medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

DentalSegmentator: robust deep learning-based CBCT image segmentation

Original Research Report

Gauthier Dot[1,2,3]*, Akhilanand Chaurasia[4], Guillaume Dubois[3,5], 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[3,14], Laurent Gajny[3]

Abstract

Objectives: Segmentation of anatomical structures on dento-maxillo-facial (DMF) computed tomography (CT) or cone beam computed tomography (CBCT) scans is increasingly needed in digital dentistry. The main aim of this research was to propose and evaluate a novel open source tool called DentalSegmentator for fully automatic segmentation of five anatomic structures on DMF CT and CBCT scans: maxilla/upper skull, mandible, upper teeth, lower teeth, and the mandibular canal.

Methods: 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 in two 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.

Results: The mean overall results in the internal test dataset (n = 133) were a Dice similarity coefficient (DSC) of 92.2 ± 6.3% and a normalised surface distance (NSD) of 98.2 ± 2.2%. The mean overall results on the external test dataset (n = 123) were a DSC of 94.2 ± 7.4% and a NSD of 98.4 ± 3.6%.

Conclusions: The results obtained from this highly diverse dataset demonstrate that this tool can provide fully automatic and robust multiclass segmentation for DMF CT and CBCT scans. To encourage the clinical deployment of DentalSegmentator, the pre-trained nnU-Net model has been made publicly available along with an extension for the 3D Slicer software.

Clinical Significance: DentalSegmentator open source 3D Slicer extension provides a free, robust, and easy-to-use approach to obtaining patient-specific three-dimensional models from CT and CBCT scans. These models serve various purposes in a digital dentistry workflow, such as visualization, treatment planning, intervention, and follow-up.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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

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

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

[3] Institut de Biomecanique Humaine Georges Charpak, Arts et Metiers Institute of Technology, 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

1. Introduction

The clinical practice of dentistry has radically evolved in the last few years, partly because of 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, and facial scanners. This data has improved diagnosis, treatment planning, intervention, and patient follow-up in several areas of dentistry [1–4]. More specifically, patient-specific 3D models derived from CT or CBCT scans are already used for educational purposes, computer-assisted surgical planning or navigation, tooth auto-transplantation planning, and virtual treatment planning for orthodontic treatments [5–8]. It may not be long before these 3D models become a key part of precision medicine in dental practice based on finite element methods, providing an opportunity to individually assess anatomical and biomechanical characteristics and adapt treatment options accordingly [9].

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To obtain a patient-specific 3D model from a CT or CBCT scan, the anatomical structures of interest must be carefully delineated on the 3D image slices, a process called segmentation. The most frequent workflows require the segmentation of jaws (maxilla and mandible), teeth (upper and lower), and the mandibular canal. When performed manually, this segmentation process takes an expert two to five hours to complete [10,11]. The current gold standard for 3D DMF image segmentation is the semi-automatic method where automatic segmentations are refined manually by an expert [12]. In recent years, several research reports have shown that deep learning-based (DL) methods could fully automate this task with results on a par with those of the experts [10,11,13–16]. Several commercially available solutions already claim to use DL methods for CBCT segmentation [17–19].

Despite these promising results, a recent systematic review of automatic tooth segmentation approaches using CBCT scans revealed that most of the studies were at high risk of bias in data selection leading to potential overestimation of the accuracy of the methods [20]. Most published studies report results from cross-validation approaches or small-sized hold-out test dataset (less than 50 CBCT scans), which is probably insufficient for evaluating the robustness and generalizability of the methods in real-world clinical settings [21].

In 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 [22], the sharing of DL frameworks such as nnU-Net [23], and the sharing of pre-trained DL models for various segmentation tasks such as TotalSegmentator [24]. As far as the authors know, only two pre-trained DL models for DMF CT and CBCT segmentation are currently publicly shared. The first one is integrated in 3D Slicer software (version 5.6.1 and later - http://www.slicer.org/) [25] as an extension called Slicer Automated Dental Tools and provides segmentation of 4 anatomical structures: the mandible, the maxilla, the cranial base and the cervical vertebrae [15]. Unfortunately, this tool does not delineate the teeth from the jaws, which is a critical limitation in most digital dentistry and surgery workflows. The second one is based on the nnU-Net framework and provides segmentation of the tissues of interest, but is not supposed to be used with CBCT scan data as it has been developed and tested exclusively on CT scan data [26].

The main aim of the research presented in this paper was to propose and evaluate a novel tool for multiclass DMF CT and CBCT image segmentation called DentalSegmentator. The performance of the tool was thoroughly evaluated on two hold-out test datasets acquired from routine clinical practice in seven clinical centers.

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2. Materials and Methods

A DL framework was trained on an internal dataset for automatic segmentation of DMF CT and CBCT scans. The results obtained from this DL-based method (the index test) were compared with those obtained by semi-automatic segmentation (the reference test) on two hold-out test datasets. The outcome set included both volume-based and surface-based metrics. The Institutional Review Board "*Comité d'Ethique pour la Recherche en Imagerie Médicale*" (CERIM) gave ethical approval for this research (IRB No. CRM-2001-051b), and its reporting conforms to recently published recommendations on artificial intelligence in dental research [21].

