications: Mx-C, Md-C, Mx-P, and Md-P (Figure 1a–d). Each subgroup included 25 STL files (n = 25). The scanning protocol was standardized across all test groups and performed as follows: digital scans began in the posterior area of the maxillary or mandibular arches. The scanner tip was angled at 60° toward the lingual side and moved along the dental arch to the mandibular right second molar. Subsequently, the scanner was directed occlusally from the mandibular right second molar across the dental arch back to the mandibular left second molar. Finally, the scanner was tilted 60° toward the buccal side and moved along the dental arch to complete the scanning process. The IOS (Primescan, Dentsply Sirona) and its scanning software program (Sirona Connect, v.5.1; Dentsply Sirona) had been calibrated before scanning each experimental group. All digital scans were performed by an experienced dentist (W.P.C.) with 8 years of previous experience using IOSs. The intraoral scans were performed in a room with an ambient lighting condition of 1000 lux, determined by using a luxometer (LX1330B Light Meter; Dr. Meter Digital Illuminance).

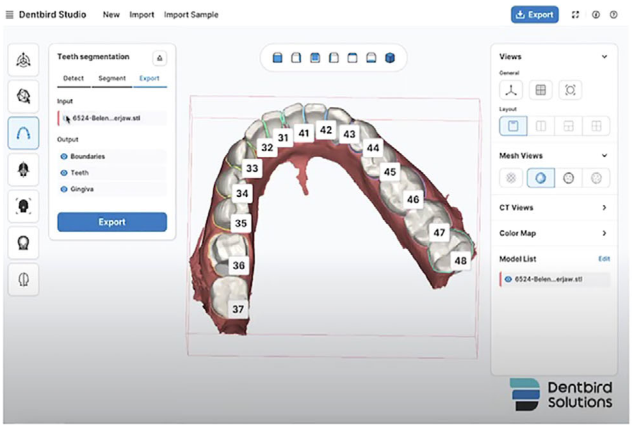
For the ground truth, an experienced clinician (W.P.C.), with 8 years of experience in using 3D computer-aided design (CAD) software programs, was responsible for manually annotating and registering the entire dataset using a CAD program specifically developed for this study in a general-purpose programming language (Python, Python Software Foundation). During the annotation process, each tooth was segmented and labeled according to its corresponding FDI notation, whereas other structures—including gingiva, scan bodies, orthodontic appliances, and artifacts—were labeled as "no-tooth" (Figure 1a–d). The inclusion of intraoral scans containing such artifacts, even if not explicitly classified as separate entities, allowed the evaluation of algorithm performance under realistic prosthodontic conditions rather than under idealized, artifact-free scenarios. The annotations were subsequently reviewed and refined, as needed, by an independent operator (C.O.M.). This curated dataset served as the reference standard for evaluating the accuracy of AI algorithms for tooth segmentation.

Five AI-based segmentation algorithms were selected for evaluation: the Tooth Group Network (Team CGIP), Dentbird Studio (Dentbird), Medit Ortho Simulation (Medit Corp), NemoSmile 3D (Nemotec), and MovumStudio (MovumTech) (Table 1, Figure 2a–c). Depending on the algorithm under evaluation, the software program may have required the user to specify whether the scan was mandibular or maxillary and/or to provide the opposing scan if only a

TABLE 1 Overview of AI-based tooth segmentation algorithms included in the study.

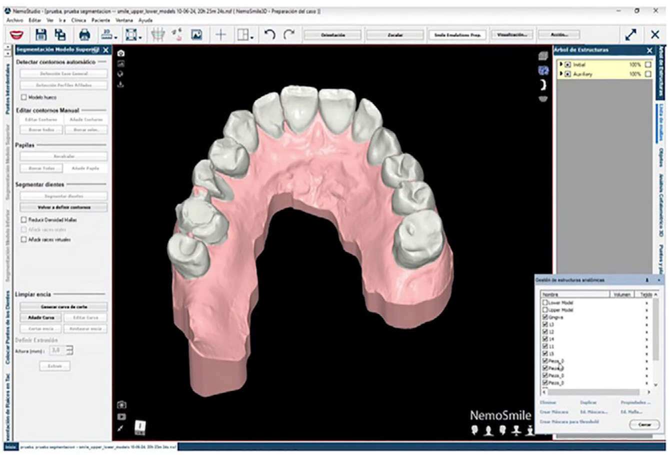
Developer	Algorithm	Version	Software program type
Team CGIP	Tooth Group Network	Final_v1	GitHub repository
Dentbird	Dentbird Studio	2.4.2	SaaS
Medit	Medit Ortho Simulation	1.3.2.48	Desktop
Nemotec	NemoSmile 3D	24.0.0.3	Desktop
Movumtech	MovumStudio	v1.0	SaaS

Abbreviations: Desktop, locally installed software program; SaaS, software-as-a-service.



(a)

(b)



(c)

FIGURE 2 User interface for tooth segmentation depending on software program. (a) Dentbird Studio. (b) Medit Ortho Simulation. (c) NemoSmile 3D.

single STL file was imported (Table 2).[25] The entire dataset was segmented using the AI-driven algorithms (Figure 3). After the dataset had been processed, segmented teeth were exported in a digital file format (STL or PLY files), depending on the software used. Subsequently, the outputs from each segmentation algorithm were standardized using the Python code to allow direct comparison with the ground truth.

The predictions on the test dataset were compared with the expert clinicians' annotations. Five performance metrics were measured: accuracy, precision, Intersection over Union (IoU), recall, and dice score. Performance metrics were defined as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

$$\text{precision} = \frac{1}{C}\sum_{i=1}^{c}\frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{IoU} = \frac{1}{C}\sum_{i=1}^{c}\frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}},$$

$$\text{recall} = \frac{1}{C}\sum_{i=1}^{c}\frac{\text{TP}}{\text{TP} + \text{FN}},$$

TABLE 2 Summary of requirements for each AI-based tooth segmentation algorithm.

