**DentalSegmentator: robust deep learning-based CBCT image segmentation**

*Original Research Report*

Gauthier Dot[1,2,3]*, Akhilanand Chaurasia[4], Guillaume Dubois[1,5], Charles Savoldelli[6], Sara Haghighat[7], Sarina Azimian[8], Ali Rahbar Taramsari[9], Gowri Sivaramakrishnan[10], Julien Issa[11,12], Abhishek Dubey[13], Thomas Schouman[1,14], Laurent Gajny[1]

## Abstract

Delineation of anatomical structures on dento-maxillo-facial (DMF) computed tomography (CT) or cone beam computed tomography (CBCT) scans is greatly needed for an increasing number of digital dentistry tasks. Following this process, called segmentation, three-dimensional (3D) patient-specific models can be exported for visualization, treatment planning, intervention, and follow-up purposes. Although several methods based on deep learning (DL) have been proposed for automating this task, there is no thoroughly evaluated publicly available tool offering segmentation of the anatomical structures needed for digital dentistry workflows. In this work, we propose and evaluate DentalSegmentator, a tool based on the nnU-Net deep learning framework, for fully automatic segmentation of 5 anatomic structures on DMF CT and CBCT scans: maxilla and upper skull, mandible, upper teeth, lower teeth and mandibular canal. A retrospective sample of 470 CT and CBCT scans was used as a training/validation set. The performance and generalizability of the tool was evaluated by comparing segmentations provided by experts and automatic segmentations on 2 hold-out test datasets: an internal dataset of 133 CT and CBCT scans acquired before orthognathic surgery and an external dataset of 123 CBCT scans randomly sampled from routine examinations in 5 institutions. In our internal test dataset ($n$ = 133), the mean overall results were a Dice similarity coefficient (DSC) of 92.2 ± 6.3% and a normalized surface distance (NSD) of 98.2 ± 2.2%. In our external test dataset ($n$ = 123), the mean overall results were a DSC of 94.2 ± 7.4% and a NSD of 98.4 ± 3.6%. The results obtained on this highly diversified dataset demonstrate that our tool can provide fully automatic and robust multiclass segmentation for DMF (CB)CT scans. To encourage the clinical deployment of DentalSegmentator, our pretrained nnU-Net model is made publicly available along with an extension for the 3D Slicer software.

**Keywords**: dental informatics, patient-specific modeling, cone-beam computed tomography, computer-assisted radiographic image interpretation, computer-assisted surgery, artificial intelligence

[1] Institut de Biomecanique Humaine Georges Charpak, Arts et Metiers Institute of Technology, Paris, France;

[2] UFR Odontologie, Universite Paris Cité, Paris, France;

[3] AP-HP, Hopital Pitie-Salpetriere, Service de Medecine Bucco-Dentaire, Paris, France;

[4] Department of Oral Medicine and Radiology, Faculty of Dental Sciences, King George Medical University, Lucknow, Uttar Pradesh, India;

[5] Materialise France, Malakoff, France;

[6] Head and Neck Institute, department of oral and maxillofacial surgery. University Hospital of Nice;

[7] School of Dentistry, Shiraz University of Medical Sciences, Shiraz, Iran;

[8] Research Committee, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran;

[9] Private dentist practitioner, Rasht, Iran;

[10] Speciality Dental Residency Program, Primary Health Care centers, Bahrain;

[11] Chair of Practical Clinical Dentistry, Department of Diagnostics, Poznan University of Medical Sciences, Poznan, Poland;

[12] Doctoral School, Poznan University of Medical Sciences, Poznan, Poland;

[13] Department of Oral Medicine and Radiology, Maharana Pratap Dental College, Kanpur, India ;

[14] Medecine Sorbonne Universite, AP-HP, Hopital Pitie-Salpetriere, Service de Chirurgie Maxillo-Faciale, Paris, France.

* Corresponding author

## Introduction

The clinical practice of dentistry has radically evolved in the last years, due to the increasing use of digital three-dimensional (3D) data that can be gathered from dento-maxillo-facial (DMF) computed tomography (CT) scans, cone beam computed tomography (CBCT) scans, intraoral scanners or facial scanners. Those data have been shown to improve diagnosis, treatment planning, intervention and patient follow-up in several domains of dentistry (SEDENTEXCT project 2012; Kapila and Nervina 2015; Assiri et al. 2020; Chogle et al. 2020). More specifically, patient-specific 3D models derived from (CB)CT scans are already used for educational purposes, computer-assisted surgical planning or navigation, tooth auto-transplantation planning or virtual treatment planning for orthodontics treatments (Alkhayer et al. 2020; Borohovitz et al. 2021; Hu et al. 2023; Lejnieks et al. 2023). In the near future, those 3D models could be a key part of precision medicine in dental practice based on finite element technologies, offering the opportunity to individually assess anatomical and biomechanical characteristics and adapt treatment options accordingly (Lahoud et al. 2022).