2.1. Dataset

2.1.1. Patient selection

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

Data from the external test dataset was randomly sampled retrospectively from routine CBCT examinations in five private centers located in India. All of 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, and there was no minimum age. The only exclusion criterion was refusal to participate in the research. 123 subjects (123 CBCT scans) were included in the external test dataset.

2.1.2. Data characteristics

All the scans in the internal dataset 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 between

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

a training/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). 91% of the scans in the internal test set exhibited metal artefacts.

The FOV of the scans in the external test dataset ranged from full-head to being 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*0.16*0.16mm^3. The scans were acquired using five 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 metal artefacts.

Due to the anonymization process, the age of the subjects could not be retrieved. Descriptive characteristics of the dataset are shown in Table 1.

	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 * 0.16 * 0.16
Number of scans by CT Device			
GEHC Discovery CT 750 HD	353	75	
Other CT Device	21	4	
Number of scans by CBCT Device			
Carestream CS 9600 CBCT	96	54	
Carestream CS 9300 CBCT			29
Vatech Smart Plus CBCT			25
Dentium Rainbow CBCT			27
Planmeca Promax 3D CBCT			12
Sirona Orthophos XG 3D CBCT			30
Metal artifacts, no. (%)			
Orthodontic materials	364 (78.1)	107 (80.5)	3 (2.4)
Metal dental filling/crown	168 (35.7)	44 (33.1)	49 (39.8)
No metal artefact	57 (12.1)	12 (9.0)	71 (57.7)
Skeletal deformity, no. (%)			
Class I	61 (13.0)	13 (9.8)	
Class II[a]	246 (52.3)	69 (51.9)	
Class III[b]	163 (34.7)	51 (38.3)	

[a]Prognathic maxilla and/or retrognathic mandible evaluated on the 3D models. [b]Retrognathic maxilla and/or prognathic mandible evaluated on the 3D models.

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

2.1.3. Ground truth segmentation process (Reference Test)

The treatments carried out on patients in the internal dataset involved segmentation of the 3D scans prior to study. The ground truth segmentations were used for diagnosis, computer-aided surgical planning, and manufacture of personalized 3D-printed surgical guides and fixation implants. This was carried out by Materialise (Leuven, Belgium) according to a certified internal procedure which cannot

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

be fully described here for reasons of confidentiality. The two-step procedure started with semi-automatic patch-based segmentations using expert priors [27] which were manually refined by an initial operator from a trained team [Step 1]. The segmentations were then verified slice-by-slice for validation by a senior operator from a trained team [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.

The CBCT scans in the external dataset were segmented specifically for this study by using a semi-automatic three-step approach. First, the CBCT scans were segmented automatically using a publicly available deep-learning model trained on CT scans [26] [Step 1]. Second, the proposed segmentations were then corrected manually by five dentists familiar with 3D image visualization and trained for the task in 3D Slicer software (version 5.6.0) [Step 2]. Finally, the segmentations were verified slice-by-slice and corrected where necessary by a senior expert (a dentist with more than five years of experience in 3D image evaluation) in 3D Slicer software [Step 3]. This process resulted in five segmentation masks: maxilla/upper skull, mandible, upper teeth, lower teeth, and both mandibular canals.

2.2. Deep-Learning based segmentation (Index Test)

2.2.1. Training

The nnU-Net deep learning framework (version 2.2.1) was used as an out-of-the-box tool according to instructions given by its authors [23]. Our raw training/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, and the target spacing of the model was $0.31*0.43*0.43mm^3$. Training time was about 24 hours on our laboratory workstation (CPU AMD Ryzen 9 3900X 12-Core; 128Gb RAM; GPU Nvidia Titan RTX 24Gb).

2.2.2. Inference

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

2.3. Evaluation

Quantitative evaluation of the model performance was carried out on the internal and external test datasets by comparing ground truth segmentations (reference test) with DL-based segmentations

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(index test) for each of the five segmentation masks. The recommendations of the Metrics Reloaded project [28] were followed by using both volume-based Dice similarity coefficient (DSC) and surface-based normalized surface distance (NSD). The tolerance for NSD was set at 1 mm, consistent with recent international challenges in biomedical imaging [22]. Furthermore, 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 [29].

2.4. 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 (%). The results were nonparametric (Shapiro-Wilk normality test). The Wilcoxon-Mann-Whitney test was used to compare results from internal and external test datasets. The Kruskal-Wallis test was used to compare DSC and NSD results from different CT/CBCT devices; when significant, post-hoc Dunn's test was used to compare each group. p values <0.05 were considered to be statistically significant. All of the data was analysed using Python (v.3.7) and R Statistical Software (v4.2.2; R Core Team 2022).

3. Results

3.1. Quantitative evaluation

Inference time was approximately 1 to 2 minutes for one 3D scan performed on the laboratory workstation described above. The mean overall results in the internal test dataset (n = 133) were a DSC of 92.2 ± 6.3% and an NSD of 98.2 ± 2.2% (Table 2). The mean overall results in the external test dataset (n = 123) were a DSC of 94.2 ± 7.4% and an NSD of 98.4 ± 3.6% (Table 3). The distribution of results is shown in Figure 1.

