

Influence of dental fillings and tooth type on the performance of a novel artificial intelligence-driven tool for automatic tooth segmentation on CBCT images – A validation study

Rocharles Cavalcante Fontenele^{a,b,*}, Maurício do Nascimento Gerhardt^{a,c}, Jáder Camilo Pinto^{a,d}, Adriaan Van Gerven^e, Holger Willems^e, Reinhilde Jacobs^{a,f,g}, Deborah Queiroz Freitas^b

^a OMFS IMPATH Research Group, Department of Imaging and Pathology, Faculty of Medicine, KU Leuven, Leuven, Belgium

^b Department of Oral Diagnosis, University of Campinas, Piracicaba, Brazil

^c Department of Prosthodontics, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

^d Department of Restorative Dentistry, São Paulo State University, School of Dentistry, Araraquara, Brazil

^e ReLU, Innovatie-en incubatiecentrum KU Leuven, Leuven, Belgium

^f Department of Oral and Maxillofacial Surgery, University Hospitals Leuven, Leuven, Belgium

^g Department of Dental Medicine, Karolinska Institute, Stockholm, Sweden

ARTICLE INFO

Keywords:

Artificial intelligence
Fillings
Tooth
Cone-beam computed tomography
Convolutional neural network

ABSTRACT

Objectives: To assess the influence of dental fillings on the performance of an artificial intelligence (AI)-driven tool for tooth segmentation on cone-beam computed tomography (CBCT) according to the type of tooth.

Methods: A total of 175 CBCT scans (500 teeth) were recruited for performing training (140 CBCT scans - 400 teeth) and validation (35 CBCT scans - 100 teeth) of the AI convolutional neural networks. The test dataset involved 74 CBCT scans (226 teeth), which was further divided into control and experimental groups depending on the presence of dental filling: without filling (control group: 24 CBCT scans - 113 teeth) and with coronal and/or root filling (experimental group: 50 CBCT scans - 113 teeth). The segmentation performance for both groups was assessed. Additionally, 10% of each tooth type (anterior, premolar, and molar) was randomly selected for time analysis according to manual, AI-based and refined-AI segmentation methods.

Results: The presence of fillings significantly influenced the segmentation performance ($p < 0.05$). However, the accuracy metrics showed an excellent range of values for both control (95% Hausdorff Distance (95% HD): 0.01–0.08 mm; Intersection over union (IoU): 0.97–0.99; Dice similarity coefficient (DSC): 0.98–0.99; Precision: 1.00; Recall: 0.97–0.99; Accuracy: 1.00) and experimental groups (95% HD: 0.17–0.25 mm; IoU: 0.91–0.95; DSC: 0.95–0.97; Precision: 1.00; Recall: 0.91–0.95; Accuracy: 0.99–1.00). The time analysis showed that the AI-based segmentation was significantly faster with a mean time of 29.8 s ($p < 0.001$).

Conclusions: The proposed AI-driven tool allowed an accurate and time-efficient approach for the segmentation of teeth on CBCT images irrespective of the presence of high-density dental filling material and the type of tooth. **Clinical significance:** Tooth segmentation is a challenging and time-consuming task, mainly in the presence of artifacts generated by dental filling material. The proposed AI-driven tool could offer a clinically acceptable approach for tooth segmentation, to be applied in the digital dental workflows considering its time efficiency and high accuracy regardless of the presence of dental fillings.

1. Introduction

The conventional dentistry workflows are being constantly replaced

by digital workflows due to the incorporation of several computer-controlled tools, such as cone-beam computed tomography (CBCT), computer-aided-design/computer-assisted-manufacturing (CAD/CAM),

* Corresponding author at: OMFS IMPATH Research Group, Department of Imaging and Pathology, Faculty of Medicine, KU Leuven, Leuven, Belgium. Department of Oral Diagnosis, University of Campinas, Piracicaba, Brazil, Kapucijnenvoer 33, BE-3000 Leuven, Belgium.

E-mail address: rocharlesf@gmail.com (R.C. Fontenele).

<https://doi.org/10.1016/j.jdent.2022.104069>

Received 14 December 2021; Received in revised form 26 January 2022; Accepted 16 February 2022

Available online 18 February 2022

0300-5712/© 2022 Elsevier Ltd. All rights reserved.

three-dimensional (3D) printing, dynamic navigation and other advanced prototyping methodologies. All disciplines of dentistry have benefitted from these technological advancements and improved the efficiency of the oral healthcare system by offering a precise diagnosis, patient-specific treatment planning and follow-up evaluation [1].

One of the most critical steps in the digital dental workflows is the segmentation of teeth on CBCT images, which is a prerequisite for ensuring a correct diagnosis of tooth-related diseases and accurate virtual treatment planning [2–5]. This step is most commonly performed by semi- or fully-automatic processes that require selecting a thresholding of gray values of the voxels that better encompass the region of interest [6,7]. However, these threshold-based approaches are prone to certain limitations which might significantly impact the accuracy of tooth segmentation, such as inability to distinguish tooth root from alveolar bone due to the presence of similar intensity profile, sensitivity to image noise, human performance variability depending on the operator's expertise and requirement of manual corrections which can be laborious and time-consuming. Furthermore, gray values thresholding is unreliable in the presence of artifacts generated by high-density dental materials, such as coronal fillings, metal posts and root fillings, due to higher variability of gray values at the region closest to the artifact generating material [4,7,8].

Recent advancements in the field of artificial intelligence (AI) have allowed the inclusion of innovative tools in the digital dental practice to overcome the inherent limitations associated with the classical tooth segmentation techniques [1,2-5]. The availability of modern data-intensive computational resources and data availability have heavily contributed to developing robust deep learning models through the application of artificial neural networks inspired by the biological neural network of a human brain [9,10]. Convolutional neural network (CNN), a robust type of deep learning algorithm, has shown significant potential for the automated segmentation of pharyngeal airway [11], mandibular bone [9], and teeth [2,4,5,12,13].

Previous studies have reported the use of CNNs for performing automatic individual tooth segmentation from CBCT images with high accuracy [4,5,12,13]. However, a lack of evidence exists concerning the performance of these CNNs for the automated segmentation of teeth restored with coronal and/or root fillings. Thus, the influence of artifacts generated by high-density restorative materials on the performance of CNNs should be considered as the majority of patients do not offer a pristine dentition. Furthermore, the time required for performing segmentation has also received less attention, which is a key deciding factor for the clinical applicability of these networks.

Therefore, the aim of the present study was to assess the influence of dental fillings and the type of tooth on the performance of a CNN-based tool for automatic tooth segmentation on CBCT images.

2. Materials and methods

This study was conducted in compliance with the World Medical Association Declaration of Helsinki on medical research. Ethical approval was obtained from the Local Institutional Ethics Board (reference number: B322201525552). Informed consent was not required as patient-specific information was anonymized.