To obtain a patient-specific 3D model from a (CB)CT scan, the anatomical structures of interest must be carefully delineated on the 3D image slices, a process called segmentation. The most common workflows require the segmentation of jaws (maxilla and mandible), teeth (upper and lower) and mandibular canal. When performed manually, this segmentation process takes an expert 2 to 5 hours

to complete (Wang et al. 2021; Cui et al. 2022). Currently, the gold standard for 3D DMF image segmentation is the semi-automatic method where automatic segmentations are refined manually by an expert (Wallner et al. 2019). In recent years, several research reports have shown that deep learning-based (DL) methods could fully automate this task with results on par with experts (Shaheen et al. 2021; Wang et al. 2021; Cui et al. 2022; Dot et al. 2022; Gillot et al. 2022; Ayidh Alqahtani et al. 2023). Several commercially available solutions already claim to use DL methods for CBCT segmentation (Ezhov et al. 2021; Ileșan et al. 2023; Nogueira-Reis et al. 2024).

Despite those promising results, a recent systematic review on automatic tooth segmentation approaches from CBCT scans revealed that most of the studies were at high risk of bias regarding data selection, leading to a potential overestimation of the accuracy of the methods (Polizzi et al. 2023). Most of the published studies reported results from cross-validation approaches or small-sized hold-out test dataset (less than 50 CBCT scans), which could be insufficient to evaluate the robustness and generalizability of the methods in actual clinical settings (Schwendicke et al. 2021). In a previous work, we evaluated a deep learning-based segmentation method for preoperative orthognathic CT scans (Dot et al. 2022). We showed that this method was clinically viable on a hold-out test dataset of 153 CT scans. However, the generalizability of this research could not be assessed as our dataset did not include any external test dataset and was limited to CT scans data.

As an effort to help the deployment and broad evaluation of rapidly evolving research, the biomedical computer imaging community has relied heavily on open research. This has led to the development of international challenges such as The Medical Segmentation Decathlon (Antonelli et al. 2022), the sharing of DL frameworks such as nnU-Net (Isensee et al. 2021) and the sharing of pre-trained DL models for various segmentation tasks such as TotalSegmentator (Wasserthal et al. 2023). For DMF CBCT segmentation, to our knowledge, only one pre-trained DL model is publicly shared (Gillot et al. 2022). This model, integrated to a 3D Slicer extension called Slicer Automated Dental Tools, allows segmentation of 4 anatomical structures: mandible, maxilla, cranial base and cervical vertebrae. Unfortunately, this tool does not delineate the teeth from the jaws, which is a critical limitation for most digital dentistry and surgery workflows.

In this work, our main objective was to propose and evaluate a novel tool for multiclass DMF CT and CBCT image segmentation called DentalSegmentator. We thoroughly evaluated the performance of the tool on 2 hold-out test datasets acquired from routine clinical practice in 7 clinical centers.

## Materials and Methods

A DL framework was trained on an internal dataset for automatic segmentation of DMF CT and CBCT scans. We compared the results of our DL-based method (the index test) with those obtained by semi-automatic segmentation (the reference test) on 2 hold-out test datasets. Our outcome set included both volume-based and surface-based metrics. This study was approved by an appropriate Institutional Review Board (IRB No. CRM-2001-051b) and its reporting followed recently published recommendations on artificial intelligence in dental research (Schwendicke et al. 2021).

**Internal Dataset**

*Patient selection*

Data of our internal dataset were selected from a retrospective sample of consecutive patients having undergone orthognathic surgery in 2 French maxillofacial surgery departments. Patients referred to these public centers presented a wide variety of dentofacial deformities, came from various socioeconomic backgrounds and were ethnically diverse. Patients were considered for inclusion regardless of the dental deformity they presented, with no minimum age. Exclusion criteria were refusal to participate in the research and lack of industry-certified CT or CBCT scan segmentation. 603 subjects (453 CT scans, 150 CBCT scans) were included in our internal dataset.

*Data characteristics*

All the scans had a full-head field of view (FOV). The median in-space pixel size of the scans was $0.43*0.43mm^2$ and their median slice thickness was 0.31mm. Most CT scans (*n* = 417) were obtained using a GE Healthcare Discovery (GEHC) CT750HD scanner and all CBCT scans (*n* = 150) were obtained using a Carestream CS 9600 scanner. Scans were randomly distributed among a train/validation set (*n* = 470; 374 CT scans and 96 CBCT scans) and an internal test set (*n* = 133; 79 CT scans and 54 CBCT scans). In the internal test set, 91% of the scans exhibited metal artefacts. Descriptive characteristics of the data are detailed in Table 1.

