**Supplementary Materials**

**Clinically Applicable System For 3D Teeth Segmentation in Intraoral Scans using Deep Learning**

Jin Hao1, 2, Wen Liao1, Yueling Zhang1, Peilin Li1, Jianru Yi1, Jerry Peng3, Zeu Zhao3, Zhang Chen3, Wenxuan Shi3, Tingyu Chen3, Bowen Zhou4, Yang Feng4, Bing Fang5, Haoji Hu6, Howard H. Yang7, Erping Li6, 7, Zuozhu Liu6, 7\*, Zhihe Zhao1\*

1 State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & West China Hospital of Stomatology, Sichuan University, Chengdu, China

2 Harvard School of Dental Medicine, Harvard University, Boston, MA, USA

3 DeepAlign Tech Inc., Ningbo, China

4 Angelalign Research Institute, Angel Align Inc., Shanghai, China

5 Ninth People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai Research Institute of Stomatology, National Clinical Research Center of Stomatology, Shanghai, China

6 College of Information Science and Electrical Engineering, Zhejiang University, Hangzhou, China

7 Zhejiang University-University of Illinois at Urbana-Champaign Institute, Zhejiang University, Haining, China

\*Correspondence: zuozhuliu@intl.zju.edu.cn (Z.L.) &zhzhao@scu.edu.cn (Z.Z.)

**Supplementary Methods**

**Neural Network Architecture.**

The detailed settings of the architecture are as follows. The transformation net is composed of: Conv2D[64]-> Conv2D[128]->Conv2D[128]->Conv2D[1024]->maxpool->fc[512]->fc[256]. Conv2D denotes the 2D convolution layer, maxpool denotes the max-pooling operation, and fc denotes the fully-connected layer. The number inside the bracket denotes the number of filters for Conv2D or output units for fc, e.g., Conv2D[64] means a convolutional layer with 64 filters. Unless otherwise indicated, all the Conv2D layers in this section use a kernel size of [1,1] and a stride size of [1,1] with batch normalization and ReLU activation.

We use the Edge-Conv block to learn the local geometrical relationship and global semantic relationship for the point clouds. Each Edge-Conv block is composed as: kNN feature extractor->Conv2D[64]->Conv2D[64]->Conv2D[64]. The kNN feature extractor first computes the pairwise distance for each pair of points, and then constructs explicit local graphs for each point with its k nearest neighbours. After that, we will learn the features for edges between each point and its neighbours. Specifically, for a point $p\_{i}$ and its neighbors $p\_{j}, j =1 to k$, associated with features $h\_{i},h\_{j}\in R^{15}$, we will compute the edge feature $e\_{ij}= H\_{θ}(h\_{i}, h\_{j}) $for the edge $<p\_{i}, p\_{j}>$ as: $e\_{ij} =ReLU(β\_{1}⋅(h\_{i}-h\_{j}) +β\_{2}⋅h\_{i} ),$ where $θ= \{β\_{1}\in R^{15}, β\_{2}\in R^{15}\}$are the learnable parameters, and $⋅$ is the dot product for two vectors. Regarding each $θ$ as a learnable filter, we will define a set of $M $filters as $Θ= \{θ\_{1},θ\_{2}...θ\_{M}\}$ to encode information of this local graph from different perspectives. In practice, we will implement this edge feature computation procedure with a 2D convolutional layer, as the parameters in $Θ$ are shared by all points. Finally, we achieve permutation invariance for our learned representation by taking the max- or mean-pooling operation over the learned edge features for each point. The updated feature vector for the point $p\_{i}$ is computed as $h\_{i}^{'}= max/mean(e\_{ijm}),$ for $m$ from 1 to $M.$ In our architecture, we use both the max- and mean-pooling for the Edge-Conv block to learn richer local representations which could be subsequently synthesized to learn a better global representation.

Lastly, we also encode the 1-hot categorical information about the maxillary and mandible with a 2D convolutional layer and feed it into the network. This will act as a prior in our segmentation model, avoiding misclassifying teeth in maxilla to mandible or vice versa. As shown in Figure 5, the concatenated outputs will be fed into a series of 2D convolution and dropout layers to synthesize global features for segmentation.

