Clinical Oral Investigations
https://doi.org/10.1007/s00784-020-03544-6

ORIGINAL ARTICLE

Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs

André Ferreira Leite[1,2] · Adriaan Van Gerven[3] · Holger Willems[3] · Thomas Beznik[3] · Pierre Lahoud[1] · Hugo Gaêta-Araujo[1,4] · Myrthel Vranckx[1] · Reinhilde Jacobs[1,5]

Received: 22 June 2020 / Accepted: 20 August 2020

Abstract

Objective To evaluate the performance of a new artificial intelligence (AI)-driven tool for tooth detection and segmentation on panoramic radiographs.

Materials and methods In total, 153 radiographs were collected. A dentomaxillofacial radiologist labeled and segmented each tooth, serving as the ground truth. Class-agnostic crops with one tooth resulted in 3576 training teeth. The AI-driven tool combined two deep convolutional neural networks with expert refinement. Accuracy of the system to detect and segment teeth was the primary outcome, time analysis secondary. The Kruskal-Wallis test was used to evaluate differences of performance metrics among teeth groups and different devices and chi-square test to verify associations among the amount of corrections, presence of false positive and false negative, and crown and root parts of teeth with potential AI misinterpretations.

Results The system achieved a sensitivity of 98.9% and a precision of 99.6% for tooth detection. For segmenting teeth, lower canines presented best results with the following values for intersection over union, precision, recall, F1-score, and Hausdorff distances: 95.3%, 96.9%, 98.3%, 97.5%, and 7.9, respectively. Although still above 90%, segmentation results for both upper and lower molars were somewhat lower. The method showed a clinically significant reduction of 67% of the time consumed for the manual.

Conclusions The AI tool yielded a highly accurate and fast performance for detecting and segmenting teeth, faster than the ground truth alone.

Clinical significance An innovative clinical AI-driven tool showed a faster and more accurate performance to detect and segment teeth on panoramic radiographs compared with manual segmentation.

Keywords Artificial intelligence · Machine learning · Panoramic radiography · Tooth · Classification

✉ André Ferreira Leite
andreleite@unb.br

[1] OMFS IMPATH Research Group, Department of Imaging and Pathology, Faculty of Medicine, University of Leuven and Oral & Maxillofacial Surgery, University Hospitals Leuven, KU Leuven, Kapucijnenvoer 33, 3000 Leuven, Belgium

[2] Department of Dentistry, Faculty of Health Sciences, Campus Universitario Darcy Ribeiro, University of Brasília, Brasília 70910-900, Brazil

[3] Relu, Innovatie-en incubatiecentrum KU Leuven, Leuven, Belgium

[4] Division of Oral Radiology, Department of Oral Diagnosis, Piracicaba Dental School, University of Campinas, Piracicaba, São Paulo, Brazil

[5] Department of Dental Medicine, Karolinska Institutet, Stockholm, Sweden

Introduction

Panoramic radiographs enable generating a 2D visualization of both dental arches. They are routinely used to present an overview of dental status and potential, surely if no prior radiograph is present [1, 2]. On the other hand, it presents limitations related to its two-dimensional nature, such as overlapping of anatomic structures and distortion [3]. Furthermore, human interpretation tends to be subjective because some dentists may not have enough specialized training or not enough time devoted to detailed diagnosis [4].

Artificial intelligence (AI)-based methods may be used in order to help dentists to interpret images. In this way, automated methods may enable faster identification and classification of data and eliminate errors associated with human fatigue. Deep learning algorithms have been investigated in

Published online: 26 August 2020

dentomaxillofacial radiology for the detection, classification, or diagnosis of diseases or anatomical structures, such as classification of teeth and mandibular morphology [5–7]; differentiation of jaw tumors [8]; and detection of root fractures [9], Sjögren's syndrome [10], maxillary sinusitis [11], calcified carotid atheroma's [12], caries [13], and periodontal diseases [14]. Although the results of previous AI research have been extremely promising, the studies are still preliminary [15].

When evaluating panoramic radiographs, one possible task susceptible to human failures is the identification of teeth and their exact shapes and boundaries. This process called tooth segmentation is of paramount importance for visual pattern recognition [16]. Therefore, the detection of anatomical structures can be considered as the first crucial step for radiographic recognition of pathologies [17]. Accurate detection and segmentation of each tooth on panoramic radiographs can help general dental practitioners in early diagnosis, making a better clinical decision [18]. Automated methods of dental age estimation might also improve forensic practice [19]. Moreover, tooth segmentation is helpful to visualize not only crowns but also roots' positions, when carrying out orthodontic treatment [20]. However, manual segmentation is time-consuming and can be tedious. Additionally, as it depends on the operator, it is prone to high inter-observer variability [21].

Artificial intelligence (AI)-driven methods have also been investigated to detect [22] and segment teeth on panoramic radiographs [21, 23–25]. Regarding segmentation studies, previous AI tools allowed to achieve results with accuracy metrics below 90%. The aim of this study was to evaluate the performance and clinically validate a novel AI-driven tool for detection and segmentation of teeth on panoramic radiographs. The accuracy of the tool was the primary outcome and time consumed for performing such a task was the secondary outcome. The hypothesis of the present study was that this novel AI-based method could increase the capacity of detecting teeth and achieve excellent and timely performance for tooth segmentation on such imaging modality. Such accurate automated tooth detection and segmentation tool might assist in treatment planning and contribute to future development of automated dental charting and radiographic reporting.

Materials and methods

Dataset

Panoramic radiographs were collected from a previous study for third molars evaluation [6] (M3BE database). The selected images were from patients older than 18 years old. With a unified dimension of 2880 × 1504 pixels, panoramic radiographs from the training set were acquired with the same device VistaPano S Ceph (Dürr Dental, Bientigheim-Bissingen, Germany). The test set was composed from panoramic radiographs acquired by two different devices: the aforementioned used for training and Promax 2D (Planmeca, Helsinki, Finland) with a dimension of 2931 × 1435 pixels. All radiographs from the training and validation test set were randomly selected and included extractions, fillings, root treatment, orthodontic brackets, retention wire, and patients with agenesis. This study was performed within the guidelines of the World Medical Association Helsinki Declaration for biomedical research involving human subjects and STROBE guidelines and was approved by the Institutional Review Board of our institution (approval number B322201525552). All image data were anonymized prior to analysis.