*Ground truth segmentation process (Reference Test)*

The patients' treatments involved segmentation of the 3D scans prior to our study. The ground truth segmentations were used for diagnosis, computer-aided surgical planning, and manufacturing of 3D-printed personalized surgical guides and fixation implants. This was done according to a certified internal procedure by Materialise (Leuven, Belgium), which cannot be fully described here for reasons of confidentiality. The 2-step procedure involved semi-automatic segmentations refined manually by an initial operator [Step 1]. The segmentations were then verified slice-by-slice for validation by a senior operator [Step 2], with a focus on the regions of interest (the external surface of the bones, teeth, and mandibular canals). Steps 1 and 2 were repeated until the segmentations were approved

and certified for clinical use. This process resulted in five segmentation masks: maxilla/upper skull; mandible; upper teeth; lower teeth and both mandibular canals.

| | Train/Validation | Internal Test | External Test |
|---|---|---|---|
| Number of scans | 374 CT + 96 CBCT | 79 CT + 54 CBCT | 123 CBCT |
| Median voxel size ($mm^3$) | 0.43 * 0.43 * 0.31 | 0.43 * 0.43 * 0.31 | 0.16 |
| Number of scans by Device | | | |
| GEHC Discovery CT750 HD | 353 | 75 | |
| Other CT Device | 21 | 4 | |
| Carestream CS 9600 CBCT | 96 | 54 | |
| Carestream CS 9300 CBCT | | | 29 |
| Vatech Smart Plus CBCT | | | 25 |
| Dentium Rainbow CBCT | | | 27 |
| Planmeca Promax 3D | | | 12 |
| Sirona Orthophos XG 3D | | | 30 |
| Metal artifacts, no. (%) | | | |
| Orthodontic materials | 364 (78.1) | 107 (80.5) | 3 (2.4) |
| Metallic dental filling/crown | 168 (35.7) | 44 (33.1) | 49 (39.8) |
| No metallic artifact | 57 (12.1) | 12 (9.0) | 71 (57.7) |

**Table 1**: Characteristics of the data in the train/validation, internal test and external test datasets.

**External Test Dataset**

*Patient selection*

Data of our external test dataset were retrospectively, randomly sampled from routine CBCT examinations in 5 private centers located in India. All the subjects were referred for a CBCT scan for various reasons such as surgical planning, orthodontic management of impacted teeth, temporomandibular joint (TMJ) disorders or diagnosis of cysts of the jaws. Patients were considered for inclusion regardless of the condition they presented, with no minimum age. The only exclusion criterion was refusal to participate in the research. 123 subjects (123 CBCT scans) were included in our external test dataset.

*Data characteristics*

The FOV of the scans ranged from full-head to localized on anatomical parts (maxilla and mandible or only part of the maxilla or mandible). The median voxel size of the CBCT scans was 0.16$mm^3$. The scans were acquired using 5 CBCT devices: Vatech Smart Plus (*n* = 25), Carestream CS 9300 (*n* = 29), Dentium

Rainbow CBCT (*n* = 27), Planmeca Promax 3D (*n* = 12) and Sirona Orthophos XG 3D (*n* = 30). 42% of the CBCT scans showed metallic artefacts. Descriptive characteristics of the data are detailed in Table 1.

*Ground truth segmentation process (Reference Test)*

The CBCT scans were segmented specifically for this study, following a semi-automatic 3-step approach. First, the CBCT scans were segmented automatically using a previously published deep-learning model (Dot et al. 2022) [Step 1]. The proposed segmentations were then corrected manually by 5 dentists familiar with 3D image visualization and trained for the task [Step 2]. Finally, the segmentations were verified slice-by-slice and corrected where necessary by a senior expert (a dentist with over five years of experience in 3D image evaluation) [Step 3]. This process resulted in five segmentation masks: maxilla/upper skull; mandible; upper teeth; lower teeth and both mandibular canals.

**Deep-Learning based segmentation (Index Test)**

*Training*

The nnU-Net deep learning framework (version 2.2.1) was used as an out-of-the-box tool, following instructions given by its authors (Isensee et al. 2021). Our raw train/validation internal dataset was used to automatically configure preprocessing, network architecture, and 3D full resolution U-Net training pipelines. No modifications were made in setting the nnU-Net hyperparameters and data augmentation strategy. Training time was about 24 hours on our laboratory workstation (CPU AMD Ryzen 9 3900X 12-Core; 128Gb RAM; GPU Nvidia Titan RTX 24Gb).

*Inference*

Inference (prediction made by the trained model) was performed once on our internal and external test datasets following nnU-Net guidelines.

**Evaluation**

Quantitative evaluation of the model performance was performed on our internal and external test datasets by comparing ground truth segmentations (reference test) with DL-based segmentations (index test) for each of the 5 segmentation masks. We followed the recommendations of the Metrics Reloaded project (Maier-Hein et al. 2023) by using both volume-based Dice similarity coefficient (DSC) and surface-based normalized surface distance (NSD). We set the tolerance for NSD at 1mm, consistent with recent international challenges in biomedical imaging (Antonelli et al. 2022). Recent studies have shown that NSD was more strongly correlated with the amount of time needed to correct a segmentation for clinical use compared to classic metrics such as DSC (Nikolov et al. 2021).

### Statistical Analysis

Continuous variables were presented as mean ± standard deviation and categorical variables were expressed as numbers and percentages. DSC and NSD results were presented as percentages (%). As the results were nonparametric (Shapiro-Wilk normality test), the Kruskal-Wallis test was used to compare DSC and NSD results from different CT/CBCT devices; when significant, we used post-hoc Dunn's test for comparing each group; $p$ values <0.05 were considered statistically significant. All data were analysed with Python (v.3.7) and R Statistical Software (v4.2.2; R Core Team 2022).