2.1. Dataset

A total of 175 CBCT scans (500 teeth) were recruited for training (140 CBCT scans - 400 teeth) and validation (35 CBCT scans - 100 teeth) of the AI networks. The teeth were randomly selected and belonged to all teeth groups (incisor, canine, premolar, and molars). The scans were obtained from the Hospital database, which were acquired for dental implant planning, third molars evaluation, endodontic treatment follow-up, and CBCT-guided tooth autotransplantation. The CBCT imaging devices consisted of NewTom VGi Evo (Cefla, Imola, Italy) and 3D Accuitomo 170 (J Morita, Kyoto, Japan) with different scanning

parameters: 110 kilovoltage peak (kVp), 3 – 20 milliampere (mA), field of view (FOV): 8 × 8, 10 × 10, 12 × 8, 16 × 16 and 24 × 19 cm, voxel size: 0.125 – 0.300 mm³ for the NewTom VGi EVO and 90 kVp, 5 mA, FOV: 8 × 8, 10 × 10, 14 × 10, and 17 × 12 cm, voxel size: 0.125 – 0.250 mm³ for the 3D Accuitomo. The inclusion criteria consisted of patients aged 18 years or older, presence of complete and unrestored permanent dentition or small edentulous zones (up to two consecutive missing teeth), and teeth with coronal and/or root filling. Exclusion criteria were large edentulous areas (more than two successive missing teeth), large amount of metal artifacts due to the presence of dental implants and/or brackets, and motion artifacts.

The test dataset involved 74 CBCT scans (226 teeth), which was further divided into a control and experimental group depending on the presence of dental filling: without filling (control group: 24 CBCT scans – 113 teeth) and with coronal and/or root filling (experimental group: 50 CBCT scans – 113 teeth). Both groups were composed by maxillary and mandibular teeth. The influence of dental fillings on the automated segmentation was assessed by comparing the segmentation accuracy of the control and experimental group. Each group was formed based on the similar number and type of tooth group, i.e. 33 anterior teeth, 40 premolars, and 40 molars ($n = 113$ per group). The control group included CBCT scans with a complete and unrestored natural dentition. On the other hand, the experimental group (Table 1) consisted of CBCT scans with at least one tooth restored with coronal filling (amalgam and composite restorations) and/or root filling (endodontic treatment with or without metal post). To avoid synergism of artifacts produced by different teeth selected in the same scan, these teeth were selected from different sextants and teeth types. The flowchart of the complete dataset distribution is summarized in Fig. 1.

The ground-truth was generated using a previously validated AI tooth segmentation tool [4,5], where the operator firstly imported the CBCT dataset and determined the region of interest of the tooth to be segmented. Thereafter, the tool generated the 3D borders and contours were manually adjusted, if needed [14]. All segmentations were performed by a single expert and later verified by another expert.

2.2. AI architecture

The AI architecture has been previously described by Shaheen et al [5], which was further developed by a software company (Relu, Leuven, Belgium) [15]. The tooth segmentation pipeline consisted of multiple configured 3D U-Net network architectures [16]. As described in a previous study, the networks worked at different spatial resolutions with variable subtasks to create a high-resolution multi-class teeth segmentation [5].

To summarize, a two-step approach was used to perform the segmentation of each tooth since the convolutional deep neural networks have limitations regarding the use of large images (e.g. full FOV CBCT image). The first step was worked by combining U-net models to

Table 1
Sample characterization of the experimental group according to the type of filling within each type of tooth.

Types of filling	Type of tooth		
	Anterior (n)	Premolars (n)	Molars (n)
Only coronal filling in the proximal tooth surface	11	5	6
Only coronal filling in the occlusal tooth surface	NA	9	10
Coronal filling in multiples tooth surfaces (more than one surface restored)	12	15	15
Coronal filling + root filling (without metal post)	5	6	4
Coronal filling + root filling (with metal post)	5	5	5

NA, not applicable; n, number of teeth.

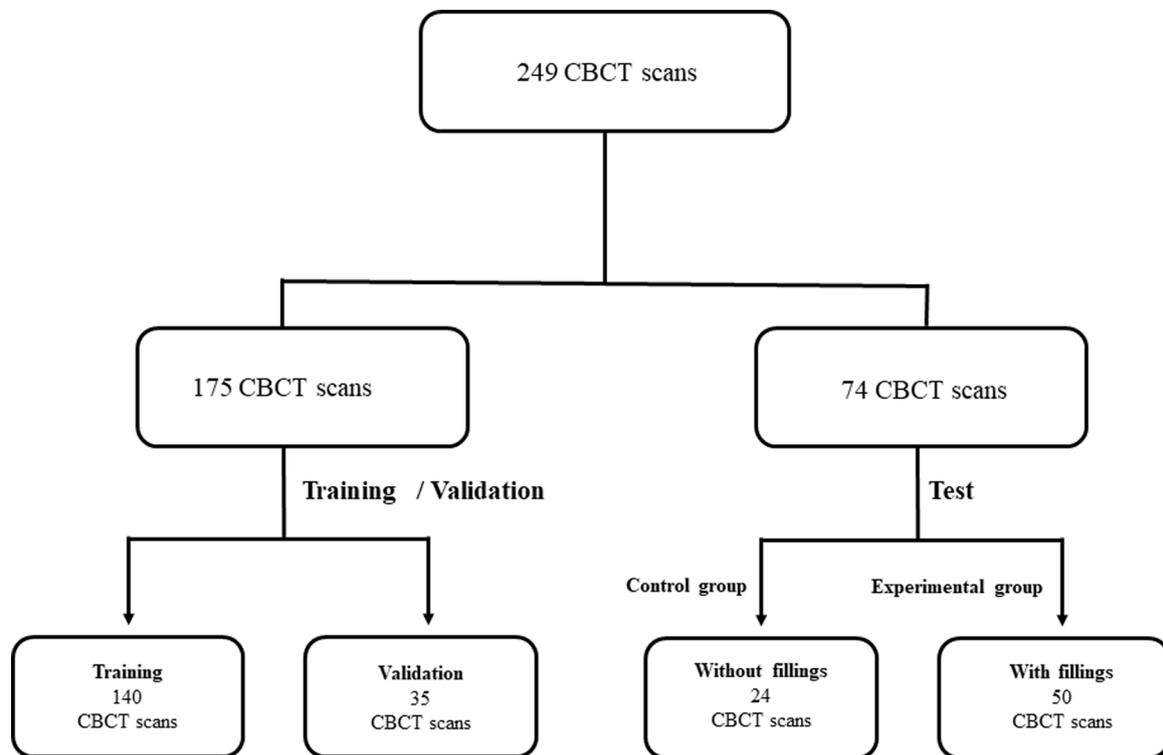


Fig. 1. Flowchart of the dataset used for the training, validation and testing of the AI-driven tool.

perform a rough segmentation and classify the different tooth instances. These models are made such that they can handle various types of FOVs. In the second step, a combination of U-net models was used to refine the rough tooth segmentation to produce a full-resolution tooth segmentation Fig. 2. summarizes the workflow to perform the automatic tooth

segmentation.

The models were implemented in PyTorch and optimized with Adam optimization for decreasing the learning rate and early stopping on the validation set. In addition, random spatial augmentations techniques were also applied, such as rotation, scaling and elastic deformation.