The statistical analysis showed similar results for both DSC and NSD metrics. Overall, the results obtained on the external and internal test dataset were statistically different. There was no statistical difference in the internal test dataset when scans obtained using the various devices were compared. In the external test dataset, the results obtained for scans acquired using Carestream 9300 and Sirona Orthophos XG 3D showed statistically significant differences.

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

Table 2: DSC and NSD results from the internal test dataset (n = 133). *SD: Standard Deviation.*

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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

Table 3: DSC and NSD results from the external test dataset (*n* = 123). *SD: Standard Deviation*

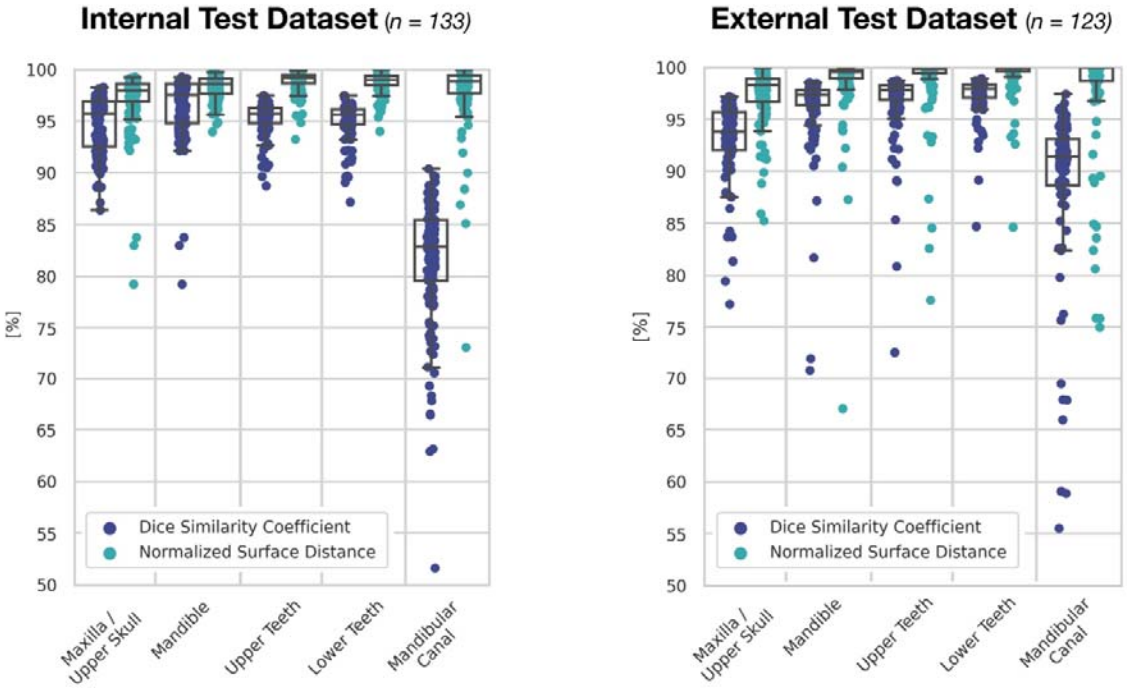

Figure 1: DSC and NSD results from the internal (left) and external (right) test datasets.