## Results

### Quantitative evaluation

Inference time was about 1 to 2 minutes for one 3D scan when performed on our laboratory workstation. In our internal test dataset ($n$ = 133), the mean overall results were a DSC of 92.2 ± 6.3% and a NSD of 98.2 ± 2.2%. In our external test dataset ($n$ = 123), the mean overall results were a DSC of 94.2 ± 7.4% and a NSD of 98.4 ± 3.6%. The mean results for each label and the distribution of results are shown in Figure 1.

Our statistical analysis showed similar results for both DSC and NSD metrics. In the internal test dataset, there was no statistical difference when comparing scans obtained with various devices. In the external test dataset, scans obtained with Carestream 9300 showed statistically superior segmentation results than those obtained with Sirona Orthophos XG 3D. Overall, the results obtained on the external test dataset were statistically superior to those obtained on the internal test dataset.

### Three-Dimensional Visualization

Six subjects representative of our test dataset were chosen to illustrate the segmentation results and the diversity of our CBCT data (Figure 2). When segmentation failures occurred, they were most of the time under-segmentations of thin bony parts (resulting in holes in the maxilla or mandible inferior border) and missing mandibular canal parts (Figure 3).

| Metric, mean ± SD [%] | | Maxilla / Upper Skull | Mandible | Upper Teeth | Lower Teeth | Mandibular Canal |
|---|---|---|---|---|---|---|
| | DSC | 94.6 ± 2.7 | 94.5 ± 1.6 | 95.3 ± 1.6 | 95.3 ± 1.6 | 81.4 ± 6.2 |
| | NSD | 97.1 ± 2.9 | 98.3 ± 1.1 | 98.9 ± 1.0 | 98.7 ± 1.0 | 98.0 ± 3.2 |

| Metric, mean ± SD [%] | | Maxilla / Upper Skull | Mandible | Upper Teeth | Lower Teeth | Mandibular Canal |
|---|---|---|---|---|---|---|
| | DSC | 93.0 ± 3.8 | 96.2 ± 4.0 | 96.7 ± 3.5 | 97.3 ± 1.9 | 87.0 ± 12.9 |
| | NSD | 97.2 ± 2.8 | 98.6 ± 3.5 | 98.8 ± 3.3 | 99.4 ± 1.9 | 97.7 ± 5.4 |

**Figure 1**: DSC and NSD results on our internal (left) and external (right) test datasets. *SD: Standard Deviation.*

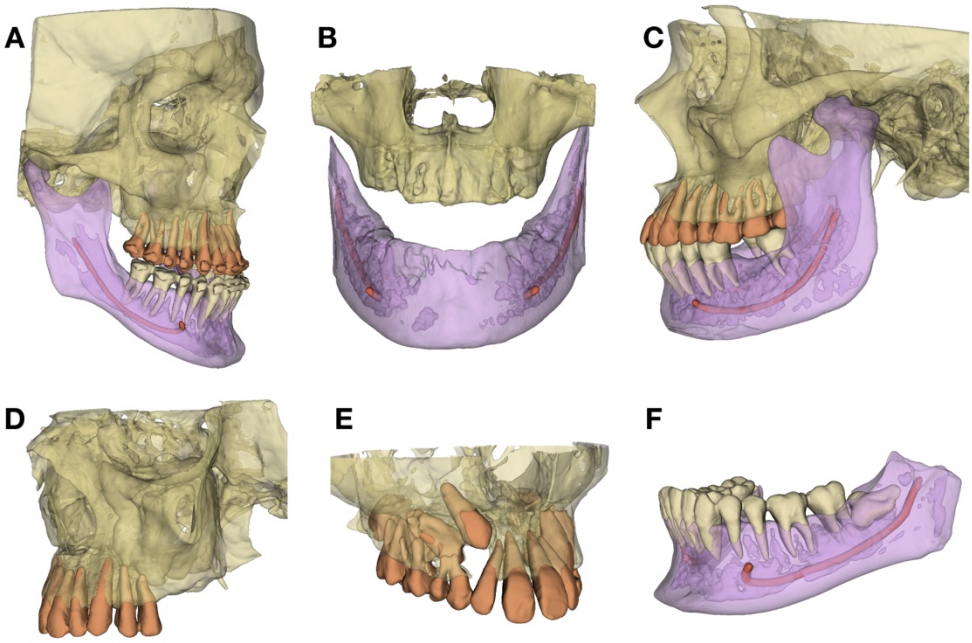


**Figure 2**: 3D surface models for 6 subjects representative of the diversity and the challenges arising from our test CBCT dataset. (**A**) Class III maxillo-mandibular deformity, before orthognathic surgery; (**B**) Edentulous jaws; (**C**) Left condylar hyperplasia; (**D**) Upper posterior dental gap; (**E**) Maxilla with impacted teeth; (**F**) Mandible with impacted third molar.