Algorithm	Arch differentiation (Maxilla/Mandible)	Opposing arch required	Alignment
Tooth Group Network	Required	Not required	Required
Dentbird Studio	Not required	Not required	Not required
Medit Ortho Simulation	Required	Required	Required
NemoSmile 3D	Required	Required	Not required
MovumStudio	Not required	Not required	Not required

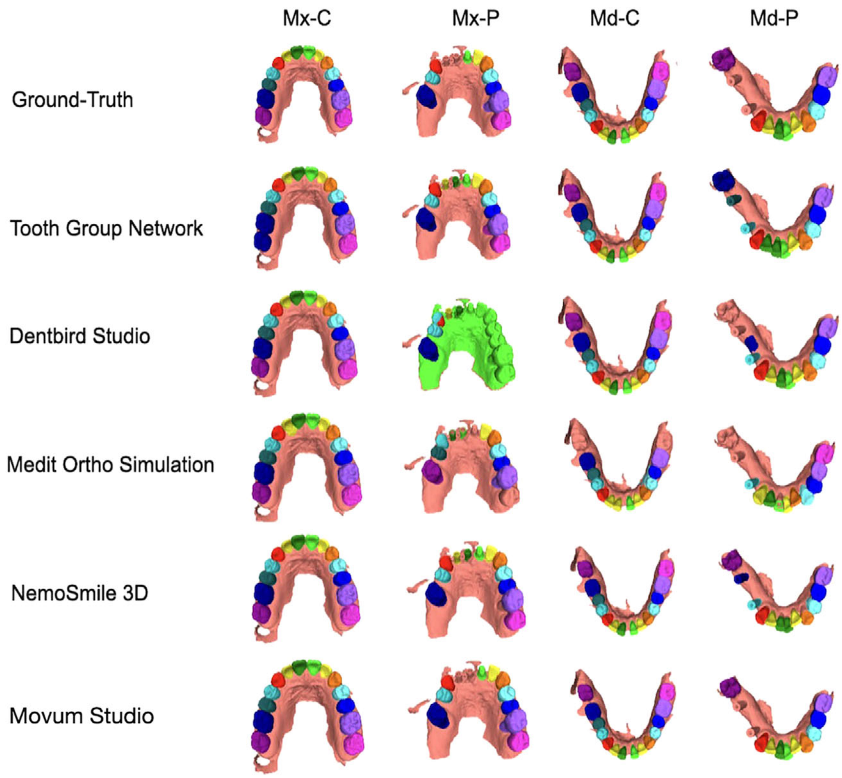


FIGURE 3 Representative results of tooth segmentations for each software program, algorithm, and group. Rows correspond to software program and algorithms. Columns correspond to groups.

and

$$\text{dice} = \frac{1}{C} \sum_{i=1}^{c} \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}},$$

where TP, TN, F, and FN are true positive, true negative, false positive, and false negative, respectively, and where C refers to the number of classes (teeth and gingiva) present in each scan (Figure 4a–e).

Once the performance metrics for each segmentation were calculated, two statistical analyses were conducted for each metric using permutation tests. The first analysis assessed whether there were significant performance differences between pairs of algorithms (10 pairwise repeated-measures comparisons). The second analysis evaluated whether performance differed significantly between the C and P groups for each software (four independent sample comparisons). In the first case, the test statistic was the median of paired differences, whereas in the second case, it was the difference in medians. All hypothesis tests were conducted with 10,000 permutations. The null hypothesis stated that both samples originated from the same population, while the alternative suggested they came from different populations. The median was preferred over the mean due to its robustness to outliers (especially failed segmentations) and its suitability under heteroscedasticity. For the first analysis, a Bonferroni–Holm correction was applied to adjust for multiple comparisons. A significance level of 0.05 was used in all tests.[55,56]

Finally, an additional analysis was performed to evaluate whether the IoU—the most representative metric of medical image segmentation—of the best performing algorithm was comparable to that of a human expert annotator. A subset of 12 intraoral scans from the initial dataset (3 Mx-C, 3 Md-C, 3 Mx-P, and 3 Md-P) was randomly selected and re-annotated by a second human expert (C.O.M.). A pairwise permutation test identical to that described in the first analysis was then applied to compare the segmentations produced by the best performing algorithm and the second human annotator.

All statistical analysis calculations were performed using the proprietary Python software code (Python v3.8, Python Software Foundation).[26–29]