In total, the dataset consisted of 153 radiographs of adult patients (mean age 26 years old). The dataset was split into a training set of 70 images (images used to train the model). The validation set was composed by 18 panoramic images used to evaluate the model during training. Finally, 65 panoramic radiographs from two different devices composed the test set, from which the final performance of the model was calculated. The labeling and manual segmentation of the full panoramic images were performed by a dentomaxillofacial radiologist with 20 years of experience, serving as the ground truth. Data augmentation strategies were used during training to add more variability to the data set by applying random cropping, affine transformations (rotation, scaling, translation), contrast variations, and addition of noise. Afterwards, another dataset was constructed containing class-agnostic crops with one tooth that resulted in a dataset with segmentations of 3576 training teeth.

Segmentation AI models

Two different deep learning models were used. The first aimed at detecting and classifying the teeth. The second algorithm focused on fine-tuning the segmentation map of the detector. The detector performed a rough segmentation on panoramic radiographs and the classification of each tooth. The generated segmentation was used to crop out a region of interest around the tooth on the full resolution image. These full-resolution crops were then given to a class agnostic segmentation model to perform a detailed segmentation of each tooth. After detector resizing, images dimensions were 348 × 400 pixels. Images of 100 × 150 pixels were used for fine-tuning, chosen based on teeth shape.

Both networks were trained with the Adaptive Moment Estimation optimizer with early stopping based on the validation set. The detector has a deeplab-v3 architecture with a pre-trained resnet-101, based on a previous study [26]. This network was pre-trained on the Common Objects in Context and Pascal Visual Object Classes dataset resulting in powerful features which resulted in faster and better convergence of the training. This architecture was trained using the Pytorch framework for 100 epochs with a decaying learning rate where

the model with the best validation loss was chosen for testing. The detector should predict 16 class maps: 2 incisors, 1 canine, 2 premolars, 3 molars for both the mandible and the maxilla. The fine tuner was based on a previously developed fully convolutional neural network [27], also with the same pre-trained resnet-101 backbone. Figure 1 demonstrates the workflow of the system by combining two different neural networks for detecting and segmenting the teeth.

Segmentation software

The used tool to collect the data and perform the validation was an adapted version of the opensource Labelme project on GitHub [28]. In the new tool, several functionalities have been included, such as access to one tooth fine tuner to speed up the labeling process, cutting the segmentation contour at desired points, adjusting contrast and brightness, and performing histogram normalization, and the ability to load in the predictions of the segmentation models.

Accuracy metrics for tooth detection and segmentation

For tooth detection, the same metrics from a previous study were used [22]: sensitivity = true positive TP/TP + FN and precision = TP/TP+FP, where TP, FP, and FN represent true-positive, false-positive, and false-negative results, respectively. The TP represented correctly labeled teeth by the AI algorithm. The FP results consisted of misinterpreted teeth. On the other hand, FN results were existent teeth that were not annotated. In other words, the system's ability to detect the presence and absence of teeth was recorded.

A segmentation map was then created for each image, containing the classes of each pixel in the image. Before computing the metrics, the following numbers were first calculated:

- TP: the number of pixels that the method predicted to be part of the class that actually belonged to the class.
- FP: the number of pixels that the method predicted to be part of the class that did not belong to the class.
- True negative (TN): the number of pixels that the method predicted to not be part of the class that actually did not belong to the class.
- FN: the number of pixels that the method predicted to not be part of the class that actually belonged to the class.

After calculating TP, FP, TN, and FN, the following metrics were calculated:

- Intersection over union (IoU): TP/(TP + FN + FP). This ratio represents the area of overlap and the area of union. An IoU of 1 is a perfect segmentation.
- Precision: TP/(TP + FP)
- Recall: TP/(TP + FN)
- F1-score = 2*(precision*recall)/(precision + recall). The F1-score is used to balance precision and recall.

In addition, the Hausdorff distance was also calculated. It is equivalent to the maximum difference, measured in pixels between the ground truth and AI prediction per tooth. The accuracy measurements were calculated for the test set. For such purpose, this data set was split into three different datasets (65 radiographs). The first one compared the AI prediction without expert refinement to the ground truth (25 panoramic radiographs). The second one compared the fully AI with the AI combined with expert refinement (20 panoramic radiographs). These two data sets were composed of radiographs from the same device (VistaPano, Dürr Dental, Bientigheim-Bissingen, Germany). The third data set was composed of 20 panoramic images acquired by other devices (Promax 2D, Planmeca, Helsinki, Finland). In the second and third data set, after the AI-based algorithm predicted the segmentation, the radiologist performed all necessary corrections.

Time analysis

Time consumed and amount of corrections were recorded. Both AI-consumed time and tome for manual optimizations

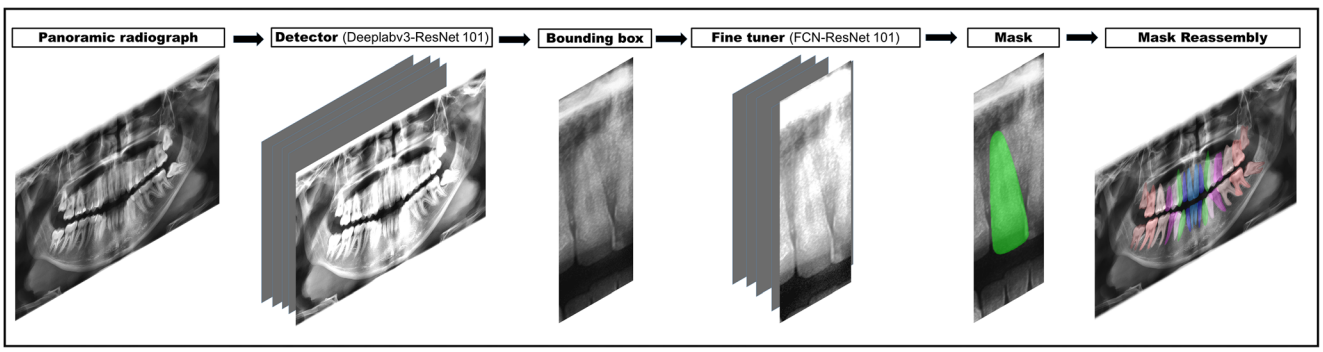


Fig. 1 Workflow of the system by combining two different neural networks for detecting and segmenting the teeth

to the segmentation were calculated. A three-grade classification of the amount of corrections was developed, based on the time consumed for doing this task: (a) no correction (time consumed only for judging the AI segmentation results without the need for manual corrections), (b) minor correction (up to 15 s, including the time for checking the image and for making minor adjustments on the tooth contour), (c) major correction (above 15 s, including the time for evaluating the image and for correcting the tooth contours). After the corrections, the quantitative metrics were calculated.