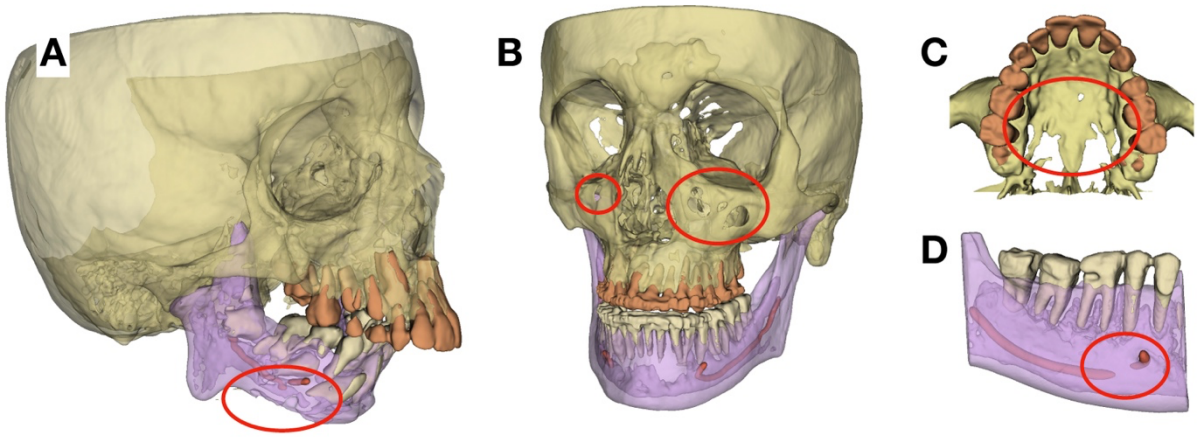


**Figure 3**: 3D surface models for 4 subjects, exhibiting some typical failures (red circles). (**A**) Under-segmentation of the mandibular inferior border and mandibular canal; (**B**) Under-segmentation of the anterior maxillary sinus walls; (**C**) Under-segmentation of the palate; (**D**) Discontinuity of the mandibular canal.

**DentalSegmentator model sharing and 3D Slicer extension**

Our pre-trained nnU-Net model is made publicly available (Dot 2024 Mar 18). This model can be used out-of-the-box via the nnU-Net version 2.2.1 command-line interface.

To encourage the use of our DL method by clinicians, we also propose an implementation in a user-friendly interface. DentalSegmentator is a publicly available extension for the 3D Slicer software, which is a free, open source software for visualization, processing, and analysis of medical 3D images (Kikinis et al. 2014). The extension, downloadable from 3D Slicers v5.7.0 extension, offers an easy-to-use approach for DMF (CB)CT scans automatic segmentation and 3D patient-specific model export (Figure 4). When needed, slice-by-slice verification and manual refinement of the segmentations can be performed directly in the 3D Slicer software. More information about the extension and its code are shared on Github platform: https://github.com/gaudot/SlicerDentalSegmentator

Those tools work on all computer platforms but require a compatible graphics processing unit (GPU) with at least 4Gb of Ram for fast results. On a laptop personal computer (CPU Intel Core i7-13850HX; 32Gb RAM; GPU NVIDIA RTX 2000 8Gb), the segmentation with the 3D Slicer extension required a mean time of 178 ± 100 seconds on 10 randomly selected CT and CBCT scans from our test dataset.