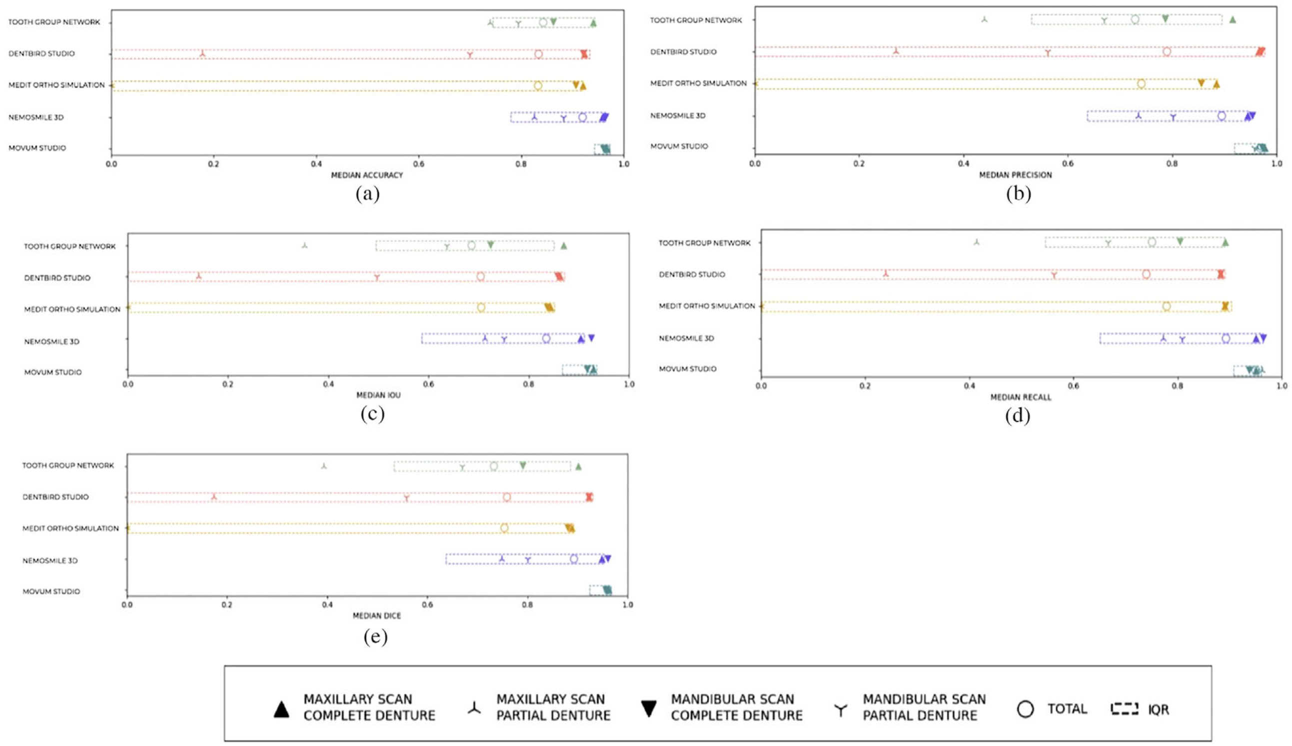


FIGURE 4 Median performance for each software program and subset. (a) Accuracy. (b) Precision. (c) IoU. (d) Recall. (e) Dice Score. *X*-axis: CSS. *Y*-axis: Software.

TABLE 3 Median performance of AI-based tooth segmentation algorithms.

Algorithm	Successful segmentations	IoU [IQR]	Accuracy [IQR]	Precision [IQR]	Recall [IQR]	Dice [IQR]
Tooth Group Network	100/100	0.686 [0.495–0.85]	0.843 [0.744–0.944]	0.729 [0.53–0.895]	0.751 [0.546–0.89]	0.733 [0.533–0.886]
Dentbird Studio	74/100	0.704 [0–0.871]	0.834 [0–0.934]	0.79 [0–0.977]	0.74 [0–0.891]	0.759 [0–0.93]
Medit Ortho Simulation	64/100	0.705 [0–0.851]	0.833 [0–0.92]	0.741 [0–0.886]	0.779 [0–0.903]	0.754 [0–0.892]
NemoSmile 3D	82/100	0.835 [0.586–0.911]	0.92 [0.78–0.965]	0.895 [0.637–0.948]	0.893 [0.651–0.964]	0.893 [0.637–0.953]
Movum Studio	100/100	0.925 [0.867–0.936]	0.966 [0.943–0.973]	0.969 [0.919–0.977]	0.95 [0.908–0.961]	0.96 [0.924–0.967]

RESULTS

The overall median performance of each algorithm is summarized in Table 3. The MovumStudio algorithm consistently outperformed all others across all measured metrics, achieving the highest median values for accuracy (0.966 [IQR: 0.943–0.973]), precision (0.969 [IQR: 0.919–0.977]), IoU (0.925 [IQR: 0.867–0.936]), recall (0.950 [IQR: 0.908–0.961]), and dice score (0.960 [IQR: 0.924–0.967]). NemoSmile 3D had the second best performance, with median values of accuracy (0.920 [IQR: 0.780–0.965]), precision (0.895 [IQR: 0.637–0.948]), IoU (0.835 [IQR: 0.586–0.911]), recall (0.893 [IQR: 0.651–0.964]), and dice score (0.893 [IQR: 0.637–0.953]), followed by the Tooth Group Network, while Dentbird Studio and Medit Ortho Simulation exhibited lower performance and a significant number of unsuccessful segmentations.

Pairwise statistical analyses revealed significant differences in performance among algorithms (Table 4). MovumStudio demonstrated statistically significant superiority across all metrics ($p < 0.01$). Notable results included precision (median of difference [MoD] = 0.1634 compared with Medit Ortho Simulation) and IoU (MoD = 0.2079 compared with the Tooth Group Network). NemoSmile 3D outperformed the Tooth Group Network and Medit Ortho Simulation in precision (MoD = 0.0563 compared with Tooth Group Network, $p < 0.01$), IoU (MoD = 0.0622, $p < 0.01$), and dice score (MoD = 0.0703, $p < 0.01$). Dentbird Studio

TABLE 4 Pairwise comparisons of AI-based tooth segmentation algorithms for accuracy, precision, Intersection over Union (IoU), recall, and dice scores. Values expressed as median of differences (MoD).

Algorithm 1	Algorithm 2	IoU (MoD)	Accuracy (MoD)	Precision (MoD)	Recall (MoD)	Dice (MoD)
Dentbird Studio	Movum Studio	−0.1193**	−0.0832**	−0.0994**	−0.1156**	−0.1005**
Dentbird Studio	Medit Ortho Simulation	0.0003	0.0000	0.0488**	0.0000	0.0007
Dentbird Studio	NemoSmile 3D	−0.0353**	−0.0293**	0.0193	−0.0629**	−0.0206**
Dentbird Studio	Tooth Group Network	−0.0203	−0.0162	0.0097	−0.0274	−0.0111
Movum Studio	Medit Ortho Simulation	0.1598**	0.1156**	0.1634**	0.1389**	0.1464**
Movum Studio	NemoSmile 3D	0.0445**	0.0207**	0.0374**	0.0104	0.0249**
Movum Studio	Tooth Group Network	0.2079**	0.1078**	0.2084**	0.1564**	0.1843**
Medit Ortho Simulation	NemoSmile 3D	−0.0623**	−0.0481**	−0.0656**	−0.0597**	−0.0620**
Medit Ortho Simulation	Tooth Group Network	−0.0631**	−0.0455**	−0.0843**	−0.0457**	−0.0642**
NemoSmile 3D	Tooth Group Network	0.0622**	0.0308**	0.0563**	0.0713**	0.0703**