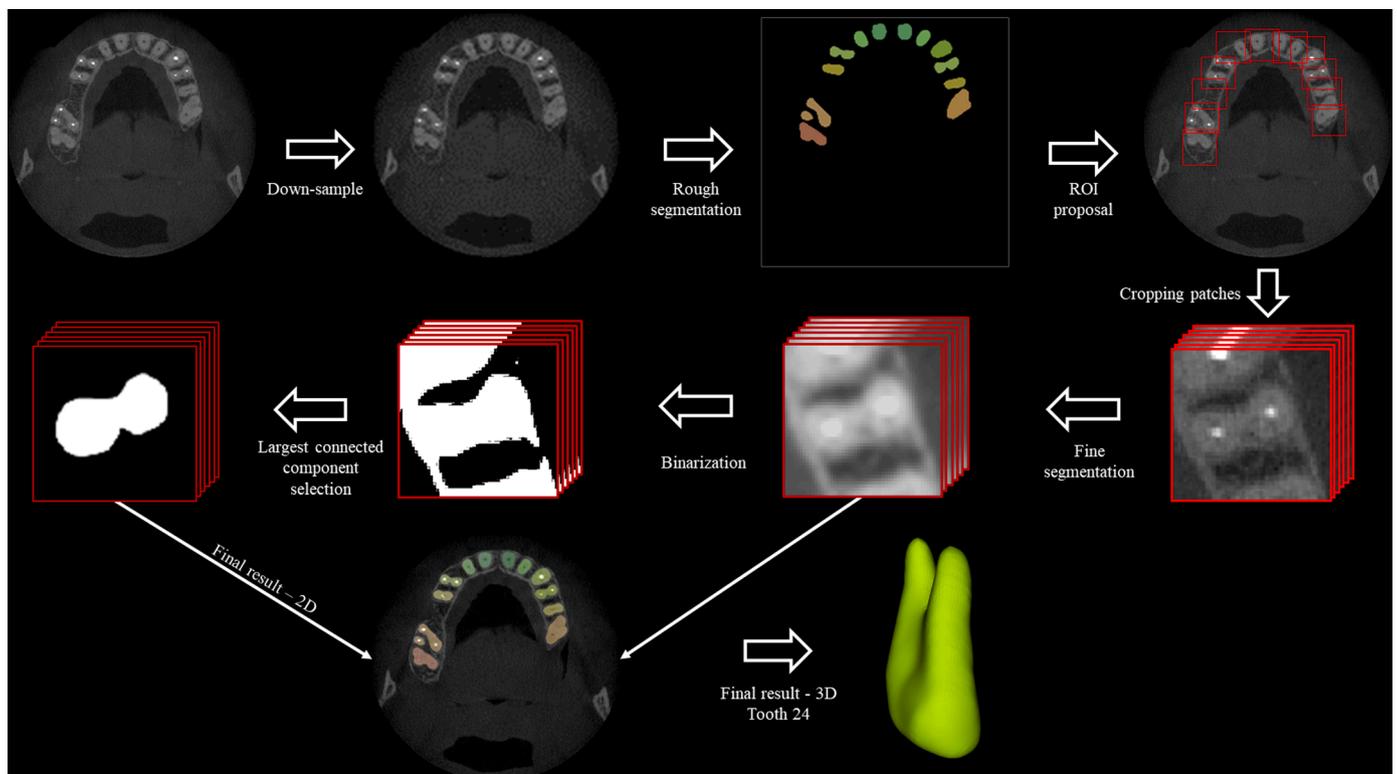


Fig. 2. Workflow of the multiple 3D U-net networks for automated tooth segmentation.

The AI model is available on an online cloud-based platform, the Virtual Patient Creator platform (Relu, Leuven, Belgium) [15]. This user-interactive platform is specialized to perform automated segmentation of craniomaxillofacial structures (maxilla, mandible, teeth, pharyngeal airway space, and maxillary sinus) from CBCT scans. Furthermore, access to the platform is granted following registration by the users.

2.3. Automated tooth segmentation

The automated AI-driven segmentation of teeth was performed using the Virtual Patient Creator platform. Following import of the CBCT image to the online platform in Digital Imaging and Communications in Medicine (DICOM) format, a segmentation map and the 3D model of each tool was automatically generated in Standard Triangle Language (STL) format. Additionally, the AI tool also provided the time in seconds needed to obtain the final segmented model.

2.4. Refinement of the automated tooth segmentation

The refinement of the automated tooth segmentation was performed by one expert (RCF) with an experience of over 5 years in oral and maxillofacial radiology. First, the expert evaluated the need for refinement using a binary scale: 0 (no refinement needed) and 1 (refinement needed). In cases where refinements were required, these were performed by the same user on the aforementioned online platform, which offered different tools such as brush (adding or removing voxels) and smart brush (grouping voxels based on their intensity) for the correction of potential mistakes in the automated segmentation. Following initial refinement, the platform provided a refined AI (R-AI) segmentation map with a 3D model of each tooth in STL format. The time spent for performing manual refinements was recorded with a digital stopwatch. Intra-observer agreement for the time duration required for the refinement task was assessed by refining 30% of the teeth twice at a period of 30 days.

For the assessment of the intra-examiner agreement, thirty days after the refinements were performed, the same expert re-evaluated 30% of the sample of teeth ($n = 80$) regarding the need for refinements.

2.5. Validation metrics

The AI and R-AI automatic segmentation maps were compared and evaluated using a confusion matrix. Four variables were obtained from the voxel-wise comparison:

- False positive (FP): voxels included in the segmentation map obtained from automated segmentation, but removed by the examiners after refinement.
- False negative (FN): voxels not included in the segmentation map obtained after automated segmentation, but included by the examiners after refinement.
- True positive (TP): voxels that should be segmented and correctly included in the final segmentation map.
- True negative (TN): voxels that should not be segmented and thus not included in the final segmentation map.

Based on these variables, the following metrics were utilized for assessing the accuracy of automated segmentation:

- 95% Hausdorff Distance (HD): Represents the 95 percentile of the maximum distance from a point in the segmentation map from the CNN model to its closest point in the segmentation map obtained after the refinement of the automated segmentation. The aim of using the 95 percentile of the HD is to eliminate a subset of outliers.

$$P_{95} \left(\min_{g \in G} \|p - g\|_2 \mid \min_{p \in P} \|g - p\|_2 \right)$$

- Intersection over union (IoU): Represents the degree of overlap between the segmentation map obtained from the CNN model and the segmentation map obtained after the refinement of the automated segmentation. An IoU of 1 means a perfect segmentation.

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

- Dice similarity coefficient (DSC): Represents the amount of intersection between the AI and R-AI segmentation maps. A DSC of 1 means a perfect segmentation.

$$\text{DSC} = \frac{2 \times \text{IoU}}{1 + \text{IoU}}$$

- Precision: Represents the fraction of correctly identified voxels among all the voxels considered as belonging to the tooth by CNN.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- Recall: Represents the rate of voxels that belonged to the tooth and was ideally identified by the CNN model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Accuracy: This ratio represents the rate of voxels that were correctly detected (TP and TN) among all voxels observed.

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

2.6. Timing analysis

The time duration was compared amongst manual, AI, and R-AI methods of tooth segmentation. A sample of 10% ($n = 27$) of teeth was randomly selected and segmented using the previously described online platform for assessing the timing of each approach:

- Manual method: One expert (RCF) performed the manual segmentation using the online platform. The time duration was recorded with a digital stopwatch from CBCT data import till the generation of a segmentation map. All teeth were segmented twice at a period for 30 days for assessing the intra-operator reliability of the following metrics: 95% HD, IoU, precision, recall, and accuracy.
- AI method: The online platform automatically provided the time required for obtaining the final segmentation map.
- R-AI method: The time needed to perform the refinement was recorded and added to the time taken by the AI method.