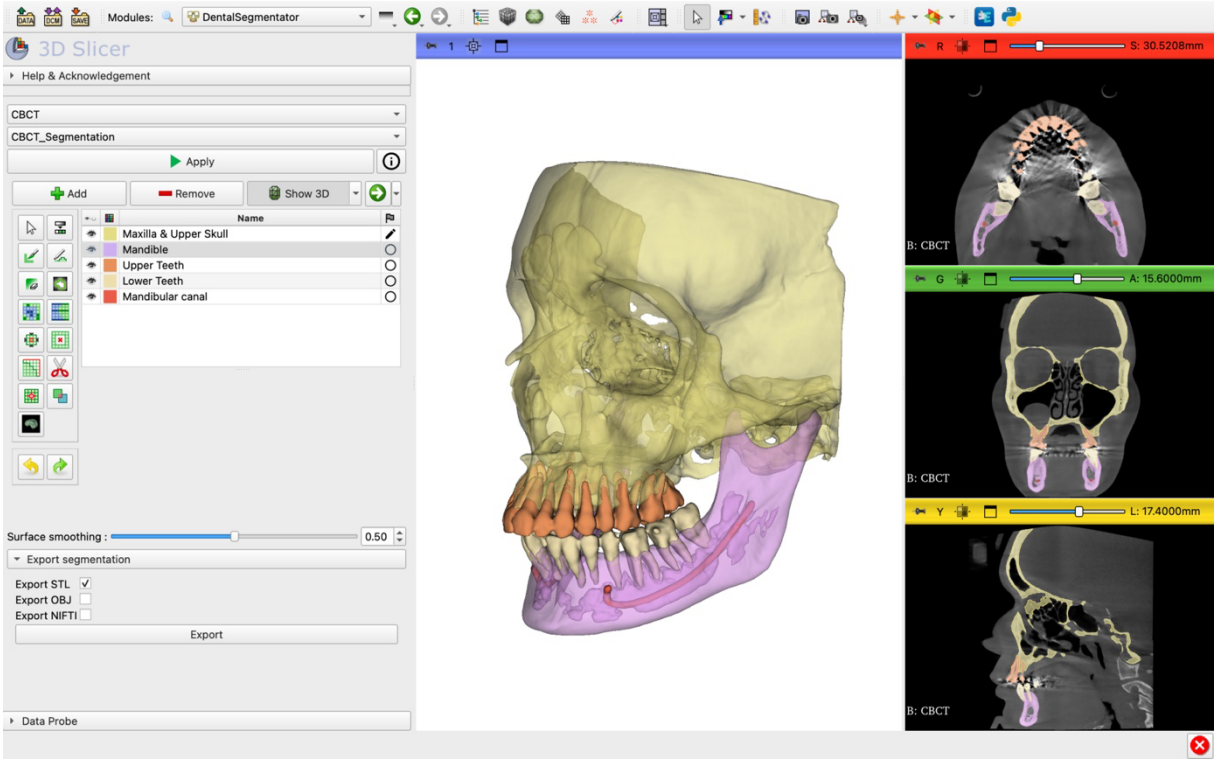


**Figure 4**: Screenshot of DentalSegmentator 3D Slicer extension.

## Discussion

We introduced DentalSegmentator, a deep learning-based tool for multiclass segmentation of DMF CT and CBCT images. This tool, based on the nnU-Net framework, was evaluated on a highly diversified test dataset of 79 CT and 177 CBCT scans from 7 institutions. Our comprehensive evaluation, comprising both volume-based and surface-based metrics, demonstrated that DentalSegmentator was able to provide fully automatic robust segmentation results for the 5 segmentation labels: maxilla/upper skull; mandible; upper teeth; lower teeth; and mandibular canal. The pre-trained model is publicly available, alongside a 3D Slicer extension with an easy-to-use graphical interface.

Due to the lack of publicly available DMF (CB)CT segmentation dataset, it is difficult to directly compare our results to previously published works. A recent systematic review highlighted that the DSC results obtained in the 23 selected studies ranged from 90 ± 3% to 97.9 ± 1.5% (Polizzi et al. 2023). Few studies in dentistry reported NSC results, as this metric has been proposed only recently by the biomedical community (Nikolov et al. 2021; Maier-Hein et al. 2023). This systematic review also pointed out the heterogeneity in the methods employed for dataset construction and model evaluation, a frequent problem in DL studies (Schwendicke et al. 2021). Some of the studies excluded patients with metal artifacts or significant skeletal deformities, while most of the models were evaluated on cross-validation datasets or on hold-out test datasets of less than 50 CBCT scans. The main risk of these approaches is that they may yield over-optimistic results which could be difficult to

reproduce in routine clinical care (i.e. poor generalizability of the model). The only study, to our knowledge, reporting results of DMF segmentation on a large-scale external test dataset (*n* = 407) had a mean DSC result of 93.8% (Cui et al. 2022), a result very close to ours. However, this model is not publicly available, which limits further evaluation and dissemination. As our dataset was randomly selected from clinical practice, most of our test images (67.6%) showed metallic artefacts. The images in our external dataset exhibited less metallic artefacts than our internal test dataset, which might explain the statistically significant difference between the results obtained on these two datasets.

Our results demonstrated the robustness and generalizability of our model for routine (CB)CT scans acquired in several use cases, ranging from orthognathic surgery planning to guided implant surgery, impacted teeth visualization or digital orthodontics. Methods in health data science are evolving at a very fast pace, with growing dataset sizes and constantly improving results (Schwendicke et al. 2022). However, we believe automatic segmentation for DMF (CB)CT scans is now mature enough to be proposed to the dental practitioners. This is the reason why we publicly share our pre-trained nnU-Net network and the DentalSegmentator extension for the 3D Slicer software. We hope that this effort will help disseminate the use of 3D models in dentistry and encourage the sharing of open datasets and improved methods. It has to be stated that while quantitative evaluation is necessary to assess the models' performance, such evaluation is not always clinically relevant (Dot et al. 2022). A clinical application such as personalized implant manufacturing will be particularly demanding in terms of segmentation precision, while computer-aided diagnosis or other digital dentistry tasks may not require such precision. The more demanding the clinical situation, the more human oversight must be incorporated into the workflow for validation and correction (Schwendicke et al. 2020).

This work has several limitations, the first one being its retrospective and relatively small-scale nature. A large prospective multi-center study would be needed to fully evaluate the generalizability of the tool. Our model was tested on 6 CBCT devices, a small number when compared to the 47 CBCT devices marketed by 20 companies that were available in 2012 (Nemtoi et al. 2013). Secondly, the construction of the reference test was a major difficulty due to the lack of a hard "gold standard" like dry skulls. We did our best to provide a solid ground-truth segmentation process with industry-certified segmentations (for the internal dataset) and a multistep approach involving experts (for the external dataset), but some biases remain possible. Finally, we did not seek to segment and label each tooth separately as it is proposed in several other methods (Ezhov et al. 2021; Shaheen et al. 2021; Cui et al. 2022; Ayidh Alqahtani et al. 2023; Nogueira-Reis et al. 2024), which could be a limitation for some applications.