3.2. Three-Dimensional Visualization

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

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

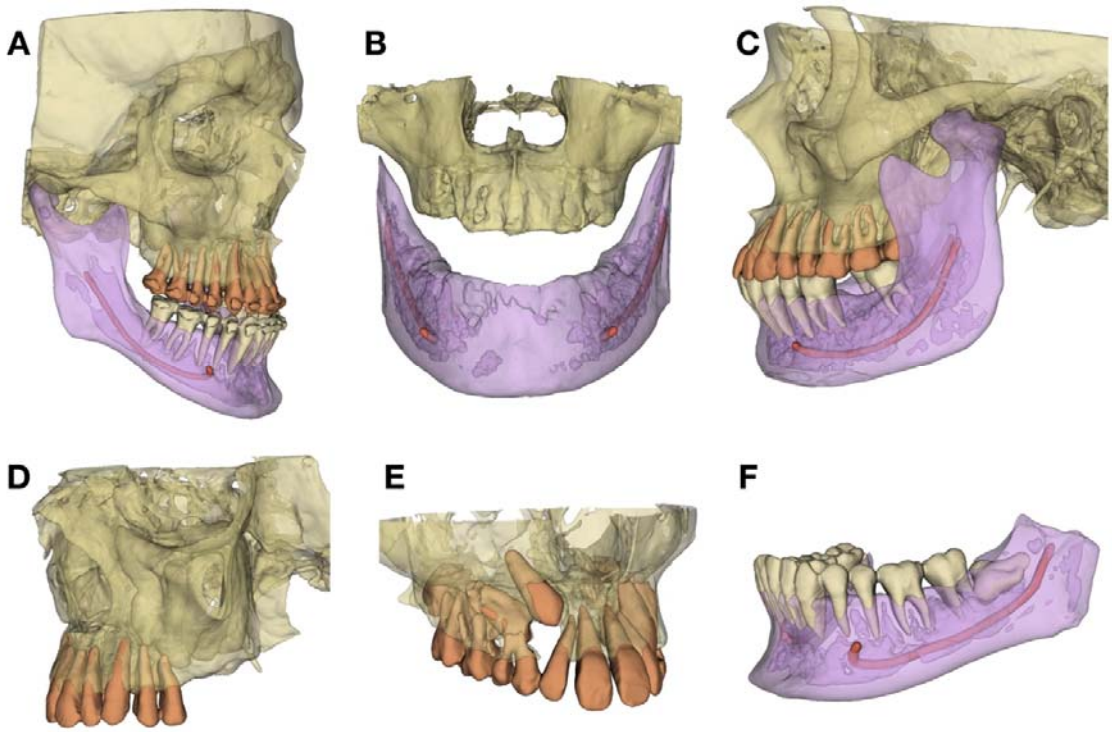

Figure 2: 3D surface models for six subjects representative of the diversity and the challenges arising from the test CBCT dataset. (**A**) Class III maxillo-mandibular deformity, before orthognathic surgery; (**B**) Edentulous jaws; (**C**) Left condylar hyperplasia; (**D**) Upper posterior edentulous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

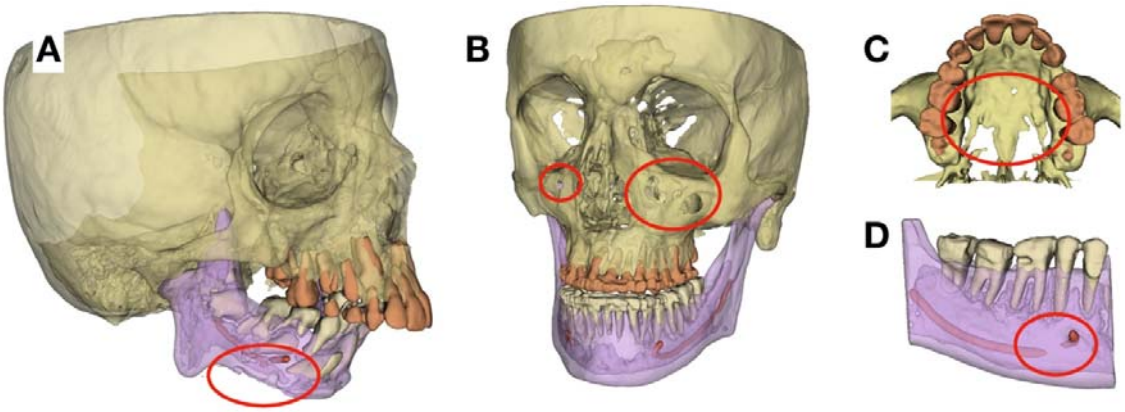

Figure 3: 3D surface models for four subjects, exhibiting some typical failures (red circles). (**A**) Under-segmentation of the mandibular inferior border and mandibular canal; (**B**) Under-segmentation of the anterior maxillary sinus walls; (**C**) Under-segmentation of the palate; (**D**) Discontinuity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

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3.3. DentalSegmentator model sharing and 3D Slicer extension

The pre-trained nnU-Net model is now publicly available [30]. This model can be used out-of-the-box via the nnU-Net version 2.2.1 command-line interface.

Implementation in a user-friendly interface to encourage clinicians to use the DL method is also proposed. DentalSegmentator is an open source extension for the 3D Slicer software (version 5.7.0 and later), which is a free, open source software for visualization, processing, and analysis of medical 3D images (http://www.slicer.org/) [25]. The extension, downloadable from the extension manager of 3D Slicer, offers an easy-to-use approach for DMF CT and CBCT scans automatic segmentation and 3D patient-specific model export (Figure 4). 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 the Github platform: https://github.com/gaudot/SlicerDentalSegmentator.