A visual evaluation was also conducted on the 20 panoramic radiographs from the second training data set. Such evaluation used illustrated colored mask boundaries on the AI segmentation results and the reference standard, similarly to a previous study [21]. The green color corresponds to the TP (i.e., the regions which overlap between the AI predicted results and the ground truth). Figure 2 a shows an example of the three-grade classification of the amount of corrections on the colored images. Besides the aforementioned classification, the radiologist also evaluated if the AI predictions were overestimated (FP, blue color) or underestimated (FN, red color) which is illustrated in Fig. 2b. The visual evaluation also aimed to record whether the corrections were made on the crown and/or the root portion of the teeth (Fig. 2c).

Statistical analysis

The differences of performance metrics among the different teeth groups and among different panoramic devices were assessed using the Kruskal-Wallis test. The chi-square test was used to verify associations among the amount of corrections, the presence of FP and FN, and the parts of the teeth (crown and root) with such AI misinterpretations. A p value less than 0.05 was considered statistically significant. All statistical analyses were performed with MedCalc Statistical Software version 19.2.1 (MedCalc Software Ltd, Oostende, Belgium).

Results

Accuracy metrics for tooth detection and segmentation

Considering only the present teeth in the testing dataset of 40 panoramic radiographs from two different devices, the system was capable of distinguishing 1179 teeth, with 15 failures of detections. There were four FP and 13 FN. Therefore, the system yielded 98.9% of sensitivity and 99.6% of precision. When the two tested devices were considered separately, the

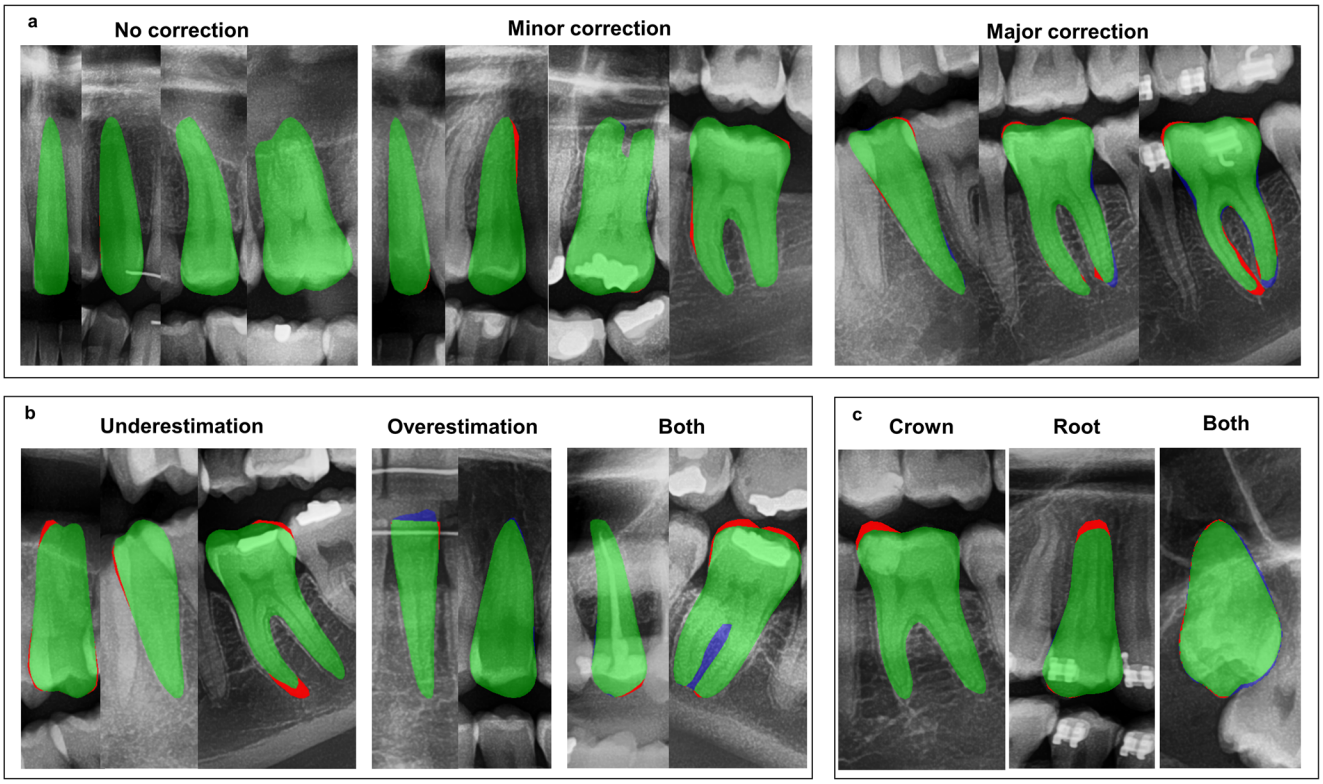


Fig. 2 Visual evaluation of the AI predictions. True positive, false negative, and false positive represented by the following colors: green, red, and blue, respectively. **a** Three-grade classification of the amount of corrections on the colored images. **b** Examples of AI underestimations (false negative) and overestimations (false positive). **c** Examples of underestimations and overestimations only at the crown, the root, and at both parts of the tooth

first panoramic device used in the training set yielded 99.6% of sensitivity and precision (with 2FP and 2FN). Values of sensitivity and precision of 98.1% and 99.6%, respectively, were achieved for the second panoramic device (11FN and 2FP). Furthermore, the module was also capable of identifying 96 missing teeth on images from both devices. Figure 3 a discloses some examples of the AI predictions compared with the ground truth and shows examples of detection and segmentation failures (Fig. 3b–e).

Table 1 shows the high performance of the new AI system for segmenting all teeth on panoramic radiographs compared with the ground truth. Considering the highest mean IoU values and the lowest mean Hausdorff distances, the lower canines presented the best performance. Although still presenting a very good performance (with mean IoU above 90%), segmentation results in both upper and lower molars were found to be somewhat lower. When evaluating images from the different panoramic devices separately, no differences were found concerning IoU median values, except for the lower incisors in which median IoU was slightly lower for images acquired by the device only used in the test set (Table 2).