2.7. Quality of the automated segmentation analysis

Based on the overall mean time consumed for performing the automated segmentation (157 s) and amount of adjustments required for the refinement process, the automated segmentation was graded into three classes: no correction, minor correction, and major correction, where:

- a) No corrections: No adjustments were needed to the automated segmentation.
- b) Minor correction: The time needed for refinement was less than the overall mean time (< 157 s).
- c) Major correction: The time needed for refinement was more than the overall mean time (> 157 s).

A qualitative visual analysis was also performed for assessing under- and/or over-estimation of the automated segmentation by overlapping the STL files of the AI and R-AI segmented teeth in MeVisLab software (MeVis Research, Bremen, Germany). Fig. 3. illustrates the comparison between the AI and R-AI segmentation maps of teeth belonging to both control and experimental group. The reproducibility of the analysis was assessed by re-evaluating 30% of the sample at a period of 30 days.

2.8. Statistical analysis

Data were analyzed using SPSS v.24.0 (IBM Corp., Armonk, NY, USA). Intra-Class Correlation Coefficient (ICC) and weighted Kappa were applied for assessing the intra-observer agreement of the quantitative and qualitative variables, respectively. The descriptive analysis of data concerning the need for refinement was represented by absolute (n) and relative (%) frequency of the teeth that required correction following segmentation within each tested group (control and experimental) and each group of teeth (anterior, premolars and molars). The normality of the data was assessed with the Shapiro-Wilk test and the accuracy metrics were represented by mean and standard deviation (SD) values.

A two-way analysis of variance (ANOVA) with post-hoc Tukey test was applied to compare the mean values of each accuracy metric for assessing the effect of the studied factors (presence of filling and tooth group) and their interactions. Also, two-way ANOVA was applied to compare the mean values of the time duration to perform the tooth segmentation according to the segmentation method within each type of tooth. Finally, chi-square test was applied to evaluate the association between the groups tested and the type of tooth with the amount of corrections and the type of refinement needed. All analyses were performed with a significance level of 5% ($\alpha = 0.05$). For each statistical test, the power analysis was measured considering the minimum

difference between groups, their standard deviation, and the number of teeth within each group for ANOVA, and considering the chi-square value and the degree of freedom for chi-square test, which achieved a statistical power ranging from 70% to 99%.

3. Results

The manual segmentation demonstrated optimal values for all accuracy metrics, thereby indicating that the expert was optimally calibrated (95% HD - 0.27 mm, IoU - 0.92, DSC - 0.96; Precision - 0.95, Recall - 0.96, and Accuracy - 1.00). The ICC for the time required to perform both manual segmentation and refinements was 0.97. In addition, the intra-examiner agreements for the need for refinements on the automatic segmentation (weighted Kappa=0.92) and the qualitative visual estimation of the automated segmentation were almost perfect (weighted Kappa=0.86) [17].

Table 2 shows the frequency of teeth that required refinement following automated AI-driven segmentation based on the control and experimental groups and type of tooth. In the control group, molars required significantly more refinement than the anterior and premolars ($p = 0.001$). Whereas, the experimental group showed no significant difference amongst the type of tooth ($p = 0.08$). However, a higher percentage of teeth with fillings (81.8% of anterior, 57.5% of premolars, and 67.5% of molars teeth) required refinements.

Apart from the precision, all accuracy metrics of the automated segmentation showed a significant difference between the control and experimental groups ($p < 0.05$). The control group had a lower 95% HD value and higher IoU, DSC, recall and accuracy values compared to the experimental group. Based on the effect of the tooth type within each group, none of the accuracy metrics showed a significant difference ($p > 0.05$). However, anterior teeth in the experimental group showed higher 95% HD and lower IoU, DSC, and recall values compared to premolars and molars (Table 3).

Fig. 4 exhibits the time required by each segmentation technique (manual, AI, R-AI) according to the type of tooth group. The AI-driven segmentation showed the lowest time consumed to complete the tooth segmentation than the others methods regardless of the type of tooth ($p < 0.001$). The type of tooth did not influence the time for AI-segmentation ($p > 0.05$). However, the time consumed to perform the manual segmentation of molars was significantly higher than for anterior and premolars teeth ($p < 0.001$). Additionally, the time consumed for AI-R segmentation of molars was significantly higher than for anterior teeth ($p < 0.001$), but not significantly different than for premolars ($p > 0.05$).

Fig. 5A illustrates the grading of the automated segmentation based on the amount of required corrections, where a time of 157 s acted as the cut-off value for differentiating between minor and major corrections. The majority of teeth in the control group required no or minor corrections irrespective of the tooth type, whereas, the experimental group observed an increased need for refinement of the anterior and premolar teeth ($p < 0.001$). However, the correction requirement for molars showed an even distribution in both groups. Additionally, under-estimation of the segmentation was significantly higher compared to over-estimation in both groups ($p = 0.02$) (Fig. 5B). The errors found in the automated segmentation of the experimental group are illustrated in Fig. 6.

4. Discussion

Based on the hypothesis that the presence of restorative material artifacts could reduce the quality of the automated segmentation performed by an AI-driven tool, the present study was conducted to validate an innovative AI-driven tool for an accurate and time-efficient automated segmentation of teeth with and without dental filling material. Our results showed that the present AI-driven tool tested showed high accuracy and fast performance to generate 3D tooth models, even in the

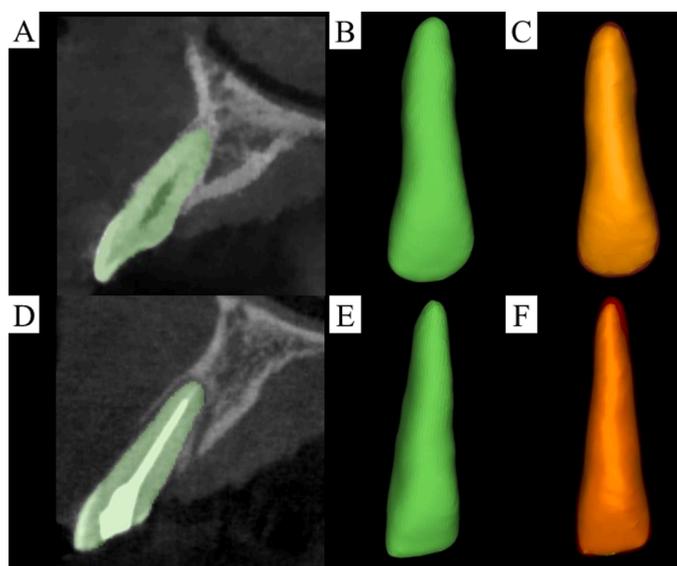


Fig. 3. Automated AI-driven segmentation of teeth without filling (A-C) and with filling (D-F). A and D – Cropped sagittal views of upper lateral incisors on CBCT images; B and E – three-dimensional models acquired from the AI-driven tool; C and F – Comparison of the AI segmentation (in white color) and refined-AI segmentation (in red color) models.