To conclude, this work demonstrated the robustness of DentalSegmentator, a publicly available DL-based tool for segmentation of DMF CT and CBCT scans. The short-term perspectives of this work will depend on the adoption of the tool by the dental community. Thanks to the open source

nature of the nnU-Net framework, our model could easily be finetuned with more (CB)CT data to meet specific needs. Detection of specific pathologies like periapical lesions or bone lesions could be added to the method (Yeshua et al. 2023; Fu et al. 2024). In the medium term, it is likely that other DL methods will exceed the classical 3D U-Net used in the study. For example, foundation models like the recently proposed MedSAM could allow for universal image segmentation, improving the generalizability of the current methods (Ma et al. 2024).

**Author Contributions**

G. Dot contributed to the conception, design, data acquisition, analysis and interpretation, performed all statistical analyses, drafted the manuscript. A. Chaurasia, T. Schouman and L. Gajny contributed to the conception, design and data interpretation, and critically revised the manuscript. G. Dubois, C. Savoldelli, S. Haghighat, S. Azimian, A. Rahbar Taramsari, G. Sivaramakrishnan, J. Issa and A. Dubey contributed to the data analysis and critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

**Acknowledgments**

The authors would like to acknowledge the work of Thibault Pelletier and Roman Fenioux (Kitware SAS) for the development of DentalSegmentator 3D Slicer extension. Many thanks to the nnU-Net framework's team for sharing their great work.

**Declaration of Conflicting Interests**

G. Dubois and T. Schouman declared relationships with the following company: Materialise. The other authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

This study has received funding by the "Fédération Française d'Orthodontie" and the "Fondation des Gueules Cassées" (grant number 28–2020).

**Figure and table legends**

**Table 1**: Characteristics of the data in the train/validation, internal test and external test datasets.

**Figure 1**: DSC and NSD results on our internal (left) and external (right) test datasets. *SD: Standard Deviation.*

**Figure 2**: 3D surface models for 6 subjects representative of the diversity and the challenges arising from our test CBCT dataset. (**A**) Class III maxillo-mandibular deformity, before orthognathic surgery; (**B**) Edentulous jaws; (**C**) Left condylar hyperplasia; (**D**) Upper posterior dental gap; (**E**) Maxilla with impacted teeth; (**F**) Mandible with impacted third molar.

**Figure 3**: 3D surface models for 4 subjects, exhibiting some typical failures (red circles). (**A**) Under-segmentation of the mandibular inferior border and mandibular canal; (**B**) Under-segmentation of the anterior maxillary sinus walls; (**C**) Under-segmentation of the palate; (**D**) Discontinuity of the mandibular canal.

**Figure 4**: Screenshot of DentalSegmentator 3D Slicer extension.

**References**


Alkhayer A, Piffkó J, Lippold C, Segatto E. 2020. Accuracy of virtual planning in orthognathic surgery: a systematic review. Head Face Med. 16(1):34.

Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, Litjens G, Menze B, Ronneberger O, Summers RM, et al. 2022. The Medical Segmentation Decathlon. Nat Commun. 13(1):4128.

Assiri H, Dawasaz AA, Alahmari A, Asiri Z. 2020. Cone beam computed tomography (CBCT) in periodontal diseases: a Systematic review based on the efficacy model. BMC Oral Health. 20(1):191.

Ayidh Alqahtani K, Jacobs R, Smolders A, Van Gerven A, Willems H, Shujaat S, Shaheen E. 2023. Deep convolutional neural network-based automated segmentation and classification of teeth with orthodontic brackets on cone-beam computed-tomographic images: a validation study. Eur J Orthod. 45(2):169–174.

Borohovitz CL, Abraham Z, Redmond WR. 2021. The diagnostic advantage of a CBCT-derived segmented STL rendition of the teeth and jaws using an AI algorithm. J Clin Orthod JCO. 55(6):361–369.

Chogle S, Zuaitar M, Sarkis R, Saadoun M, Mecham A, Zhao Y. 2020. The Recommendation of Cone-beam Computed Tomography and Its Effect on Endodontic Diagnosis and Treatment Planning. J Endod. 46(2):162–168.

Cui Z, Fang Y, Mei L, Zhang B, Yu B, Liu J, Jiang C, Sun Y, Ma L, Huang J, et al. 2022. A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat Commun. 13(1):2096.

Dot G. 2024 Mar 18. DentalSegmentator nnU-Net pretrained model for CBCT image segmentation. [accessed 2024 Mar 18]. https://zenodo.org/doi/10.5281/zenodo.10829675.

Dot G, Schouman T, Dubois G, Rouch P, Gajny L. 2022. Fully automatic segmentation of craniomaxillofacial CT scans for computer-assisted orthognathic surgery planning using the nnU-Net framework. Eur Radiol. 32(6):3639–3648.

Ezhov M, Gusarev M, Golitsyna M, Yates JM, Kushnerev E, Tamimi D, Aksoy S, Shumilov E, Sanders A, Orhan K. 2021. Clinically applicable artificial intelligence system for dental diagnosis with CBCT. Sci Rep. 11(1):15006.

Fu WT, Zhu QK, Li N, Wang YQ, Deng SL, Chen HP, Shen J, Meng LY, Bian Z. 2024. Clinically Oriented CBCT Periapical Lesion Evaluation via 3D CNN Algorithm. J Dent Res. 103(1):5–12.