Note: Statistically significant differences are indicated by an asterisk (*).
Abbreviation: MoD, median of differences.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

performed comparably with the Tooth Group Network for complete dentitions, with no significant differences in IoU or dice score ($p > 0.05$). However, it demonstrated significantly lower robustness for partially edentulous arches, with reductions in accuracy (MoD = −0.0832, $p < 0.01$) and IoU (MoD = −0.1193, $p < 0.01$), compared with MovumStudio against NemoSmile 3D. Dentbird Studio showed lower precision (MoD = −0.0293, $p < 0.01$), IoU (MoD = −0.0353, $p < 0.01$), and dice score (MoD = −0.0206, $p < 0.01$). The Tooth Group Network was statistically outperformed by both NemoSmile 3D and MovumStudio. Median differences included lower precision (MoD = 0.0563, $P < 0.01$) and IoU (MoD = 0.0622, $p < 0.01$) compared with NemoSmile 3D and significant deficits in accuracy (MoD = −0.1078, $p < 0.01$), precision (MoD = −0.2084, $p < 0.01$), and dice score (MoD = −0.1843, $p < 0.01$) compared with MovumStudio.

Statistically significant differences in performance were observed between complete (C) and partially edentulous (P) dentition groups for all algorithms except MovumStudio (Table 5). MovumStudio showed no significant difference in accuracy between the C and P groups (DiM = −0.0001, $p > 0.05$). In contrast, Medit Ortho Simulation exhibited pronounced decreases in accuracy (DiM = 0.9148, $p < 0.001$), precision (DiM = 0.8815, $p < 0.001$), and dice score (DiM = 0.8846, $p < 0.001$). Dentbird Studio showed significant reductions in accuracy (DiM = 0.5679, $p < 0.001$), precision (DiM = 0.4581, $p < 0.001$), and IoU (DiM = 0.4831, $p < 0.001$). Tooth Group Network also displayed significant reductions in accuracy (DiM = 0.1528, $p < 0.001$), precision (DiM = 0.2942, $p < 0.001$), and dice score (DiM = 0.2912, $p < 0.001$). NemoSmile 3D exhibited a minor but statistically significant decrease in accuracy (DiM = 0.0860, $p < 0.01$), with corresponding reductions in precision (DiM = 0.1944, $p < 0.001$) and dice score (DiM = 0.1945, $p < 0.001$). For maxillary segmentations in the P group, MovumStudio achieved the highest accuracy at 0.967 (IQR: 0.933–0.982), followed by NemoSmile 3D (0.735 [IQR: 0.143–0.870]), Tooth Group Network (0.440 [IQR: 0.320–0.727]), Dentbird Studio (0.270 [IQR: 0–0.811]), and Medit Ortho Simulation (0.000 [IQR: 0–0.476]). For mandibular segmentations in the P group, MovumStudio again demonstrated the highest accuracy at 0.963 (IQR: 0.909–0.976), followed by NemoSmile 3D (0.802 [IQR: 0–0.908]), Tooth Group Network (0.669 [IQR: 0.530–0.743]), Dentbird Studio (0.561 [IQR: 0–0.790]), and Medit Ortho Simulation (0.000 [IQR: 0–0.489]).

Regarding the comparison between human and automatic segmentation, the second human annotator achieved a median IoU of 0.9312 [0.895–0.951] when compared with the main annotator (W.P.C.). The best-performing algorithm, MovumStudio, achieved a median IoU of 0.9221 [0.874–0.946] for the same subset of 12 intraoral scans, showing a performance comparable to that of the second human annotator. The pairwise permutation test, identical to that described in Table 4, yielded a Median of Differences (MoD) of 0.0004 and a p-value (rounded to four decimal places) of 1, indicating no statistically significant difference between the two.

TABLE 5 Differences in performance metrics (DiM) for AI-based tooth segmentation algorithms between complete and partially edentulous patients.

Algorithm	IoU (DiM)	Accuracy (DiM)	Precision (DiM)	Recall (DiM)	Dice (DiM)
Tooth Group Network	0.2809***	0.1528***	0.2942***	0.2553***	0.2912***
Dentbird Studio	0.4831***	0.5679***	0.4581***	0.4345***	0.5091***
Medit Ortho Simulation	0.8404***	0.9148***	0.8815***	0.8916***	0.8846***
NemoSmile 3D	0.1869***	0.0860**	0.1944***	0.1694***	0.1945***
MovumStudio	0.0032	−0.0001	0.0149***	−0.0109	0.0022

Note: Statistically significant differences are indicated by an asterisk (*)
Abbreviation: DiM, difference in medians.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

DISCUSSION

This study evaluated the performance of five AI algorithms for automated tooth segmentation on intraoral scans. Although there have been some prior studies on a limited number of participants,[39,40] the authors are unaware of a previous study that compared multiple AI algorithms for tooth segmentation under actual clinical conditions. It is important to note that this study does not consider the internal architecture of the analyzed algorithms, nor does it include any information regarding the data used for their training, as most of these systems are commercial and proprietary. Therefore, the evaluation is strictly limited to the performance of the algorithms on previously unseen real clinical data.

AI-powered software programs for automated tooth segmentation and labeling play a crucial role in digitally assisted dentistry, offering considerable time savings compared with the traditional manual methods commonly used in prosthodontic, orthodontic, and surgical treatments. Previous studies on tooth segmentation have indicated that AI-assisted techniques generally achieve higher accuracy than manual approaches.[27–40] However, accuracy can vary depending on the operator's expertise, with more experienced users often achieving better results. The findings from the present research demonstrated that the performance of AI-based tooth segmentation algorithms was significantly influenced by the number of teeth present in the intraoral scans and by the specific algorithm used. Based on the findings of this study, MovumStudio consistently outperformed all other algorithms across all metrics, demonstrating the highest level of performance in every tested scenario (Table 6). Consequently, the null hypothesis that no significant differences would be found among the tested algorithms and between the complete dentition and partially edentulous groups was rejected.