Table 2

Absolute (n) and relative (%) frequency of teeth that required corrections based on the tested groups (control and experimental) and tooth groups (anterior, premolars and molars).

Teeth group	Control			p-value	Experimental			p-value
	No corrections n (%)	Needed corrections n (%)	Total n (%)		No corrections n (%)	Needed corrections n (%)	Total n (%)	
Anterior	29 (87.9)	4 (12.1)	33 (100)	0.001*	6 (18.2)	27 (81.8)	33 (100)	0.085
Premolars	38 (95)	2 (5)	40 (100)		17 (42.5)	23 (57.5)	40 (100)	
Molars	26 (65)	14 (35)	40 (100)		13 (32.5)	27 (67.5)	40 (100)	

* Indicates statistically significant association between the tooth group and need for refinement in the control group. Statistical power analysis of 0.70 for the control group and 0.95 for the experimental group.

Table 3

Accuracy metrics of the automated AI-driven segmentation based on the tested groups (control and experimental) and tooth groups (anterior, premolars and molars).

Metrics	Teeth group	Control Mean ± SD	Experimental Mean ± SD
95% Hausdorff distance (HD) (mm)	Anterior	0.03 ± 0.16 Ba	0.25 ± 0.34 Aa
	Premolars	0.01 ± 0.09 Ba	0.17 ± 0.38 Ab
	Molars	0.08 ± 0.39 Aa	0.19 ± 0.43 Ab
Intersection over union (IoU)	Anterior	0.98 ± 0.04 Aa	0.91 ± 0.05 Bb
	Premolars	0.99 ± 0.03 Aa	0.94 ± 0.06 Ba
	Molars	0.97 ± 0.04 Aa	0.95 ± 0.04 Aa
Dice similarity coefficient (DSC)	Anterior	0.99 ± 0.02 Aa	0.95 ± 0.03 Bb
	Premolars	0.99 ± 0.02 Aa	0.97 ± 0.03 Ba
	Molars	0.98 ± 0.02 Aa	0.97 ± 0.03 Ba
Precision	Anterior	1.00 (0)	1.00 (0)
	Premolars	1.00 (0)	1.00 (0)
	Molars	1.00 (0)	1.00 (0)
Recall	Anterior	0.99 ± 0.04 Aa	0.91 ± 0.05 Bb
	Premolars	0.99 ± 0.03 Aa	0.94 ± 0.05 Ba
	Molars	0.97 ± 0.04 Aa	0.94 ± 0.04 Aa
Accuracy	Anterior	0.9997 ± 0.0017 Aa	0.9952 ± 0.0051 Bab
		0.9998 ± 0.0016 Aa	0.9968 ± 0.0047 Ba
	Premolars	0.9997 ± 0.0017 Aa	0.9952 ± 0.0051 Bab
		0.9998 ± 0.0016 Aa	0.9968 ± 0.0047 Ba
	Molars	0.9973 ± 0.0045 Ab	0.9940 ± 0.0050 Bb
		0.9973 ± 0.0045 Ab	0.9940 ± 0.0050 Bb

Different uppercases mean significant difference between control and experimental groups ($p < 0.05$)

Different lowercases mean significant difference between groups of teeth within each group tested ($p < 0.05$)

Statistical power analysis achieved for all metrics was higher than 0.90, except for precision and accuracy, which showed a power statistical analysis of 0.70.

presence of dental fillings.

Previously, various studies have developed and validated different CNN-based algorithms for automated tooth segmentation [2,4,5,8,12,13,18-20]. Although some of the aforementioned studies [2,4,5] have reported on the inclusion of teeth with fillings in the testing dataset, a lack of information exists related to the number of teeth with fillings. Therefore, the actual impact of the artifacts generated by fillings on the performance of CNNs is unknown. The present study focused towards assessing the influence of artifacts derived from dental fillings on the performance of an AI-driven tool by comparing control and experimental groups. Based on our findings, the presence of dental filling significantly impacted the performance of the AI-driven tool compared to the non-restored teeth. However, the observed statistical difference was quite low and cannot be considered clinically significant. Additionally, even in the presence of a coronal and/or root filling, the accuracy metrics were still higher (95% HD \leq 0.25 mm; IoU \geq 0.91; DSC \geq 0.95) than the CNN based automatic tooth segmentation approaches reported in the literature [2,4,5,8,20]. The high performance for the segmentation of teeth with fillings could be attributed to the fact that the training and validation datasets consisted of cases with dental fillings, enabling the CNN to learn and deal with these artifacts. Thus, confirming the robustness of the AI tool for the generation of accurate 3D

teeth models even in the presence of restorative material artifacts.

The AI-driven tool tested in the present study was based on a multiclass segmentation approach which has the ability to accurately segment the complete dental arch at the same time-point. Previous studies [4,5,12,13] have proposed different methods for automated tooth segmentation on CBCT scans with CNN's algorithms. Lahoud et al. developed and validated an AI-driven tool for a precise and fast automated tooth segmentation with the application of a feature pyramid network [4]. Although excellent results were obtained (IoU=0.87; DSC=0.93), the AI-tool was not trained to segment molar teeth. Duan et al. proposed a multiclass 3D tooth segmentation technique for tooth and pulp segmentation [13]. However, the generalizability of their approach is questionable owing to the acquisition of CBCT scans with a single device and similar scanning parameters. Recently, Shaheen et al [5]. also observed high accuracy of an automated AI-driven tool based on a multiclass approach (95% HD=0.56 mm; IoU=0.82; DSC=0.90). Nonetheless, the authors failed to assess its performance for the segmentation of restored dentition, where the presence of artifacts might inhibit optimal functioning of the tool. Thereby, comparison with prior evidence was not possible owing to the lack of a standardized approach and study design.

Furthermore, accuracy metrics have been widely accepted as an essential aspect for assessing the segmentation quality. However, the time required for segmentation has received less attention, which is a key factor when considering the clinical applicability of these tools. To the best of our knowledge, this is the first study assessing and comparing the time required for 3D individual tooth segmentation by manual, AI, and R-AI approaches of all types of teeth. The time analysis showed that the AI had the fastest performance (mean time: 29.8 s) compared to the other methods, and 37 times faster than the manual segmentation of a single tooth.

Regarding the influence of the tooth type on the performance of the AI-driven tool, the segmentation of the molars required more refinements compared to the other teeth in the control group without any impact on the tool's performance. This behavior could be attributed to the complex molar anatomy. It is important to highlight that the morphological complexity of molars might have influenced the segmentation more than the filling material, as observed from the comparison between both control and experimental groups, which showed an even distribution for the type of required correction. Compared to the control group, more refinements were required for the experimental group irrespective of the tooth type. The performance of the tool was lowest for segmenting anterior teeth (95% HD=0.25 mm; IoU=0.91; DSC=0.95) in the experimental group compared to premolars (95% HD=0.17 mm; IoU=0.94; DSC=0.97) and molars (95% HD=0.19 mm; IoU=0.95; DSC=0.97). These findings were in accordance with a previous study [5] which also showed a lower performance of the AI-tool for segmenting incisors. The distinct expression of the artifacts generated by high-density materials depending on their anatomical position in the dental arch might justify our findings, since the segmentation accuracy of the anterior teeth was lower than the posterior teeth in the experimental group [21,22].