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

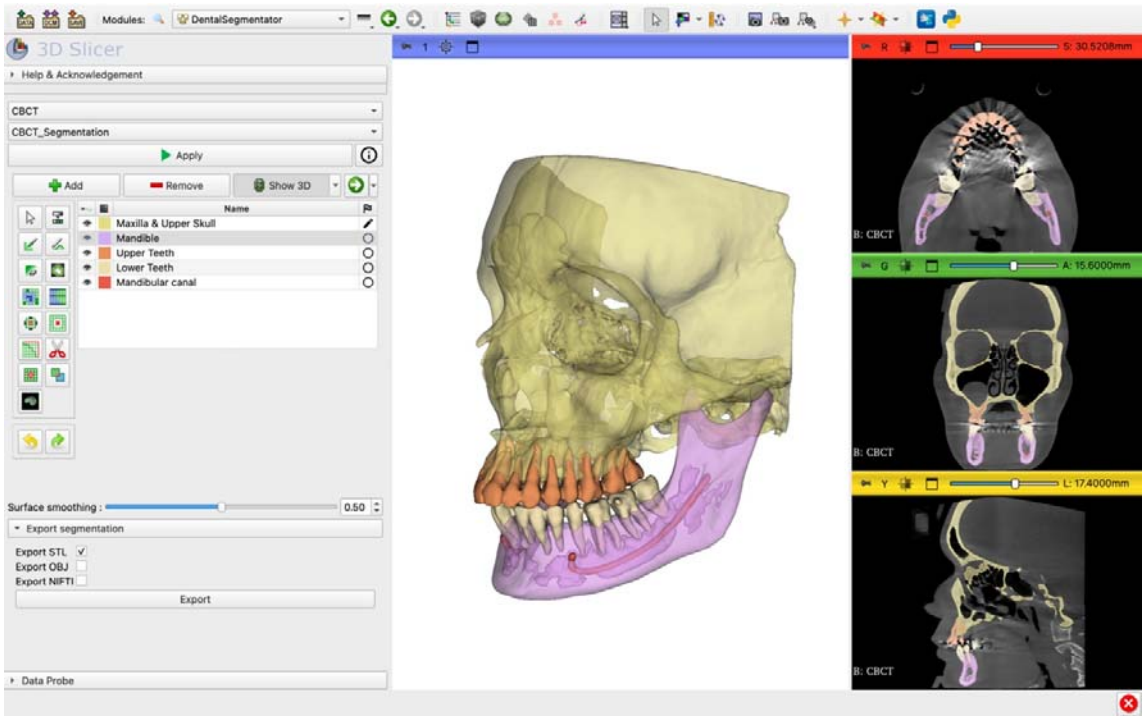

Figure 4: Screenshot of DentalSegmentator 3D Slicer extension.

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4. Discussion

This article introduces 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 diverse test dataset of 79 CT and 177 CBCT scans from seven institutions. The comprehensive evaluation was composed of both volume-based and surface-based metrics, and demonstrated that DentalSegmentator was able to provide fully automatic robust segmentation results for the five segmentation labels: maxilla/upper skull, mandible, upper teeth, lower teeth, and the mandibular canal. The pre-trained model is publicly available along with an open source 3D Slicer extension with an easy-to-use graphic interface.

Due to the lack of a publicly available DMF CT and CBCT segmentation dataset, it is difficult to compare the results obtained here directly with those previously published. A recent systematic review highlighted that the DSC results obtained in the 23 selected studies ranged from 90 ± 3% to 97.9 ± 1.5% [20]. Only a few studies in dentistry report NSC results because this metric has recently been proposed by the biomedical community [28,29]. This systematic review also pointed out that the heterogeneity in the methods employed for dataset construction and model evaluation is a frequent problem in DL studies [21]. Some of the studies excluded patients with metal artefacts or significant skeletal deformities, while most of the models were evaluated on cross-validation datasets or on hold-out test datasets of fewer 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). As far as the authors know, the only study reporting results of DMF segmentation on a large-scale external test dataset (n = 407) had a mean DSC result of 93.8% [11], a result very close to that found in the present study. However, this model is not publicly available, which limits further evaluation and dissemination. As the dataset used in the present study was randomly selected from clinical practice, most of the test images (67.6%) showed metal artefacts. The images in our external dataset exhibited fewer metal artefacts than our internal test dataset, which might explain the statistically significant difference between the results obtained on these two datasets. Despite these statistical results, the clinical impact of the variations observed between our internal and external test results is questionable. Part of the DSC errors found in the external dataset may be caused by the resolution of the external dataset as the median voxel size of this data was half the target spacing of the model presented in this paper.

In recent years, several deep learning architectures have been proposed for the segmentation of 3D biomedical data. The study presented in this paper used the open source nnU-Net framework, which was introduced in 2018 [23]. This framework is well supported, has been shown to provide state-of-the-art results on numerous datasets, and does not necessarily require

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

expert knowledge. New data recently published shows that in 2024 the framework was still providing highly competitive results compared to more modern architectures [30]. It is likely that slightly better results could be obtained by tailoring the deep learning architecture to the task, but that would require additional expertise.

The results demonstrated the robustness and generalizability of the model for the segmentation of routine CT and CBCT scans acquired in several cases of use such as orthognathic surgery planning, guided implant surgery, impacted teeth visualization, and digital orthodontics. Methods in health data science are evolving at a very fast pace, with growing dataset sizes and constantly improving results [32]. The results presented in this paper should improve if more training data is obtained from different CT and CBCT machines, or if post-processing steps are added (for example, to fill in some of the under-segmentations shown in Figure 3). However, automatic segmentation for DMF CT and CBCT scans is now mature enough for dental practitioners and researchers to use. This is why the pre-trained nnU-Net network and the DentalSegmentator extension for the 3D Slicer software have been publicly shared. It is hoped 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 said that while quantitative evaluation is necessary to assess the performance of the models, such evaluation is not always clinically relevant 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 of validation and correction must be incorporated into the workflow [33].

The results presented in this paper have several limitations. The first one is their retrospective and relatively small-scale nature. A large prospective multi-center study is needed to fully evaluate the generalizability of the tool. The model was tested on 6 CBCT devices, a small number compared to the 47 CBCT devices marketed by 20 companies that were available in 2012 [34]. Moreover, it was not possible to retrieve the age of the patients in the dataset, which calls into question the applicability of the solution for subjects with primary teeth. Secondly, the construction of the reference test was a major problem due to the lack of a hard “gold standard” like dry skulls. A solid segmentation process with industry-certified segmentations (for the internal dataset) and a multi-stage approach involving experts (for the external dataset) have been provided, but bias remains possible. Finally, each tooth was not segmented and labelled separately as proposed in several other methods [11,13,16,17,19], which could be a limitation in some applications.