Gillot M, Baquero B, Le C, Deleat-Besson R, Bianchi J, Ruellas A, Gurgel M, Yatabe M, Turkestani NA, Najarian K, et al. 2022. Automatic multi-anatomical skull structure segmentation of cone-beam computed tomography scans using 3D UNETR. PLOS ONE. 17(10):e0275033.

Hu L, Rong R, Song W, Wu H, Jia S, He Z, Sa Y. 2023. Patient-specific 3D printed models for enhanced

learning of immediate implant procedures and provisionalization. Eur J Dent Educ Off J Assoc Dent Educ Eur. 27(3):765–772.
Ileșan RR, Beyer M, Kunz C, Thieringer FM. 2023. Comparison of Artificial Intelligence-Based Applications for Mandible Segmentation: From Established Platforms to In-House-Developed Software. Bioengineering. 10(5):604.
Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 18(2):203–211.
Kapila SD, Nervina JM. 2015. CBCT in orthodontics: assessment of treatment outcomes and indications for its use. Dentomaxillofacial Radiol. 44(1):20140282.
Kikinis R, Pieper SD, Vosburgh KG. 2014. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: Jolesz FA, editor. Intraoperative Imaging and Image-Guided Therapy. New York, NY: Springer. p. 277–289. [accessed 2023 Sep 26]. https://doi.org/10.1007/978-1-4614-7657-3_19.
Lahoud P, Jacobs R, Boisse P, EzEldeen M, Ducret M, Richert R. 2022. Precision medicine using patient-specific modelling: state of the art and perspectives in dental practice. Clin Oral Investig. 26(8):5117–5128.
Lejnieks M, Akota I, Jākobsone G, Neimane L, Radzins O, Uribe SE. 2023. Effect of 3D printed replicas on the duration of third molar autotransplantation surgery: A controlled clinical trial. Dent Traumatol. n/a(n/a). [accessed 2024 Feb 15]. https://onlinelibrary.wiley.com/doi/abs/10.1111/edt.12905.
Ma J, He Y, Li F, Han L, You C, Wang B. 2024. Segment anything in medical images. Nat Commun. 15(1):654.
Maier-Hein L, Reinke A, Godau P, Tizabi MD, Buettner F, Christodoulou E, Glocker B, Isensee F, Kleesiek J, Kozubek M, et al. 2023. Metrics reloaded: Recommendations for image analysis validation. [accessed 2024 Feb 15]. http://arxiv.org/abs/2206.01653.
Nemtoi A, Czink C, Haba D, Gahleitner A. 2013. Cone beam CT: a current overview of devices. Dentomaxillofacial Radiol. 42(8):20120443.
Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, De Fauw J, Patel Y, Meyer C, Askham H, Romera-Paredes B, et al. 2021. Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. J Med Internet Res. 23(7):e26151.
Nogueira-Reis F, Morgan N, Suryani IR, Tabchoury CPM, Jacobs R. 2024. Full virtual patient generated by artificial intelligence-driven integrated segmentation of craniomaxillofacial structures from CBCT images. J Dent. 141:104829.
Polizzi A, Quinzi V, Ronsivalle V, Venezia P, Santonocito S, Lo Giudice A, Leonardi R, Isola G. 2023. Tooth automatic segmentation from CBCT images: a systematic review. Clin Oral Investig. 27(7):3363–3378.
Schwendicke F, Marazita ML, Jakubovics NS, Krois J. 2022. Big Data and Complex Data Analytics: Breaking Peer Review? J Dent Res. 101(4):369–370.
Schwendicke F, Samek W, Krois J. 2020. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res. 99(7):769–774.
Schwendicke F, Singh T, Lee J-H, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J. 2021. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. J Dent. 107:103610.
SEDENTEXCT project. 2012. Cone Beam CT for dental and maxillofacial radiology (evidence based guidelines). European Commission.
Shaheen E, Leite A, Alqahtani KA, Smolders A, Van Gerven A, Willems H, Jacobs R. 2021. A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study. J Dent. 115:103865.
Wallner J, Schwaiger M, Hochegger K, Gsaxner C, Zemann W, Egger J. 2019. A review on multiplatform evaluations of semi-automatic open-source based image segmentation for cranio-maxillofacial surgery. Comput Methods Programs Biomed. 182:105102.
Wang H, Minnema J, Batenburg KJ, Forouzanfar T, Hu FJ, Wu G. 2021. Multiclass CBCT Image

Segmentation for Orthodontics with Deep Learning. J Dent Res. 100(9):943–949.
Wasserthal J, Breit H-C, Meyer MT, Pradella M, Hinck D, Sauter AW, Heye T, Boll DT, Cyriac J, Yang S, et al. 2023. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiol Artif Intell. 5(5):e230024.
Yeshua T, Ladyzhensky S, Abu-Nasser A, Abdalla-Aslan R, Boharon T, Itzhak-Pur A, Alexander A, Chaurasia A, Cohen A, Sosna J, et al. 2023. Deep learning for detection and 3D segmentation of maxillofacial bone lesions in cone beam CT. Eur Radiol. 33(11):7507–7518.