When classifying the quality of the AI-segmentation based on the

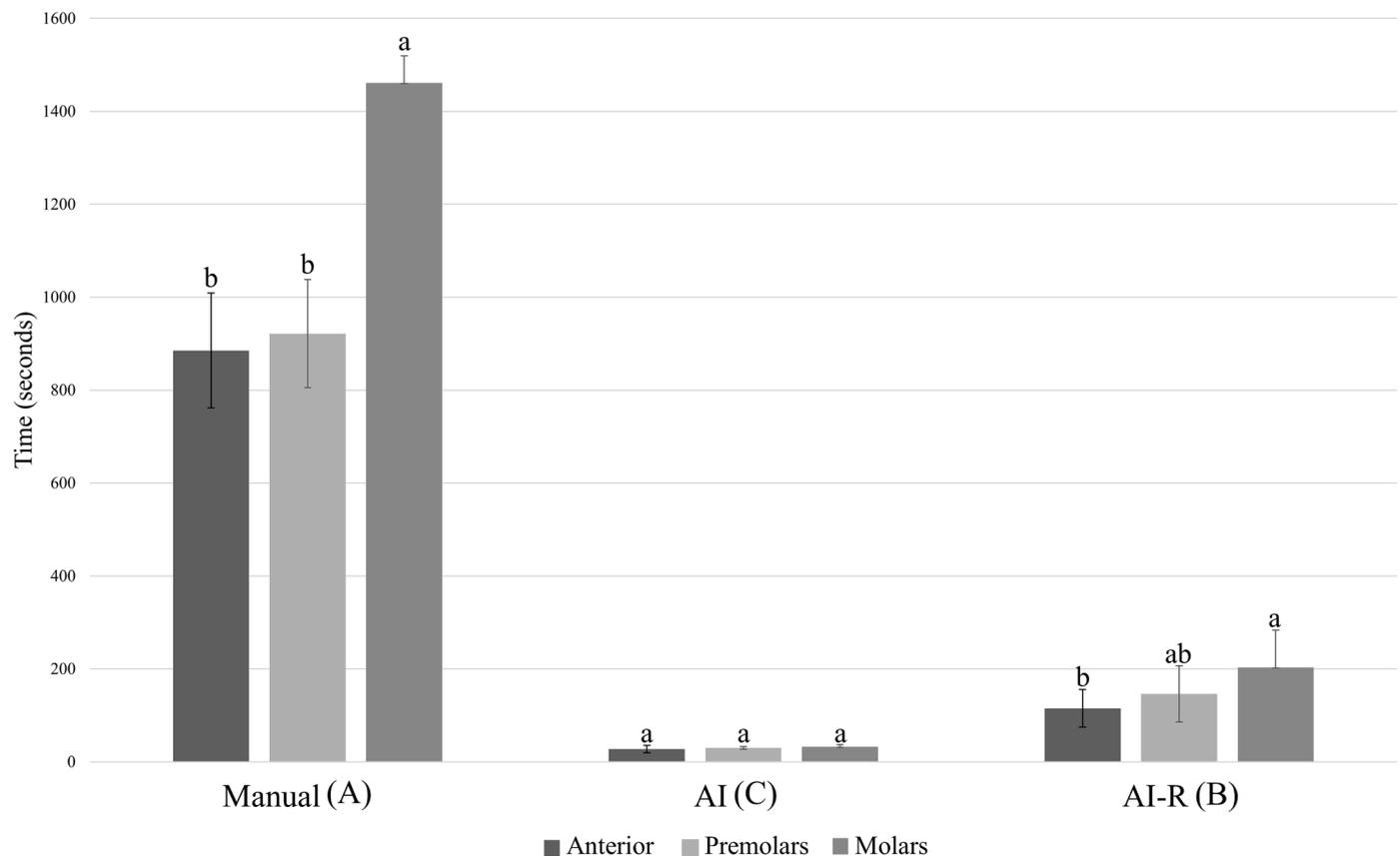


Fig. 4. Mean and standard deviation of time consumed for tooth segmentation according to manual, artificial intelligence (AI), and refined-AI method within each tooth group. Different uppercases represent a statistically significant difference between the segmentation methods ($p < 0.001$), and different lowercases represent a statistically significant difference between the teeth groups within each segmentation method ($p < 0.001$). Statistical power analysis of 0.99.

time consumed and amount of correction as proposed by Leite et al [3], the experimental group presented a higher frequency of minor and major corrections. However, the time taken to perform the refinement task was still low compared to manual segmentation. Moreover, it was not possible to compare our findings with the aforementioned study as their AI-driven tool was dedicated for automated tooth segmentation on panoramic radiographs.

The main error observed for the automated segmentation for both groups tested was the under-estimation. These results agree with a previous study's findings that also noticed that under-estimation was the main error for premolars and molars automated segmentation from panoramic radiography images using a CNN-based tool [3]. The lower signal to noise ratio, intensity inhomogeneity of CBCT images and the lack of definition between the edges of the tooth root and alveolar socket might justify the under-estimation observed for the control group of teeth in our study [23]. Additionally, to these inherent limitations aforementioned, the under-estimation observed in teeth with fillings also might be explained due to the expression of hypodense artifacts from the fillings that affect the recognition of the voxels that would encompass the teeth; consequently, a higher number of false-negative voxels was obtained. However, the presence of these errors in AI-driven segmentation is not expected to potentially cause adverse impacts in a clinical scenario, given that the present AI tool allows refinements to be carried out in automated segmentation. Based on the high accuracy metrics and the low time required to perform the refinements in the present study, the AI-R method still proves to be a clinically efficient method, especially in challenging cases, such as in the presence of fillings.

The present study is the pioneer for clinically proving the robustness of an AI-driven tool to perform automatic tooth segmentation even in the presence of coronal and root fillings, as seen through the high

accuracy metrics obtained (e.g. precision and accuracy above 0.99). These high accuracy results can be attributed to the constant search for more robust computational models based on CNN networks. First, this increase of computational knowledge can be justified by the heterogeneity of training and validating datasets (i.e. different image resolutions, FOV size, CBCT devices, and presence of artifacts). Furthermore, the current AI-driven tool is constantly supervised by experts in dentistry and retrained to provide increasingly accurate results, as shown in the present study.

The proposed AI-driven tool could offer a clinically acceptable alternative for tooth segmentation, to be applied in the digital dental workflow considering its time efficiency and high accuracy. However, as the CBCT dataset was acquired from only two devices, reported results should not be overstated or generalized to other devices. Therefore, future studies are required to test the performance of the AI tool with recruitment of data from different devices. Additionally, it is also recommended to assess the impact of other types of high-density materials, such as metallic crowns, orthodontic brackets, and dental implants on the performance of the tool. Finally, it is essential to highlight that our main aim was to evaluate the influence of fillings on the performance of AI-tool to generate 3D tooth models within each type of tooth. Thus, the testing sample was composed of different types of fillings common in dentistry. For these purposes, our sample showed sufficient statistical power ranging from 70% to 99%, the lower value being justified by the low variability between the groups tested in specific situations. However, the dataset size used would not allow to compare the AI-performance amongst all different types of dental fillings, considering that a representative sample of each type of filling within each type of tooth should be needed.