The ethical implications of developing and using a tool such as DentalSegmentator have not been studied at length in the literature. A checklist for the evaluation of artificial intelligence applications in dentistry from an ethical perspective has recently been proposed [35]. When

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

evaluating the study presented in this paper from this point of view, several ethical pillars such as transparency, diversity, protection of privacy, equity, solidarity and governance were considered and addressed. More research needs to be carried out to address the pillars related to the clinical use of such a tool: wellness, respect of autonomous decision-making, accountability and responsibility, prudence, and sustainable development.

The perspectives of this study will depend on the adoption of the tool by the dental community. The main targets of this study in the short term are dental researchers and educators because 3D Slicer software is not approved for clinical use and the application distributed is intended for research use. Clinical use of this tool will require a few adaptations and further research to meet regulation such as the Artificial Intelligence Act recently adopted by the European Parliament [36]. In particular, several developments will be needed to mitigate the risk of automation bias and ensure that clinicians review and check the results before their clinical use. Thanks to the open source nature of the nnU-Net framework, the model could be easily fine-tuned with more CT and CBCT data to meet specific needs. Detection of specific pathologies like periapical lesions or bone lesions could be added to the method [37,38]. In the medium term, it is likely that other DL methods will exceed the classical 3D U-Net used in this study. For example, foundation models like the recently proposed MedSAM could allow for universal image segmentation, improving the generalizability of the current methods [39].

5. Conclusions

This study, based on a highly diverse dataset of more than 700 CT and CBCT scans, demonstrated the robustness of DentalSegmentator, a free open source tool for automatic segmentation of five key anatomical structures on DMF CT and CBCT scans.

References

[1] SEDENTEXCT project, Cone Beam CT for dental and maxillofacial radiology (evidence based guidelines), European Commission, 2012.

[2] S.D. Kapila, J.M. Nervina, CBCT in orthodontics: assessment of treatment outcomes and indications for its use, Dentomaxillofacial Radiol. 44 (2015) 20140282. https://doi.org/10.1259/dmfr.20140282.

[3] H. Assiri, A.A. Dawasaz, A. Alahmari, Z. Asiri, Cone beam computed tomography (CBCT) in periodontal diseases: a Systematic review based on the efficacy model, BMC Oral Health 20 (2020) 191. https://doi.org/10.1186/s12903-020-01106-6.

[4] S. Chogle, M. Zuaitar, R. Sarkis, M. Saadoun, A. Mecham, Y. Zhao, The Recommendation of Cone-beam Computed Tomography and Its Effect on Endodontic Diagnosis and Treatment Planning, J. Endod. 46 (2020) 162–168. https://doi.org/10.1016/j.joen.2019.10.034.

[5] A. Alkhayer, J. Piffkó, C. Lippold, E. Segatto, Accuracy of virtual planning in orthognathic

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

surgery: a systematic review, Head Face Med. 16 (2020) 34. https://doi.org/10.1186/s13005-020-00250-2.

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

[7] L. Hu, R. Rong, W. Song, H. Wu, S. Jia, Z. He, Y. Sa, 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 (2023) 765–772. https://doi.org/10.1111/eje.12928.

[8] M. Lejnieks, I. Akota, G. Jākobsone, L. Neimane, O. Radzins, S.E. Uribe, Effect of 3D printed replicas on the duration of third molar autotransplantation surgery: A controlled clinical trial, Dent. Traumatol. n/a (2023). https://doi.org/10.1111/edt.12905.

[9] P. Lahoud, R. Jacobs, P. Boisse, M. EzEldeen, M. Ducret, R. Richert, Precision medicine using patient-specific modelling: state of the art and perspectives in dental practice, Clin. Oral Investig. 26 (2022) 5117–5128. https://doi.org/10.1007/s00784-022-04572-0.

[10] 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.

[11] Z. Cui, Y. Fang, L. Mei, B. Zhang, B. Yu, J. Liu, C. Jiang, Y. Sun, L. Ma, J. Huang, Y. Liu, Y. Zhao, C. Lian, Z. Ding, M. Zhu, D. Shen, A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images, Nat. Commun. 13 (2022) 2096. https://doi.org/10.1038/s41467-022-29637-2.

[12] J. Wallner, M. Schwaiger, K. Hochegger, C. Gsaxner, W. Zemann, J. Egger, A review on multiplatform evaluations of semi-automatic open-source based image segmentation for cranio-maxillofacial surgery, Comput. Methods Programs Biomed. 182 (2019) 105102. https://doi.org/10.1016/j.cmpb.2019.105102.

[13] E. Shaheen, A. Leite, K.A. Alqahtani, A. Smolders, A. Van Gerven, H. Willems, R. Jacobs, A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study, J. Dent. 115 (2021) 103865. https://doi.org/10.1016/j.jdent.2021.103865.

[14] G. Dot, T. Schouman, G. Dubois, P. Rouch, L. Gajny, Fully automatic segmentation of craniomaxillofacial CT scans for computer-assisted orthognathic surgery planning using the nnU-Net framework, Eur. Radiol. 32 (2022) 3639–3648. https://doi.org/10.1007/s00330-021-08455-y.