There was an association between the amount of corrections and frequency of FP and FN (chi-square, $p < 0.001$). Concerning teeth needing minor corrections, the frequency of FN (74.4%) was higher than the frequency of FP (52.7%). Most of the teeth that needed major corrections presented both kinds of misinterpretation of the contours by the

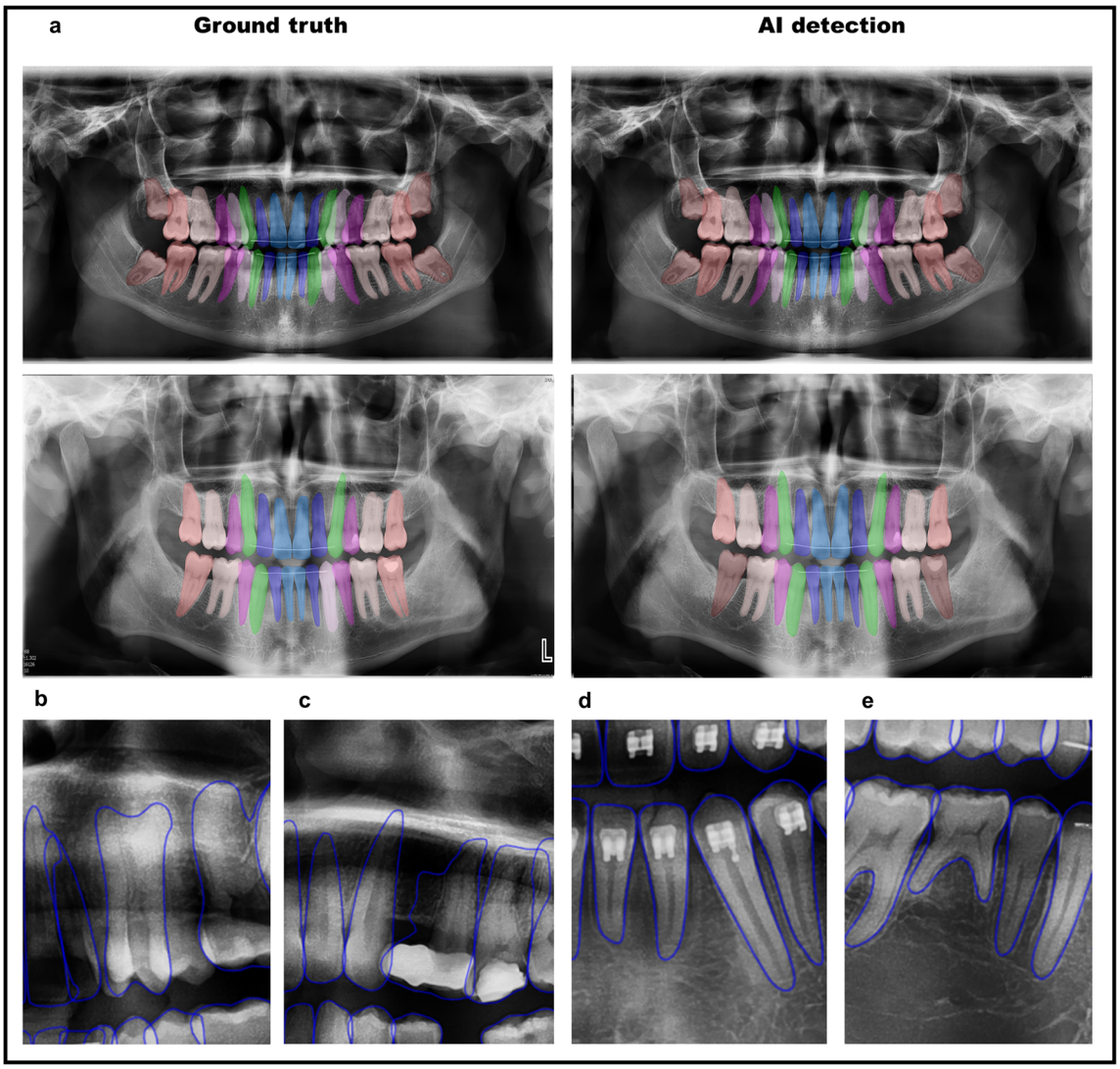


Fig. 3 **a** Comparison between AI prediction and the ground truth on two different panoramic radiographs. **b** AI was not able to detect teeth 23 and 25, probably because of the malposition of tooth 24. **d** Example of detection failure in which the AI network could detect a missing tooth, but it considered the present tooth as the 31 (but the expert considered as the 32). **c** The neural network was able to detect absence of tooth 24 and presence of tooth 25. However, it could be seen a bad segmentation of tooth 25 because the contour of the bridge that replaces tooth 24 has been incorporated to the contour of tooth 25 in the segmentation process. **e** AI was not able to distinguish a deciduous tooth in a case in which tooth 45 was absent with the prolonged permanence of the tooth 85

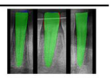
Table 1 Performance (accuracy metrics) of the new AI-driven system for segmenting all the teeth on panoramic radiographs

	IoU (%)	Precision (%)	Recall (%)	F1 (%)	Hausdorff (pixels)
Teeth 11/21	94.2	95.2	98.9	97.0	13.8
Teeth 12/22	94.4	96.6	97.7	97.1	10.7
Teeth 13/23	93.6	96.4	96.9	97.3	11.9
Teeth 14/24	94.2	96.2	97.7	96.8	13.2
Teeth 15/25	93.1	95.7	97.2	96.3	12.8
Teeth 16/26	92.4	95.1	97.0	95.9	19.5
Teeth 17/27	92.7	95.3	97.1	96.1	18.4
Teeth 18/28	93.7	96.1	97.5	96.7	14.9
Teeth 31/41	92.6	94.4	98.0	96.1	11.1
Teeth 32/42	93.7	95.5	98.1	96.7	10.2
Teeth 33/43	95.3	96.9	98.3	97.5	7.9
Teeth 34/44	94.7	97.1	97.5	97.2	9.8
Teeth 35/45	95.2	96.7	98.3	97.5	8.3
Teeth 36/46	92.8	96.1	96.4	96.2	14.1
Teeth 37/47	92.2	94.6	97.2	95.8	15.5
Teeth 38/48	92.6	95.0	97.4	96.0	16.0

AI network (FP and FN), although the frequency of FN was also slightly higher than the FP. The frequency of FP and FN was different among teeth groups (chi-square, $p < 0.001$). For all teeth groups, the frequency of FN was higher than the FP, except for incisors (Fig. 4a). The frequency of teeth that did not need correction was higher for anterior teeth. Conversely, multi-rooted teeth presented a higher need for refinement.