Average values for tooth segmentation performance have been reported to range from 0.813 to 0.98.[28,29,34,36,37] However, many of these studies lacked consistent metrics for performance evaluation, failed to differentiate between maxillary and mandibular arches, and did not take into consideration the number of teeth present in the arch. Moreover, most studies have not used actual clinical digital files for training, making comprehensive comparisons across studies challenging. The IoU metric has been widely considered the benchmark for evaluating semantic tooth segmentation because of its ability to provide a more precise assessment of segmentation quality. Unlike accuracy, which can be inflated by class imbalances, IoU focuses on the overlap between predicted and ground-truth regions, penalizing both false positives and false negatives, making IoU particularly valuable in clinical applications where spatial precision and boundary accuracy are essential. Additionally, IoU has been reported to be more robust in multiclass scenarios, such as distinguishing individual teeth, and better reflects performance across datasets with varying class distributions, offering a more reliable evaluation of segmentation algorithms.[59] Wang et al[28] reported an IoU of 92.2 ±3.8 for tooth segmentation in complete dentate intraoral scans and an IoU of 89.3 ±8.0 for partially edentulous patients using a convolutional neural network. Similarly, Vinayahalingam et al[29] achieved an average IoU score of 0.915 for tooth segmentation on intraoral scans. In the present investigation, the best average IoU was obtained by the MovumStudio algorithm, 0.929 [0.916–0.936]. For maxillary scans with a complete dentition, MovumStudio achieved an IoU of 0.929 [0.916–0.936], outperforming other algorithms, including NemoSmile 3D (0.904 [0.848–0.91]) and Tooth Group Network (0.87 [0.763–0.922]). In partially edentulous maxillary arches, MovumStudio's IoU remained stable at 0.928 [0.793–0.954], while other algorithms, including Dentbird Studio and Medit Ortho Simulation, exhibited significant declines, with IoUs of 0.142 [0–0.721] and 0 [0–0.461], respectively. This discrepancy highlights the limitations of less robust algorithms when processing complex or incomplete dentitions. For mandibular scans, similar trends were observed. MovumStudio achieved an IoU of 0.917 [0.874–0.927] in complete dentitions and maintained strong performance in partially edentulous arches (0.916 [0.802–0.937]). By contrast, NemoSmile 3D, the second-best performer, showed a moderate decline in IoU for partially edentulous mandibular arches (0.751 [0–0.815]). The poorer performance of Medit

TABLE 6 Median performance of AI-based tooth segmentation algorithms across maxillary and mandibular intraoral scans for complete and partially edentulous patients.

Algorithm	Successful segmentations	IoU [IQR]	Accuracy [IQR]	Precision [IQR]	Recall [IQR]	Dice [IQR]
Maxillary scan, completely dentate (*N* = 25)						
Tooth Group Network	25/25	0.87 [0.763–0.922]	0.941 [0.877–0.964]	0.916 [0.808–0.978]	0.892 [0.81–0.94]	0.902 [0.804–0.959]
Dentbird Studio	21/25	0.863 [0.776–0.901]	0.924 [0.867–0.949]	0.966 [0.851–0.987]	0.883 [0.791–0.916]	0.923 [0.819–0.948]
Medit Ortho Simulation	23/25	0.844 [0.78–0.895]	0.921 [0.882–0.955]	0.885 [0.827–0.926]	0.891 [0.847–0.957]	0.889 [0.837–0.943]
NemoSmile 3D	22/25	0.904 [0.848–0.91]	0.959 [0.936–0.968]	0.945 [0.886–0.949]	0.951 [0.923–0.964]	0.949 [0.891–0.952]
Movum Studio	25/25	0.929 [0.916–0.936]	0.968 [0.965–0.972]	0.978 [0.973–0.981]	0.951 [0.94–0.957]	0.963 [0.956–0.967]
Maxillary scan, partial dentition (*N* = 25)						
Tooth Group Network	25/25	0.352 [0.302–0.682]	0.739 [0.671–0.846]	0.44 [0.32–0.727]	0.414 [0.376–0.746]	0.393 [0.329–0.731]
Dentbird Studio	15/25	0.142 [0–0.721]	0.178 [0–0.835]	0.27 [0–0.811]	0.239 [0–0.751]	0.174 [0–0.777]
Medit Ortho Simulation	9/25	0 [0–0.461]	0 [0–0.773]	0 [0–0.476]	0 [0–0.484]	0 [0–0.48]
NemoSmile 3D	19/25	0.712 [0.109–0.831]	0.825 [0.745–0.916]	0.735 [0.143–0.87]	0.772 [0.109–0.873]	0.748 [0.124–0.872]
Movum Studio	25/25	0.928 [0.793–0.954]	0.967 [0.933–0.982]	0.964 [0.808–0.973]	0.963 [0.81–0.978]	0.962 [0.813–0.976]
Mandibular scan, completely dentate (*N* = 25)						
Tooth Group Network	25/25	0.724 [0.623–0.902]	0.863 [0.787–0.956]	0.787 [0.67–0.967]	0.805 [0.72–0.929]	0.791 [0.676–0.948]
Dentbird Studio	21/25	0.858 [0.69–0.894]	0.922 [0.815–0.941]	0.972 [0.758–0.984]	0.883 [0.714–0.91]	0.923 [0.734–0.944]
Medit Ortho Simulation	23/25	0.837 [0.715–0.908]	0.907 [0.849–0.947]	0.856 [0.746–0.943]	0.892 [0.781–0.954]	0.881 [0.764–0.951]
NemoSmile 3D	23/25	0.925 [0.86–0.931]	0.965 [0.926–0.972]	0.954 [0.903–0.959]	0.965 [0.921–0.971]	0.961 [0.896–0.964]
Movum Studio	25/25	0.917 [0.874–0.927]	0.961 [0.942–0.966]	0.973 [0.964–0.978]	0.938 [0.913–0.95]	0.957 [0.932–0.962]
Mandibular scan, partial dentition (*N* = 25)						
Tooth Group Network	25/25	0.636 [0.493–0.7]	0.794 [0.758–0.848]	0.669 [0.53–0.743]	0.666 [0.534–0.749]	0.669 [0.533–0.732]
Dentbird Studio	17/25	0.497 [0–0.746]	0.7 [0–0.904]	0.561 [0–0.79]	0.562 [0.0–0.756]	0.558 [0.0–0.772]
Medit Ortho Simulation	9/25	0 [0–0.454]	0 [0–0.65]	0 [0–0.489]	0 [0–0.514]	0 [0–0.501]
NemoSmile 3D	18/25	0.751 [0–0.815]	0.883 [0–0.914]	0.802 [0–0.908]	0.809 [0–0.889]	0.8 [0–0.895]
Movum Studio	25/25	0.916 [0.802–0.937]	0.963 [0.909–0.976]	0.957 [0.828–0.968]	0.952 [0.838–0.968]	0.956 [0.836–0.967]