[15] M. Gillot, B. Baquero, C. Le, R. Deleat-Besson, J. Bianchi, A. Ruellas, M. Gurgel, M. Yatabe, N.A. Turkestani, K. Najarian, R. Soroushmehr, S. Pieper, R. Kikinis, B. Paniagua, J. Gryak, M. Ioshida, C. Massaro, L. Gomes, H. Oh, K. Evangelista, C.M.C. Junior, D. Garib, F. Costa, E. Benavides, F. Soki, J.-C. Fillion-Robin, H. Joshi, L. Cevidanes, J.C. Prieto, Automatic multi-anatomical skull structure segmentation of cone-beam computed tomography scans using 3D UNETR, PLOS ONE 17 (2022) e0275033. https://doi.org/10.1371/journal.pone.0275033.

[16] K. Ayidh Alqahtani, R. Jacobs, A. Smolders, A. Van Gerven, H. Willems, S. Shujaat, E. Shaheen, 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 (2023) 169–174. https://doi.org/10.1093/ejo/cjac047.

[17] M. Ezhov, M. Gusarev, M. Golitsyna, J.M. Yates, E. Kushnerev, D. Tamimi, S. Aksoy, E. Shumilov, A. Sanders, K. Orhan, Clinically applicable artificial intelligence system for dental diagnosis with CBCT, Sci. Rep. 11 (2021) 15006. https://doi.org/10.1038/s41598-021-94093-9.

[18] R.R. Ileșan, M. Beyer, C. Kunz, F.M. Thieringer, Comparison of Artificial Intelligence-Based Applications for Mandible Segmentation: From Established Platforms to In-House-Developed Software, Bioengineering 10 (2023) 604. https://doi.org/10.3390/bioengineering10050604.

[19] F. Nogueira-Reis, N. Morgan, I.R. Suryani, C.P.M. Tabchoury, R. Jacobs, Full virtual patient generated by artificial intelligence-driven integrated segmentation of craniomaxillofacial structures from CBCT images, J. Dent. 141 (2024) 104829. https://doi.org/10.1016/j.jdent.2023.104829.

[20] A. Polizzi, V. Quinzi, V. Ronsivalle, P. Venezia, S. Santonocito, A. Lo Giudice, R. Leonardi, G.

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Isola, Tooth automatic segmentation from CBCT images: a systematic review, Clin. Oral Investig. 27 (2023) 3363–3378. https://doi.org/10.1007/s00784-023-05048-5.

[21] F. Schwendicke, T. Singh, J.-H. Lee, R. Gaudin, A. Chaurasia, T. Wiegand, S. Uribe, J. Krois, Artificial intelligence in dental research: Checklist for authors, reviewers, readers, J. Dent. 107 (2021) 103610. https://doi.org/10.1016/j.jdent.2021.103610.

[22] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, O. Ronneberger, R.M. Summers, B. van Ginneken, M. Bilello, P. Bilic, P.F. Christ, R.K.G. Do, M.J. Gollub, S.H. Heckers, H. Huisman, W.R. Jarnagin, M.K. McHugo, S. Napel, J.S.G. Pernicka, K. Rhode, C. Tobon-Gomez, E. Vorontsov, J.A. Meakin, S. Ourselin, M. Wiesenfarth, P. Arbeláez, B. Bae, S. Chen, L. Daza, J. Feng, B. He, F. Isensee, Y. Ji, F. Jia, I. Kim, K. Maier-Hein, D. Merhof, A. Pai, B. Park, M. Perslev, R. Rezaiifar, O. Rippel, I. Sarasua, W. Shen, J. Son, C. Wachinger, L. Wang, Y. Wang, Y. Xia, D. Xu, Z. Xu, Y. Zheng, A.L. Simpson, L. Maier-Hein, M.J. Cardoso, The Medical Segmentation Decathlon, Nat. Commun. 13 (2022) 4128. https://doi.org/10.1038/s41467-022-30695-9.

[23] F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat. Methods 18 (2021) 203–211. https://doi.org/10.1038/s41592-020-01008-z.

[24] J. Wasserthal, H.-C. Breit, M.T. Meyer, M. Pradella, D. Hinck, A.W. Sauter, T. Heye, D.T. Boll, J. Cyriac, S. Yang, M. Bach, M. Segeroth, TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images, Radiol. Artif. Intell. 5 (2023) e230024. https://doi.org/10.1148/ryai.230024.

[25] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J.V. Miller, S. Pieper, R. Kikinis, 3D Slicer as an image computing platform for the Quantitative Imaging Network, Magn. Reson. Imaging 30 (2012) 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001.

[26] G. Dot, T. Schouman, G. Dubois, P. Rouch, L. Gajny, Pretrained model for segmentation of craniomaxillofacial CT scans with nnU-Net, (2023). https://doi.org/10.5281/zenodo.7548785.

[27] P. Coupé, J.V. Manjón, V. Fonov, J. Pruessner, M. Robles, D.L. Collins, Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation, NeuroImage 54 (2011) 940–954. https://doi.org/10.1016/j.neuroimage.2010.09.018.