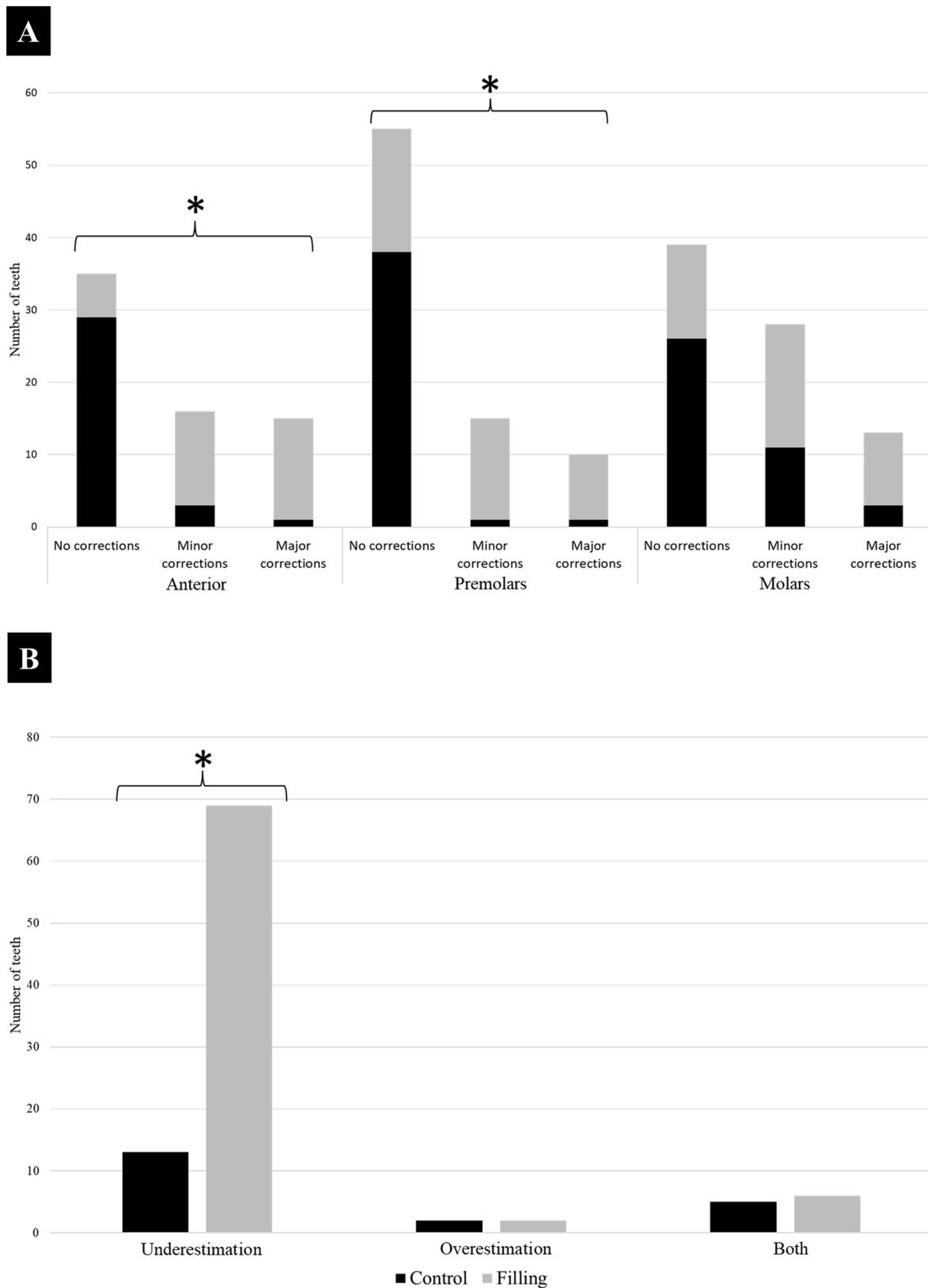


Fig. 5. Analysis of the quality of AI-driven automated tooth segmentation. A- Frequency of teeth that required correction based on a three-grade scale (no correction, minor correction, and major correction) within each tooth group (anterior, premolars, molars). Asterisk (*) indicates a statistically significant association between the type of tooth and the need for correction ($p < 0.001$). Statistical power analysis of 0.99. B- Frequency of teeth based on the type of error (under-estimation, over-estimation) within control and experimental group. The asterisk (*) indicates statistically significant association between the type of error and the group tested ($p = 0.02$). Statistical power analysis of 0.70.

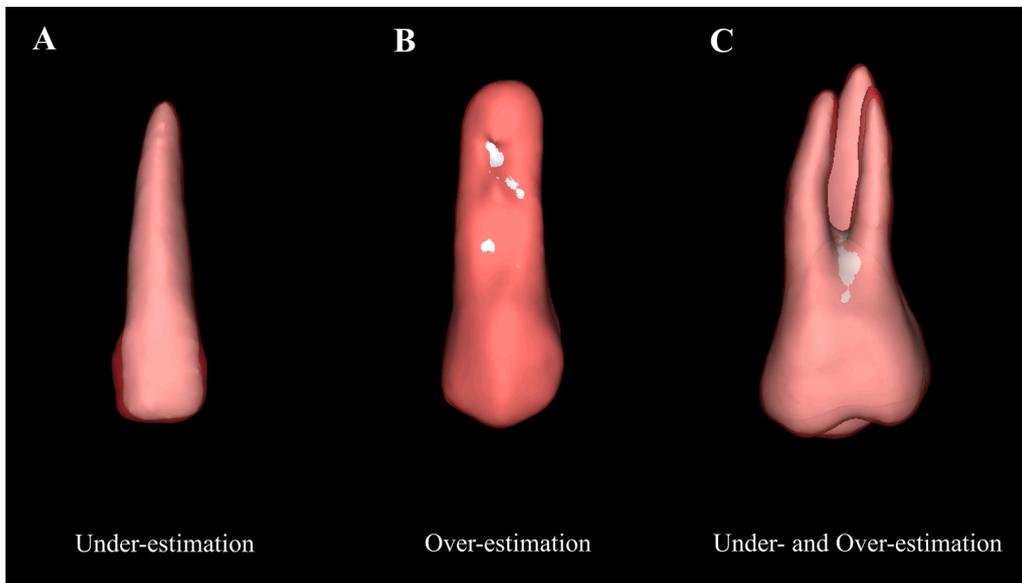


Fig. 6. Illustration of the errors on the AI-driven segmentation of teeth with fillings. The white segmentation map represents the segmentation before the refinements, and the red segmentation map represents the segmentation after the refinements are done. A – Automated segmentation underestimated in the crown of upper lateral incisor; B – Automated segmentation overestimated in the root of upper premolar; Automated segmentation with under- and overestimation in the furcation area and root of the upper molar.

5. Conclusions

The dental filling materials influenced the AI-driven tool performance, mainly for the anterior teeth. However, the AI-driven tool proposed showed high accuracy metrics and a time-efficient approach to provide 3D tooth models from CBCT images regardless of artifacts generated by these high-density materials and the type of tooth.