In general, the frequency of corrections was to some extent higher in the crowns (Fig. 4b). An association was also found between the frequency of misclassifications by the network and the portion of the teeth that needed corrections by the expert (chi-square, $p < 0.001$). The FP occurred somewhat more in the roots for canines and premolars but not for molars and incisors. From the 70 cases with such exclusive AI misclassification, 38

Table 2 Comparison of performance (medium IoU) of different teeth groups among panoramic radiographs acquired from different devices

Teeth group	Panoramic device 1 (Vistapano)	Panoramic device 2 (Promax)	P-value	Image example
Lower incisors	98.1	97.3	0.01	
Lower canines	99.3	98.1	0.13	
Lower premolars	98.0	98.2	0.50	
Lower molars	95.9	94.9	0.11	
Upper incisors	98.0	98.0	0.41	
Upper canines	99.0	99.1	0.76	
Upper premolars	98.0	98.0	0.65	
Upper molars	97.8	98.0	0.41	

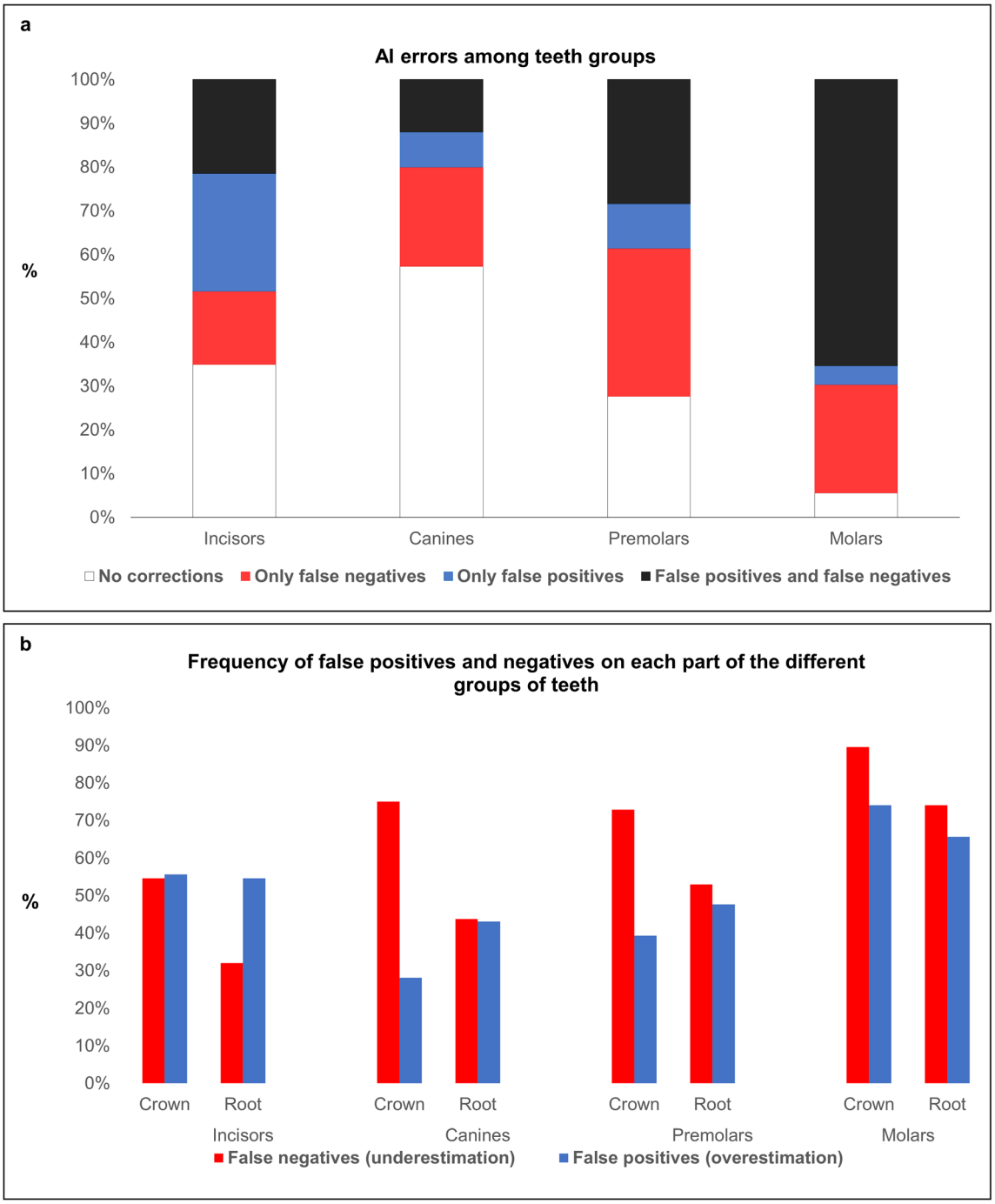


Fig. 4 **a** Graph showing the frequencies of different AI predictions (no corrections needed, false positive, and false negative) among teeth groups (incisors, canines, premolars, and molars). **b** Graph disclosing the frequencies of false positive and false negative on each part (crown and root) of the different teeth groups (incisors, canines, premolars, and molars)

(54.3%) were only in the roots, whereas 17 (24.3%) were only in the crowns. On the other hand, from the 145 cases with only FP, 91 (62.8%) occurred in the crowns.

Time analysis

An association was found between teeth groups and the amount of corrections (chi-square, $p < 0.001$). Correction times were significantly lower for the canines than for the other groups of teeth (Kruskal-Wallis, $p < 0.001$). For molars, correction time increased significantly ($p < 0.001$). Incisors and premolars presented equivalent correction times ($p > 0.05$). Table 3 shows the time consumed to perform tooth segmentation on each region comparing the combination of AI prediction and expert refinement among different devices, fully manual, and fully AI prediction. The median time to

perform manual segmentation of all teeth on panoramic radiographs was about 18 min for fully manual segmentation, with approximately 30 s for each anterior tooth, 35 s for premolars, and 40 s for molars. The median time for the AI prediction (fully automated segmentation) was 30 s. For the combination of AI prediction followed by an expert refinement, no significant differences were found among the two tested panoramic devices, with a median time of 6.3 s for performing the task. When comparing the AI prediction time to fully manual segmentation, the following time reductions could be achieved for all teeth group: 76% for incisors, 86% for canines, 75% for premolars, and 49% for molars.