Ortho Simulation and Dentbird Studio in both maxillary and mandibular partial dentition arches further emphasizes the challenges faced by these algorithms in processing complex dental configurations.

The findings suggest that the Nemotec and Tooth Group Network algorithms performed correctly in patients with a complete dentition but experienced a noticeable decline in performance when applied to partially edentulous patients. Additionally, Medit Ortho Simulation and Dentbird Studio encountered challenges in partially edentulous patients, particularly in the maxillary arch. These results emphasize the varying capabilities of AI algorithms and the critical importance of selecting the most suitable tools based on specific clinical needs and the complexity of individual patients. The differences in IoU across the tested AI algorithms can be attributed to variations in algorithm architecture, training datasets, and validation methodologies.

Finally, as the difference in median IoU between the manual segmentation performed by C.O.M. and that produced by MovumStudio was only 0.0004 and not statistically significant, intraoral AI segmentation technology appears to have reached a level of maturity suitable for routine clinical use.

Limitations of the present study include the sample size ($n = 25$ per subgroup; $N = 100$ in total), which may limit the statistical power and generalizability of the results, and the use of only one intraoral digital scanner. Larger datasets and the inclusion of scans acquired with different devices would be required to confirm the consistency of these findings across broader populations and scanning systems.

CONCLUSION

Based on the findings of this clinical study, significant differences were observed among the tested algorithms, highlighting variations in their ability to accurately segment and label teeth in intraoral scans. Partially edentulous arches exhibited significantly lower performance compared with complete dentate arches. Among the evaluated systems, the MovumStudio algorithm consistently achieved the highest scores across all performance metrics for tooth segmentation, demonstrating superior results in both maxillary and mandibular intraoral digital scans. Its performance was statistically comparable to that of a human expert annotator.

ACKNOWLEDGMENTS

The authors thank Ortodentis dental laboratory for their support with data collection.

CONFLICT OF INTEREST STATEMENT

The authors declare the following potential conflicts of interest related to the MovumTech algorithm evaluated in this study: G.C.S is a current employee of MovumTech. X.B.A holds equity or stock ownership in MovumTech. O.G.M serves on the Board membership and holds equity or stock ownership in MovumTech. J.G.A holds equity or stock ownership in MovumTech. W.P.C serves on the board membership and equity or stock ownership in MovumTech. C.O.M holds equity or stock ownership in MovumTech. All other authors declare no competing interests.

ORCID

Xavier Paolo Burgos-Artizzu B.Eng., MSc, PhD https://orcid.org/0000-0001-9681-5404
Joaquín García-Arranz RDT https://orcid.org/0009-0006-1838-3152
Óscar González-Martín DDS, MS, PhD https://orcid.org/0000-0003-3957-4872
Carlos Oteo-Morilla DDS, MS, PhD https://orcid.org/0000-0001-5293-0195
Wenceslao Piedra-Cascón DDS, MS https://orcid.org/0000-0002-2765-2207

REFERENCES

1. International Organization for Standardization. ISO/IEC 2382:2015. Information Technology—Vocabulary. Available at: https://www.iso.org/standard/63598.html. Accessed July 1, 2020.
2. International Organization for Standardization. ISO/IEC TR 24028. Information Technology—Artificial Intelligence—Overview of Trustworthiness in Artificial Intelligence. 2020. Available at: https://www.iso.org/standard/77608. html?browse=tc. Accessed July 1, 2020.
3. Das K, Behera RN. A survey on machine learning: concept, algorithms and applications. Int J Innov Res Comput Commun Eng. 2017;5:1301–9.
4. Alpaydin E. Introduction to machine learning. 4th ed. Cambridge, MA: Massachusetts Institute of Technology; 2020. 23–491.
5. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
6. Park WJ, Park JB. History and application of artificial neural networks in dentistry. Eur J Dent 2018;12:594–601.
7. Revilla-León M, Gómez-Polo M, Vyas S, Barmak BA, Galluci GO, Att W, et al. Artificial intelligence applications in implant dentistry: a systematic review. J Prosthet Dent 2023;129(2):293–300.
8. Bonfanti-Gris M, Ruales E, Salido MP, Martinez-Rus F, Özcan M, Pradies G. Artificial intelligence for dental implant classification and peri-implant pathology identification in 2D radiographs: a systematic review. J Dent 2025;153:105533.
9. Yilmaz E, Kayikcioglu T, Kayipmaz S. Computer-aided diagnosis of periapical cyst and keratocystic odontogenic tumor on cone beam computed tomography. Comput Methods Programs Biomed 2017;146: 91–100.
10. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Ozyurek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J 2020;53:680–89.
11. Joo J, Jeong S, Jin H, Lee U, Yoon JY, Kim SC. Periodontal disease detection using convolutional neural networks. In: 1st International Conference on Artificial Intelligence in Information and Communication (ICAIIC); 2019. 360–62.
12. Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, et al. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep 2019;9:8495.
13. Li W, Jiang X, Sun W, Wang SH, Liu C, Zhang X, et al. Gingivitis identificationvia multichannel gray-level co-occurrence matrix and particle swarm optimization neural network. Int J Imaging Syst Technol 2020;30:401–11.
14. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci 2018;48:114–23.
15. Yauney G, Rana A, Wong LC, Javia P, Muftu A, Shah P. Automated process incorporating machine learning segmentation and correlation