[28] L. Maier-Hein, A. Reinke, P. Godau, M.D. Tizabi, F. Buettner, E. Christodoulou, B. Glocker, F. Isensee, J. Kleesiek, M. Kozubek, M. Reyes, M.A. Riegler, M. Wiesenfarth, A.E. Kavur, C.H. Sudre, M. Baumgartner, M. Eisenmann, D. Heckmann-Nötzel, A.T. Rädsch, L. Acion, M. Antonelli, T. Arbel, S. Bakas, A. Benis, M. Blaschko, M.J. Cardoso, V. Cheplygina, B.A. Cimini, G.S. Collins, K. Farahani, L. Ferrer, A. Galdran, B. van Ginneken, R. Haase, D.A. Hashimoto, M.M. Hoffman, M. Huisman, P. Jannin, C.E. Kahn, D. Kainmueller, B. Kainz, A. Karargyris, A. Karthikesalingam, H. Kenngott, F. Kofler, A. Kopp-Schneider, A. Kreshuk, T. Kurc, B.A. Landman, G. Litjens, A. Madani, K. Maier-Hein, A.L. Martel, P. Mattson, E. Meijering, B. Menze, K.G.M. Moons, H. Müller, B. Nichyporuk, F. Nickel, J. Petersen, N. Rajpoot, N. Rieke, J. Saez-Rodriguez, C.I. Sánchez, S. Shetty, M. van Smeden, R.M. Summers, A.A. Taha, A. Tiulpin, S.A. Tsaftaris, B. Van Calster, G. Varoquaux, P.F. Jäger, Metrics reloaded: Recommendations for image analysis validation, (2023). https://doi.org/10.48550/arXiv.2206.01653.

[29] S. Nikolov, S. Blackwell, A. Zverovitch, R. Mendes, M. Livne, J. De Fauw, Y. Patel, C. Meyer, H. Askham, B. Romera-Paredes, C. Kelly, A. Karthikesalingam, C. Chu, D. Carnell, C. Boon, D. D'Souza, S.A. Moinuddin, B. Garie, Y. McQuinlan, S. Ireland, K. Hampton, K. Fuller, H. Montgomery, G. Rees, M. Suleyman, T. Back, C.O. Hughes, J.R. Ledsam, O. Ronneberger, Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study, J. Med. Internet Res. 23 (2021) e26151. https://doi.org/10.2196/26151.

[30] G. Dot, DentalSegmentator nnU-Net pretrained model for CBCT image segmentation, (2024). https://doi.org/10.5281/ZENODO.10829675.

[31] F. Isensee, T. Wald, C. Ulrich, M. Baumgartner, S. Roy, K. Maier-Hein, P.F. Jaeger, nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation, (2024). https://doi.org/10.48550/arXiv.2404.09556.

[32] F. Schwendicke, M.L. Marazita, N.S. Jakubovics, J. Krois, Big Data and Complex Data Analytics:

medRxiv preprint doi: https://doi.org/10.1101/2024.03.18.24304458; this version posted June 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Breaking Peer Review?, J. Dent. Res. 101 (2022) 369–370. https://doi.org/10.1177/00220345211070983.

[33] F. Schwendicke, W. Samek, J. Krois, Artificial Intelligence in Dentistry: Chances and Challenges, J. Dent. Res. 99 (2020) 769–774. https://doi.org/10.1177/0022034520915714.

[34] A. Nemtoi, C. Czink, D. Haba, A. Gahleitner, Cone beam CT: a current overview of devices, Dentomaxillofacial Radiol. 42 (2013) 20120443. https://doi.org/10.1259/dmfr.20120443.

[35] R. Rokhshad, M. Ducret, A. Chaurasia, T. Karteva, M. Radenkovic, J. Roganovic, M. Hamdan, H. Mohammad-Rahimi, J. Krois, P. Lahoud, F. Schwendicke, Ethical considerations on artificial intelligence in dentistry: A framework and checklist, J. Dent. 135 (2023) 104593. https://doi.org/10.1016/j.jdent.2023.104593.

[36] Texts adopted - Artificial Intelligence Act - Wednesday, 13 March 2024, (n.d.). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html (accessed May 21, 2024).

[37] T. Yeshua, S. Ladyzhensky, A. Abu-Nasser, R. Abdalla-Aslan, T. Boharon, A. Itzhak-Pur, A. Alexander, A. Chaurasia, A. Cohen, J. Sosna, I. Leichter, C. Nadler, Deep learning for detection and 3D segmentation of maxillofacial bone lesions in cone beam CT, Eur. Radiol. 33 (2023) 7507–7518. https://doi.org/10.1007/s00330-023-09726-6.

[38] W.T. Fu, Q.K. Zhu, N. Li, Y.Q. Wang, S.L. Deng, H.P. Chen, J. Shen, L.Y. Meng, Z. Bian, Clinically Oriented CBCT Periapical Lesion Evaluation via 3D CNN Algorithm, J. Dent. Res. 103 (2024) 5–12. https://doi.org/10.1177/00220345231201793.

[39] J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, Nat. Commun. 15 (2024) 654. https://doi.org/10.1038/s41467-024-44824-z.