Considering the cutoff value for the time correction of 15 s, it could be observed that only molars presented a higher frequency of major corrections need in both upper and lower jaws (Fig. 5 a and b). Conversely, in general, only minor corrections or no corrections were needed for incisors, canines, and premolars (Fig. 5b). Figure 5 c shows the comparison of time consumed by the expert to correct the AI prediction compared with the time consumed for the same expert to perform teeth segmentation manually. The time for doing manual segmentation was significantly higher than the time that the expert spent for refining AI predictions.

Fig. 5 **a** Graph comparing the time consumed for the expert to perform corrections on each group of teeth. The time was higher for molars and lower for canines and incisors. **b** Graph showing that only molars presented a higher frequency of major corrections for both upper and lower jaws. Most of the other teeth groups did not need corrections or needed only minor corrections by the expert. **c** The time consumed by the expert to correct the AI prediction was significantly lower than for the same expert to perform fully manual segmentation for all groups of teeth, especially for the anterior teeth

Discussion

The hypothesis of the present study was confirmed, as the newly proposed AI tool presented a highly accurate and fast performance for detecting and segmenting teeth. The system yielded an almost perfect sensitivity and precision for tooth detection and was capable of accurately detecting teeth on images acquired from different devices. These robust results bring new clinical perspectives for further implementation of pathology detection. Regarding teeth segmentation, a mean IoU above 90% was achieved for all teeth groups. Moreover, this study was the first to demonstrate that the time needed to perform fully manual segmentation of all teeth of a panoramic radiograph may be reduced by 67% when using a segmentation method based on deep learning algorithms. Therefore, this method may bring more efficiency to the daily workflow of the dentists.

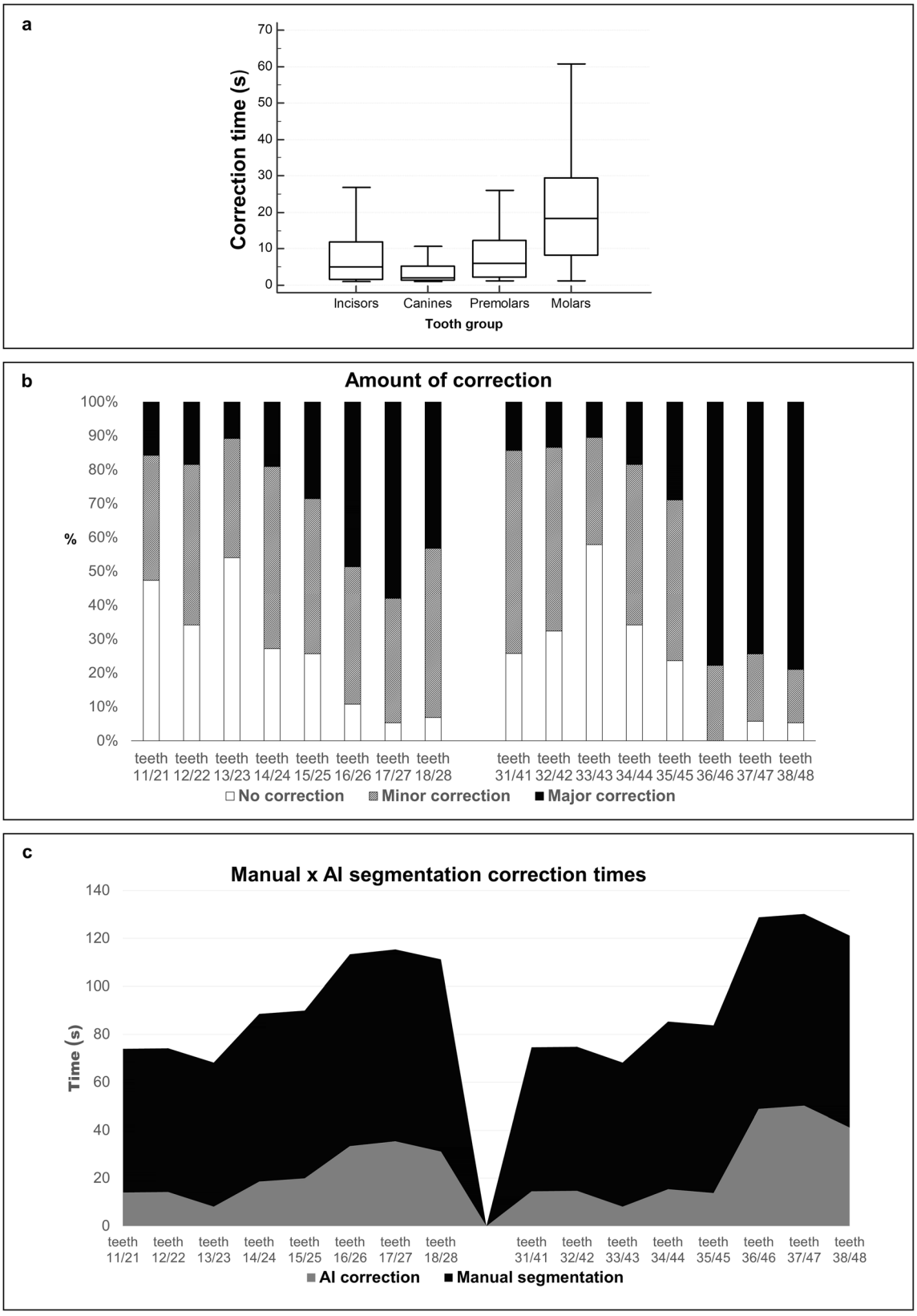
Automation of tooth detection and segmentation may be considered the first and most challenging step in the development of AI systems capable of interpreting images and differentiating pathologies from anatomical structures. Therefore, this first step should be as accurate as possible. Our system achieved excellent performance for tooth detection, similarly to a recently developed system [22]. However, the previous module aimed only at detecting and numbering the teeth, while our system was developed to detect and segment teeth.