of oral diseases with systemic health. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:3387–93.
16. Musleh D, Almossaeed H, Balhareth F, Alqahtani G, Alobaidan N, Altalag J, et al. Advancing dental diagnostics: a review of artificial intelligence applications and challenges in dentistry. Big Data Cogn Comput 2024;8:66 https://doi.org/10.3390/bdcc8060066
17. You W, Hao A, Li S, Wang Y, Xia B. Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. BMC Oral Health 2020;20:141.
18. Polizzi A, Boato M, Serra S, D'Antò V, Leonardi R. Applications of artificial intelligence in orthodontics: a bibliometric and visual analysis. Clin Oral Investig 2025;29(1):65.
19. Sadek M, Alaskari O, Hamdan A. Accuracy of web-based automated versus digital manual cephalometric landmark identification. Clin Oral Investig 2024 Nov;28(11):621.
20. Zecca PA, Caccia M, Levrini L, Carganico A, Reguzzoni M, Donadio D, et al. AI-based open-source software for cephalometric analysis from limited FOV radiographs. J Dent 2024 Dec;151:105412.
21. Revilla-León M, Gómez-Polo M, Barmak AB, Kois JC, Alonso Pérez-Barquero J. Influence of scan extension on the accuracy of maximum intercuspal position recorded by using intraoral scanners or an artificial intelligence-based program. J Prosthet Dent 2025;S0022-3913(25)00044-7.
22. Revilla-León M, Ntovas P, Barmak AB, Kois JC, Alonso Pérez-Barquero J. Influence of edentulous areas on the accuracy of the maximum intercuspal position recorded by using different intraoral scanners or an artificial intelligence-based program. J Prosthet Dent 2025;S0022-3913(24)00827-8.
23. Fan W, Zhang J, Wang N, Li J, Hu L. The application of deep learning on CBCT in dentistry. Diagnostics (Basel) 2023 Jun;13(12):2056.
24. Zheng Q, Gao Y, Zhou M, Li H, Lin J, Zhang W, et al. Semi or fully automatic tooth segmentation in CBCT images: a review. PeerJ Comput Sci 2024 Apr;10:e1994.
25. Bernauer SA, Wieland P, Zitzmann NU, Joda T. Feasibility testing of the automatic design of three-unit implant fixed dental prostheses with different dental CAD software: a pre-clinical pilot trial. J Clin Med 2025 Mar;14(1):233.
26. Tandon D, Rajawat J. Present and future of artificial intelligence in dentistry. J Oral Biol Craniofac Res 2020;10(4):391–96.
27. Jana A, Subhash HM, Metaxas D. Automatic tooth segmentation from 3D dental model using deep learning: a quantitative analysis of what can be learnt from a single 3D dental model. Proc SPIE 2023;12567:1256707.
28. Wang X, Alqahtani KA, Van den Bogaert T, Shujaat S, Jacobs R, Shaheen E. Convolutional neural network for automated tooth segmentation on intraoral scans. BMC Oral Health 2024;24:804 https://doi.org/10.1186/s12903-024-04582-2
29. Vinayahalingam S, Kempers S, Schoep J, Hsu TH, Moin DA, van Ginneken B, et al. Intra-oral scan segmentation using deep learning. BMC Oral Health 2023;23:643 https://doi.org/10.1186/s12903-023-03362-8
30. Zhuang S, Wei G, Cui Z, Zhou Y. Robust hybrid learning for automatic teeth segmentation and labeling on 3D dental models. IEEE Trans Multimedia 2023. https://doi.org/10.1109/TMM.2023.3289760
31. Xu X, Liu C, Zheng Y. 3D tooth segmentation and labeling using deep convolutional neural networks. IEEE Trans Vis Comput Graph 2019 Jul;25(7):2336–48 https://doi.org/10.1109/TVCG.2018.2839685
32. Xiong H, Li K, Tan K, Feng Y, Tianyi Zhou J, Hao J, et al. TSegFormer: 3D tooth segmentation in intraoral scans with geometry guided transformer. In: Greenspan H, Madabhushi A, Mousavi P, Salcudean S, Duncan J, Syeda-Mahmood T, et al., editors. Medical image computing and computer assisted intervention—MICCAI 2023. Lecture Notes in Computer Science. Vol. 14225. Cham: Springer; 2023. 481–90 https://doi.org/10.1007/978-3-031-43987-2_41
33. Kim T, Cho Y, Kim D, Chang M, Kim YJ. Tooth segmentation of 3D scan data using generative adversarial networks. Appl Sci 2020;10(2):490 https://doi.org/10.3390/app10020490
34. Hao J, Liao W, Zhang YL, Peng J, Zhao Z, Chen Z, et al. Toward clinically applicable 3-dimensional tooth segmentation via deep learning. J Dent Res 2022;101(3):304–11 https://doi.org/10.1177/00220345211040459
35. Almalki A, Latecki LJ. Self-supervised learning with masked autoencoders for teeth segmentation from intra-oral 3D scans. In: Proc IEEE/CVF Winter Conf Appl Comput Vis 2024. 7820–30.
36. Cui ZM, Li CJ, Chen NL, Wei GD, Chen RN, Zhou YF, et al. TSegNet: an efficient and accurate tooth segmentation network on 3D dental model. Med Image Anal 2021;69:101949.
37. Zanjani FG, Pourtaherian A, Zinger S, Moin DA, Claessen F, Cherici T, et al. Mask-MCNet: tooth instance segmentation in 3D point clouds of intraoral scans. Neurocomputing 2021;453:286–98.
38. Zhao Y, Zhang LM, Yang CS, Tan YY, Liu Y, Li PC, et al. 3D Dental model segmentation with graph attentional convolution network. Pattern Recogn Lett 2021;152:79–85.
39. Lian C, Wang L, Wu TH, Wang F, Yap PT, Ko CC, et al. Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3d intraoral scanners. IEEE Trans Med Imaging 2020;39(7):2440–50.
40. Yacout YM, Eid FY, Tageldin MA, Kassem HE. Evaluation of the accuracy of automated tooth segmentation of intraoral scans using artificial intelligence-based software packages. Am J Orthod Dentofacial Orthop 2024 Sep;166(3):282–91e1. https://doi.org/10.1016/j.ajodo.2024.05.015
41. Raju R, Tr PA. Accuracy of tooth segmentation in the digital Kesling setup of two different software programs: a retrospective study. Cureus 2024 Sep;16(9):e70306.
42. Ntovas P, Marchand L, Schnappauf A, Finkelman M, Revilla-Leon M, Att W. Comparison between conventional and artificial intelligence-assisted setup for digital implant planning: accuracy, time-efficiency, and user experience. Clin Oral Implants Res 2024;21:290–97.
43. Moreno-Pérez J, Rubio Flores D, Fernández Villares C, Nagendrababu V, Abella Sans F. Guided tooth autotransplantation of a palatally impacted maxillary canine into the site of a failed maxillary canine implant. Aust Endod J 2024;50(2):369–76.
44. Tsirogiannis P, Reissmann DR, Heydecke G. Evaluation of the marginal fit of single-unit, complete-coverage ceramic restorations fabricated after digital and conventional impressions: a systematic review and meta- analysis. J Prosthet Dent 2016;116:328–35e2.
45. Baan F, de Waard O, Bruggink R, Xi T, Ongkosuwito EM, Maal TJJ. Virtual setup in orthodontics: planning and evaluation. Clin Oral Investig 2020;24(7):2385–93.
46. Tabesh M, Nejatidanesh F, Savabi G, Davoudi A, Savabi O, Mirmohammadi H. Marginal adaptation of zirconia complete-coverage fixed dental restorations made from digital scans or conventional impressions: a systematic review and meta-analysis. J Prosthet Dent 2021;125: 603–10.
47. Hasanzade M, Aminikhah M, Afrashtehfar KI, Alikhasi M. Marginal and internal adaptation of single crowns and fixed dental prostheses by using digital and conventional workflows: a systematic review and meta-analysis. J Prosthet Dent 2021;126:360–68.
48. Tabesh M, Nejatidanesh F, Savabi G, Davoudi A, Savabi O. Marginal accuracy of lithium disilicate full-coverage single crowns made by direct and indirect digital or conventional workflows: a systematic review and meta- analysis. J Prosthodont 2022;31:744–53.
49. Michelinakis G, Apostolakis D, Tsagarakis A, Kourakis G, Pavlakis E. A comparison of accuracy of 3 intraoral scanners: a single-blinded in vitro study. J Prosthet Dent 2020;124(5):581–88 https://doi.org/10.1016/j.prosdent.2019.10.023
50. Revilla-León M, Jiang P, Sadeghpour M, Piedra Cascón W, Zandinejad A, Özcan M, et al. Intraoral digital scans. Part 1: influence of ambient scanning light conditions on the accuracy (trueness and precision) of different intraoral scanners. J Prosthet Dent 2019 Dec; S0022-3913(18)30992-2. https://doi.org/10.1016/j.prosdent.2019.06.003