Overall, the entire deep learning architecture corresponding to Figure 5 is composed of: Input->Transformation Net->Edge-Conv->max&meanpool->Edge-Conv -> max&meanpool -> Edge-Conv -> Conv2D[1024]-> maxpool-> Conv2D[256]-> Dropout-> Conv2D[256]-> Dropout-> Conv2D[128]-> Output.

We use a dropout layer with a dropout ratio of 0.6. The output layer is a Conv2D[33] layer without batch-normalization and ReLU activation. Besides, we use a Conv2D[128] layer to encode the one-hot categorical vector. This architecture is inspired by DGCNN with some modifications to adapt to the 3D teeth data which is of much higher resolution and morphological complexity. We use the standard cross-entropy loss for the 33-class classification problem.

**Training and Testing Settings.**

We did a hyper-parameter search in our experiment. For the value of $k$ in the kNN algorithm. We tried $k = \{5,10,15,20,25,30,35,40,45,50\}$ and finally set $k$ as 25 as setting $k$ larger than 25 would not bring significant improvements. We also did a hyper-parameter search for our neural network architecture, specifically for the number of Edge-Conv blocks. We tried networks with 1 to 5 Edge-Conv blocks and found that an architecture with 4 or 5 Edge-Conv blocks could not give consistently better results than that with 3 Edge-Conv blocks. While a model with 3 Edge-Conv blocks can perform much better than that with 1 or 2 blocks. Hence, we use 3 Edge-Conv blocks in our final implementation. We use the Adam optimizer with learning rate $η=0.001$, with a hyper-parameter search in $η=\{0.1, 0.01, 0.001, 0.0001\}.$ The other parameters for Adam follow the default setting in TensorFlow.

We use TensorFlow to implement the neural network. We use a NVIDIA V100 16GB GPU with a batch size of 3 to train the model. The model was trained for 100 epochs. We evaluated the trained model in each epoch in a validation set of 100 scans and selected the model with best validation accuracies for testing. During testing, we sample 30,000 points for each mesh, which will be segmented by DL. For the rest points (i.e., faces), we choose five coordinate-based nearest neighbours from the 30,000 points, and the neighbour with the highest probability will determine the label for that point. Neither the per face accuracy nor average area accuracy is sensitive to the number of neighbours from 5 to 50. Another strategy is to use DL to segment all the 150,000 points with a lower inference speed to avoid neighbour selection. Both inference methods give consistent results. The results in Table S2 for model inference speed were evaluated with a NVIDIA 1050Ti 4GB GPU and i7-7700 CPU 3.6 GHz, in order to be consistent with the baseline. We expect faster inference speed for our model with a more powerful GPU.

**Boundary Smoothing Hyper-parameter.**

The objective function in boundary smoothing is

$min\_{}\sum\_{f\_{i }\in F}^{}E\_{1}(y\_{i})+λ\sum\_{f\_{i}, f\_{j}\in F}^{}E\_{2}(y\_{i}, y\_{j}) $. After hyper-parameter search, we set $λ=50$ for adjacency between gingiva and teeth; $λ=200$ between any two of incisors，lateral incisors, and canines; and $λ=300$ between any two of canines, premolars and molars. Empirically, the geometric boundaries over adjacencies between gingiva and teeth, incisors and canines, premolars and molars become increasingly more distinguishable. Accordingly, we weigh more over the prior that adjacent faces tend to share the same label.

**Supplementary Results**

**Metrics.** The mIoU score is the averaged IoU score for all the classes. For a class $l$, we denote $F'\_{l}$ and $F\_{l}$ as the sets for the mesh faces which are labeled as $l$ by our system and human experts, respectively. The IoU score for class $l$ is defined as $IoU\_{l} = \frac{F'\_{l }∩F\_{l}}{F'\_{l}∪F\_{l}}$. Afterward, the mIoU score is defined as $mIoU = \frac{1}{33}\sum\_{l\in L}^{}IoU\_{l}$, where $L = \{ 0, 11-18, 21-28, 31-38, 41-48\}$. The per-face accuracy and average-area accuracy are defined in the main paper.