Regarding tooth segmentation, the higher performance was achieved for canines, followed by incisors, premolars, and molars. To the best of our knowledge, only four previous studies have tested the use of AI tools for automatic detection and segmentation of all teeth on panoramic radiographs [21, 23–25]. The lower values achieved by the three aforementioned studies [10, 12, 13] were probably related to smaller data set sizes or different applied methodologies. Our results are more in line with the two most recent studies that also observed a higher degree of difficulty for segmenting multi-

Table 3 Median time (seconds) for performing tooth segmentation comparing artificial intelligence (AI) prediction with expert refinement among different devices and fully manual and fully AI

	First panoramic device (VistaPano)	Second panoramic device (Promax 2D)			
Teeth group	AI prediction + expert refinement	AI prediction + expert refinement	*p* value	Fully manual	Fully AI
Lower incisors	4.9 s	6.4 s	0.82	30 s	N/A
Lower canines	2.9 s	3.0 s	0.42	30 s	N/A
Lower premolars	7.8 s	8.6 s	0.30	35 s	N/A
Lower molars	26.2 s	26.1	0.20	40 s	N/A
Upper incisors	5.5 s	6.4 s	0.24	30 s	N/A
Upper canines	4.1 s	4.0 s	0.37	30 s	N/A
Upper premolars	6.7 s	10.4 s	0.14	35 s	N/A
Upper molars	15.9	14.9	0.81	40s	N/A
Total time	367 s	387 s		1080s	30 s

N/A, not applicable



rooted teeth than for single-rooted teeth [21, 25]. One study showed the following mean IoU values for the upper and lower teeth: 0.900 for incisors, 0.889 for the canines, 0.873 for the premolars, and 0.859 for the molars [10]. Another deep learning-based system showed a diagnostic accuracy of 86.9% for the determination of the root morphology of the

mandibular first molar [7]. Accordingly, our results were superior or similar to those previous studies for all the accuracy metrics. By combining two different deep learning algorithms with expert refinement, we could greatly improve convergence speed and final accuracy for the segmentation task.

Tooth segmentation is more challenging than bone structure segmentation for several reasons, such as the number of teeth per jaw, the proximity of adjacent tooth structures, the difference in density within a tooth, and tooth development [29]. This process is even more challenging on a panoramic radiograph due to its inherent limitations [3].

Regarding tooth segmentation in our study, the frequency of FN was higher than FP, except for the incisors. This might be explained by some radiopaque images that eventually appear in such anterior region, including the superimposition of the hard palate line and the anterior nasal spine on the apical region of the upper incisor, the ghost image of the cervical spine, and the bite block used to position the patients. These images may confuse the networks. As a result, boundaries of the incisors' root and crowns have been slightly overestimated in some images. Molars presented a higher number of FN and FP. This was an expected result, considering the complexity of their anatomy on both crowns and roots. Nonetheless, even for this tooth group, the performance of the system was quite impressive.

The high performance achieved by the novel AI tool could have been influenced by the good image quality and composition of the data sets. For a 2D image, the combination of an AI-based method with expert refinement should improve the performance. Analyses were performed on high-resolution panoramic radiographs acquired on two devices with no significant differences regarding performance and time analysis. Furthermore, most of the training and validation test data was composed of radiographs from young adults, as the inclusion criteria comprised radiographs of patients older than 18 years from an epidemiological study on the management of third molars [6]. Despite the possibility of selection bias, the choice of selecting panoramic radiographs from young patients was to ensure that the system was trained for all groups of teeth. Nevertheless, the selection criteria did not exclude radiographs of low and high density, with some positioning imperfections, artifacts, presence of orthodontic appliances, and from older patients with missing teeth. The system also performed well in such cases. For AI-driven imaging approaches, the influence of factors such as image quality, artifacts, presence of implants and bridges, and variations of the teeth in between patients and patient's age should be further investigated.

Deep convolutional neural networks are increasingly applied for medical image diagnostics [30]. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. In this way, images can be used as an input for the neural networks in order to achieve several different outputs [15]. As such, deep learning methods have already shown promising results for detecting caries [13], root fractures [9], periodontal diseases [14], for differentiating cysts and jaw tumors [8], for skeletal classification on lateral cephalograms [31], and even for improving oral cancer outcomes [32]. Regarding teeth and bone segmentation, deep learning is an encouraging approach to segment anatomical structures and later on in clinical decision making [5]. Advances in the methods may improve virtual surgery planning, optimize orthodontic planning and follow-up, forensic identification, and help to differentiate normal anatomical structures from pathologies. Therefore, the combination of AI tools with expert's analyses may revolutionize oral healthcare, allowing to have real precise and personalized dentistry. Dentist's workflow will become more efficient with automated suggestions for complex cases, better treatment planning, and prediction of diseases and outcomes [14].

In conclusion, our study presented and validated a new AI-driven tool for fast and accurate tooth detection and segmentation on panoramic radiographs. Other head and neck structures should be also further investigated. Future clinical perspectives of this development include AI-based tools for automated dental charting and radiographic reporting, as well as AI-driven clinical decision support. The current developments form the basis of further developments of AI-driven tools to accurately and automatically diagnose various dental and bone diseases, saving time and minimizing human errors. The use of more robust and explainable systems is an important step for a more precise and personalized dental practice.

Funding information The authors gratefully acknowledge financial support from Fundação de Apoio à Pesquisa do Distrito Federal– FAP-DF (protocol #23106.013588/2019-05).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Sklavos A, Beteramia D, Delpachitra SN, Kumar R (2019) The panoramic dental radiograph for emergency physicians. Emerg Med J 36(9):565–571. https://doi.org/10.1136/emermed-2018-208332
2. Yeung AWK, Mozos I (2020) The innovative and sustainable use of dental panoramic radiographs for the detection of osteoporosis. Int J Environ Res Public Health 17(7). https://doi.org/10.3390/ijerph17072449