51. Piedra-Cascón W, Adhikari RR, Özcan M, Krishnamurthy VR, Revilla-León M, Gallas-Torreira M. Accuracy assessment (trueness and precision) of a confocal based intraoral scanner under twelve different ambient lighting conditions. J Dent 2023;134:104530.
52. Gehrke P, Rashidpour M, Sader R, Weigl PA. A systematic review of factors impacting intraoral scanning accuracy in implant dentistry with emphasis on scan bodies. Int J Implant Dent 2024;10:20.
53. Thomas AA, Jain RK. Influence of operator experience on scanning time and accuracy with two different intraoral scanners: a prospective clinical trial. Turk J Orthod 2023 Mar;36(1):10–14.
54. Sindhu S, Maiti S, Nallaswamy D. Factors affecting the accuracy of intraoral scanners: a systematic review. Ann Dent Spec 2023;11:40–52.
55. Kroeger CM, Ejima K, Hannon BA, Halliday TM, McComb B, Teran-Garcia M, et al. Persistent confusion in nutrition and obesity research about the validity of classic nonparametric tests in the presence of heteroscedasticity: evidence of the problem and valid alternatives. Am J Clin Nutr 2021;113(3):517–24 https://doi.org/10.1093/ajcn/nqaa357
56. St. Laurent R, Turk P. The effects of misconceptions on the properties of Friedman's test. Commun Stat Simul Comput 2013;42(7):1596–615 https://doi.org/10.1080/03610918.2012.671874
57. Giacalone M, Agata Z, Cozzucoli PC, Basso A. Bonferroni-Holm and permutation tests to compare health data: methodological and applicative issues. BMC Med Res Methodol 2018;18:81 https://doi.org/10.1186/s12874-018-0540-8
58. Sipper M. Neural networks with à la carte selection of activation functions. SN Comput Sci 2021;2:470 https://doi.org/10.1007/s42979-021-00885-1
59. Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes 2022;15:210.

How to cite this article: Cacho-Salvador G, Burgos-Artizzu XP, García-Arranz J, González-Martín Ó, Oteo-Morilla C, Gallas-Torreira M, et al. Performance assessment of five artificial intelligence-based algorithms for automated tooth segmentation and labeling on intraoral scans. J Prosthodont. 2026;1–12. https://doi.org/10.1111/jopr.70117