CRediT authorship contribution statement

Rocharles Cavalcante Fontenele: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft. **Maurício do Nascimento Gerhardt:** Conceptualization, Methodology, Validation, Software, Writing – original draft, Writing – review & editing. **Jáder Camilo Pinto:** Formal analysis, Investigation, Writing – review & editing. **Adriaan Van Gerven:** Conceptualization, Methodology, Software, Validation, Writing – review & editing. **Holger Willems:** Conceptualization, Methodology, Software, Writing – review & editing. **Reinhilde Jacobs:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision. **Deborah Queiroz Freitas:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

- [1] S. Shujaat, M.M. Bornstein, J.B. Price, R. Jacobs, Integration of imaging modalities in digital dental workflows - possibilities, limitations, and potential future developments, *Dentomaxillofac. Radiol.* 50 (2021), 20210268, <https://doi.org/10.1259/dmfr.20210268>.
- [2] S. Lee, S. Woo, J. Yu, J. Seo, J. Lee, C. Lee, Automated CNN-Based tooth segmentation in cone-beam CT for dental implant planning, *IEEE Access* 8 (2020) 50507–50518, <https://doi.org/10.1109/ACCESS.2020.2975826>.
- [3] A.F. Leite, A. Van Gerven, H. Willems, T. Beznik, P. Lahoud, H. Gaëta-Araujo, M. Vranckx, R. Jacobs, Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs, *Clin. Oral Investig.* 25 (2021) 2257–2267, <https://doi.org/10.1007/s00784-020-03544-6>.
- [4] P. Lahoud, M. EzEldeen, T. Beznik, H. Willems, A. Leite, A. Van Gerven, R. Jacobs, Artificial intelligence for fast and accurate 3-dimensional tooth segmentation on cone-beam computed tomography, *J. Endod.* 47 (2021) 827–835, <https://doi.org/10.1016/j.joen.2020.12.020>.
- [5] E. Shaheen, A. Leite, K.A. Alqahtani, A. Smolders, A. Van Gerven, H. Willems, R. Jacob, A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study, *J. Dent.* (2021), 103865, <https://doi.org/10.1016/j.jdent.2021.103865>.
- [6] B. Hassan, P.C. Souza, R. Jacobs, S. de Azambuja, P. Berti, Van der Stelt, Influence of scanning and reconstruction parameters on quality of three-dimensional surface models of the dental arches from cone beam computed tomography, *Clin. Oral Investig.* 14 (2010) 303–310, <https://doi.org/10.4047/jap.2017.9.5.381>.
- [7] L. Wang, K.C. Chen, Y. Gao, F. Shi, S. Liao, G. Li, J. Yan, P.K. Lee, B. Chow, N. X. Liu, J.J. Xia, D. Shen, Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization, *Med. Phys.* 41 (2014), 043503, <https://doi.org/10.1118/1.4868455>.
- [8] Z. Cui, C. Li, W. Wang, Toothnet: automatic tooth instance segmentation and identification from cone beam ct images, in: , 2019, pp. 6361–6370, <https://doi.org/10.1109/CVPR.2019.006653>, 2019-June.
- [9] P.J. Verhelst, A. Smolders, T. Beznik, J. Meewis, A. Vandemeulebroucke, E. Shaheen, A. Van Gerven, H. Willems, C. Politis, R. Jacobs, Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography, *J. Dent.* (2021), 103786, <https://doi.org/10.1016/j.jdent.2021.103786>.
- [10] J.J. Hwang, Y.H. Jung YH, B.H. Cho, M.S. Heo, An overview of deep learning in the field of dentistry, *Imaging. Sci. Dent.* 49 (2019) 1–7, <https://doi.org/10.5624/isd.2019.49.1.1>.
- [11] S. Shujaat, O. Jazil, H. Willems, A. Van Gerven, E. Shaheen, C. Politis, R. Jacobs, Automated segmentation of the pharyngeal airway space with convolutional neural network, *J. Dent.* 111 (2021), 103705, <https://doi.org/10.1016/j.jdent.2021.103705>.
- [12] H. Wang, J. Minnema, K.J. Batenburg, T. Forouzanfar, F.J. Hu, G. Wu, Multiclass CBCT Image Segmentation for Orthodontics with Deep Learning, *J. Dent. Res.* 100 (2021) 943–949, <https://doi.org/10.1177/00220345211005338>.
- [13] W. Duan, Y. Chen, Q. Zhang, X. Lin, X. Yang, Refined tooth and pulp segmentation using U-Net in CBCT image, *Dentomaxillofac. Radiol.* (2021), 20200251, <https://doi.org/10.1259/dmfr.20200251>.
- [14] M. EzEldeen, G. Van Gorp, J. Van Dessel, D. Vandermeulen, R. Jacobs, 3-dimensional analysis of regenerative endodontic treatment outcome, *J. Endod.* 41 (2015) 317–324, <https://doi.org/10.1016/j.joen.2014.10.023>.
- [15] relu, Virtual Patient Creator, Leuven, Belgium. (n.d.). <https://creator.relu.eu/>.
- [16] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 9901 (2016) 424–432, https://doi.org/10.1007/978-3-319-46723-8_49.LNCS.
- [17] J.R. Landis, G.G. Koch, An application of hierarchical kappa type statistics in the assessment of majority agreement among multiple observers, *Biometrics* 33 (1977) 363–374, <https://doi.org/10.2307/2529786>.
- [18] Q. Li, K. Chen, L. Han, Y. Zhuang, J. Li, J. Lin, Automatic tooth roots segmentation of cone beam computed tomography image sequences using U-net and RNN, *J. Xray. Sci. Technol.* 28 (2020) 905–922, <https://doi.org/10.3233/XST-200678>.
- [19] Y. Rao, Y. Wang, F. Meng, J. Pu, J. Sun, Q. Wang, A symmetric fully convolutional residual network with DCRF for accurate tooth segmentation, *IEEE Access* 8 (2020) 92028–92038, <https://doi.org/10.1109/ACCESS.2020.2994592>.

- [20] Y. Chen, H. Du, Z. Yun, S. Yang, Z. Dai, L. Zhong, Q. Feng, W. Yang, Automated segmentation of individual tooth in dental CBCT images from tooth surface map by a multi-task FCN, *IEEE Access* 8 (2020) 97296–97309, <https://doi.org/10.1109/ACCESS.2020.2991799>.
- [21] R.C. Fontenele, A. Farias Gomes, L.P.L. Rosado, F.S. Neves, D.Q. Freitas, Mapping the expression of beam hardening artefacts produced by metal posts positioned in different regions of the dental arch, *Clin. Oral. Investig.* 25 (2021) 571–579, <https://doi.org/10.1007/s00784-020-03494-z>.
- [22] P.M. Queiroz, G.M. Santaella, T.D. da Paz, D.Q. Freitas, Evaluation of a metal artefact reduction tool on different positions of a metal object in the FOV, *Dentomaxillofac. Radiol.* 46 (2017), 20160366, <https://doi.org/10.1259/dmfr.20160366>.
- [23] Y. Wang, S. Liu, G. Wang, Y. Liu, Accurate tooth segmentation with improved hybrid active contour model, *Phys. Med. Biol.* 64 (2018), <https://doi.org/10.1088/1361-6560/aaf441>. Article 015012.