3. Jacobs R, Quirynen M (2014) Dental cone beam computed tomography: justification for use in planning oral implant placement. Periodontol 2000 66(1):203–213. https://doi.org/10.1111/prd.12051
4. Lin PL, Huang PY, Huang PW, Hsu HC, Chen CC (2014) Teeth segmentation of dental periapical radiographs based on local singularity analysis. Comput Methods Prog Biomed 113(2):433–445. https://doi.org/10.1016/j.cmpb.2013.10.015
5. Vinayahalingam S, Xi T, Bergé S, Maal T, de Jong G (2019) Automated detection of third molars and mandibular nerve by deep learning. Sci Rep 9(1):9007. https://doi.org/10.1038/s41598-019-45487-3
6. Vranckx M, Ockerman A, Coucke W, Claerhout E, Grommen B, Miclotte A, van Vlierberghe M, Politis C, Jacobs R (2019) Radiographic prediction of mandibular third molar eruption and mandibular canal involvement based on angulation. Orthod Craniofacial Res 22(2):118–123. https://doi.org/10.1111/ocr.12297
7. Hiraiwa T, Ariji Y, Fukuda M, Kise Y, Nakata K, Katsumata A, Fujita H, Ariji E (2019) A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. Dentomaxillofacial Radiol 48(3):20180218. https://doi.org/10.1259/dmfr.20180218
8. Ariji Y, Yanashita Y, Kutsuna S, Muramatsu C, Fukuda M, Kise Y, Nozawa M, Kuwada C, Fujita H, Katsumata A, Ariji E (2019) Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. Oral Surg Oral Med Oral Pathol Oral Radiol 128(4):424–430. https://doi.org/10.1016/j.oooo.2019.05.014
9. Fukuda M, Inamoto K, Shibata N, Ariji Y, Yanashita Y, Kutsuna S, Nakata K, Katsumata A, Fujita H, Ariji E (2019) Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. Oral Radiol 2019:1617–1620. https://doi.org/10.1007/s11282-019-00409-x
10. Kise Y, Ikeda H, Fujii T, Fukuda M, Ariji Y, Fujita H, Katsumata A, Ariji E (2019) Preliminary study on the application of deep learning system to diagnosis of Sjögren's syndrome on CT images. Dentomaxillofacial Radiol 48(6):20190019. https://doi.org/10.1259/dmfr.20190019
11. Murata M, Ariji Y, Ohashi Y, Kawai T, Fukuda M, Funakoshi T, Kise Y, Nozawa M, Katsumata A, Fujita H, Ariji E (2019) Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. Oral Radiol 35(3):301–307. https://doi.org/10.1007/s11282-018-0363-7
12. Kats L, Vered M, Zlotogorski-Hurvitz A, Harpaz I (2019) Atherosclerotic carotid plaque on panoramic radiographs: neural network detection. Int J Comput Dent 22(2):163–169
13. Moutselos K, Berdouses E, Oulis C, Maglogiannis I (2019) Recognizing occlusal caries in dental intraoral images using deep learning. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2019, pp 1617–1620. https://doi.org/10.1109/EMBC.2019.8856553
14. Lee JH, Kim DH, Jeong SN, Choi SH (2018) Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci 48(2):114. https://doi.org/10.5051/jpis.2018.48.2.114
15. Leite AF, de Faria Vasconcelos K, Willems H, Jacobs R (2020) Radiomics and machine learning in oral healthcare. Proteomics Clin Appl 14(3):e1900040. https://doi.org/10.1002/prca.201900040
16. Barboza EB, Marana AN, Oliveira DT (2012) Semiautomatic dental recognition using a graph-based segmentation algorithm and teeth shapes features. In: Proceedings - 2012 5th IAPR International Conference on Biometrics, ICB 2012. https://doi.org/10.1109/ICB.2012.6199831.
17. Baksi BG, Alpöz E, Soğur E, Mert A (2010) Perception of anatomical structures in digitally filtered and conventional panoramic radiographs: A clinical evaluation. Dentomaxillofacial Radiol 39(7):424–430. https://doi.org/10.1259/dmfr/30570374
18. Hasan MM, Ismail W, Hassan R, Yoshitaka A (2016) Automatic segmentation of jaw from panoramic dental X-ray images using GVF snakes. In: World Automation Congress Proceedings, vol 1, pp 1–6. https://doi.org/10.1109/WAC.2016.7583022
19. Banar N, Bertels J, Laurent F, Boedi RM, de Tobel J, Thevissen P, Vandermeulen D (2020) Towards fully automated third molar development staging in panoramic radiographs. Int J Legal Med 134(5):1831–1841. https://doi.org/10.1007/s00414-020-02283-3
20. Galibourg A, Dumoncel J, Telmon N, Calvet A, Michetti J, Maret D (2018) Assessment of automatic segmentation of teeth using a watershed-based method. Dentomaxillofacial Radiol 47(1):20170220. https://doi.org/10.1259/dmfr.20170220
21. Lee JH, Han SS, Kim YH, Lee C, Kim I (2020) Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. Oral Surg Oral Med Oral Pathol Oral Radiol 129(6):635–642. https://doi.org/10.1016/j.oooo.2019.11.007
22. Tuzoff DV, Tuzova LN, Bornstein MM, Krasnov AS, Kharchenko MA, Nikolenko SI, Sveshnikov MM, Bednenko GB (2019) Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiol 48(4):20180051. https://doi.org/10.1259/dmfr.20180051
23. Silva G, Oliveira L, Pithon M (2018) Automatic segmenting teeth in X-ray images: trends, a novel data set, benchmarking and future perspectives. Expert Syst Appl 107:15–31. https://doi.org/10.1016/j.eswa.2018.04.001
24. Wirtz A, Mirashi SG, Wesarg S (2018) Automatic teeth segmentation in panoramic X-ray images using a coupled shape model in combination with a neural network. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 16-20, pp 712–179. https://doi.org/10.1007/978-3-030-00937-3_81
25. Jader G, Fontineli J, Ruiz M et al (2019) Deep instance segmentation of teeth in panoramic X-ray images. In: Proceedings - 31st Conference on Graphics, Patterns and Images, SIBGRAPI 2018. https://doi.org/10.1109/SIBGRAPI.2018.00058
26. Chen L-C, Zhu Y, Papandreou G, et al (2018) Rethinking atrous convolution for semantic image segmentation arXiv Prepr arXiv170605587. https://doi.org/10.1159/000018039
27. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015. https://doi.org/10.1109/CVPR.2015.7298965
28. Wada K labelme: Image polygonal annotation with Python. https://github.com/wkentaro/labelme. Accessed 1 Nov 2019
29. Shaheen E, Khalil W, Ezeldeen M, van de Casteele E, Sun Y, Politis C, Jacobs R (2017) Accuracy of segmentation of tooth structures using 3 different CBCT machines. Oral Surg Oral Med Oral Pathol Oral Radiol 23(1):123–128. https://doi.org/10.1016/j.oooo.2016.09.005
30. Schwendicke F, Golla T, Dreher M, Krois J (2019) Convolutional neural networks for dental image diagnostics: a scoping review. J Dent 91:103226. https://doi.org/10.1016/j.jdent.2019.103226
31. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J (2020) Automated skeletal classification with lateral cephalometry based on artificial intelligence. J Dent Res 99(3):249–256. https://doi.org/10.1177/0022034520901715
32. Ilhan B, Lin K, Guneri P, Wilder-Smith P (2020) Improving oral cancer outcomes with imaging and artificial intelligence. J Dent Res 99(3):241–248. https://doi.org/10.1177/0022034520902128