**Statistical Segmentation Results Analysis.**

We attach more statistical analysis results for the segmentation output from the baseline model and DLBS. The results for DL are displayed in the main paper. As we can only get the categorical outputs from the baseline and DLBS, we did not compute the AUC scores for them. The sensitivity, specificity, positive predicted values and negative predictive values are reported in Table S1-S4, which are evaluated with the same 200 scans from 100 patients as used for DL in the main paper. We regard the segmentation task as 33 independent binary classification problems to get the statistical results in these Tables.

We can notice that DLBS can achieve much higher sensitivity than the baseline in both maxillary and mandible. More specifically, DLBS can achieve mean sensitivities of 93.08% and 93.30% for maxillary and mandible, while the baseline can only get sensitivities of 87.92% and 84.07%, which are 5.16-9.23% lower than DLBS. Moreover, though DLBS shows inferior sensitivities for the third molar teeth, the baseline is even much worse than DLBS, e.g., 37.35% for tooth 18. DLBS also maintains consistently better results in terms of PPV, which is 2.05-4.02% higher than the baseline. The specificity and negative predicted values of DLBS are also higher than that of the baseline, but with a smaller gap. This is because the mesh scan contains a large portion of gingiva faces, and correctly recognizing a considerably large part of them can lead to a reasonably good specificity and negative predicted values for all the teeth.

**Visualization.**

We visualize 15 cases to show the segmentation performance of our system and the baseline model. The 15 cases are selected from the hold-out test cases with complicated dental morphologies and hence some segmentations are of low accuracies. This could help better understand the difficulties in 3D teeth segmentation. For clarity, we also attach the ground truth from human experts, and we assign specific colors for each corresponding tooth.

Figure S1 shows five cases when the baseline model usually failed to distinguish a large part of the teeth from gingiva. For example, in case 1-3, the baseline model usually failed to recognize the canine teeth. Failing to identify these canine teeth further led to mislabelling of the molar teeth, as we can notice by comparing the corresponding colors shown in baseline and ground truth. In case 4-5, it even failed to recognize a huge part of the teeth mesh faces by misclassifying them as gingiva. In contrast, our DC-Net shows much better performance, while it also made a few mistakes, such as failing to recognize the molar teeth in case 1. Overall, DC-Nets could handle most of these complicated morphologies that are hard to segment for the baseline. Moreover, in our large-scale study, DC-Net never made mistakes as the baseline, i.e., misclassifying a huge part of teeth mesh faces as gingiva.

Figure S2 shows five complicated cases when the teeth-gingiva or tooth-tooth boundary are hard to distinguish. In all of the five cases, we can notice that the baseline model failed to identify the teeth-gingiva boundaries, either for a small part of boundaries as in the canine teeth in case 9 or a significant part of the boundaries as in other cases. In case 6, the baseline model even misclassified a large part of the gingiva mesh faces as premolar teeth, leading to a completely unuseful segmentation. In comparison, DC-Net only made a minor mistake in case 10 for a molar teeth colored as light yellow. The baseline model sometimes also failed for the tooth-tooth boundary. As shown in case 9, it recognized two adjacent canine or molar teeth as a single tooth. Such mistakes are also discussed in the main paper and are not found in segmentations by DC-Net.

The last complicated dental morphology we usually encounter during segmentation is teeth scans with second- or third molars. As shown in Figure S3, we can notice that the baseline is unable to handle most of these complicated cases. In case 13-15, it completely ignored the second- or third- molars. While in case 11-12, though it successfully identified the mesh faces for molar teeth, it mistakenly labelled some gingiva as part of the teeth and failed to assign correct FDI labels to many teeth, i.e., the colors of many teeth in the baseline differ from the colors in ground truth. As discussed in the main paper, our DC-Net might make a similar mistake to ignore the molars, such as in case 11. But in most of the cases, DC-Net is able to correctly recognize the molars, as verified by a 96.9% successful rate of segmentation in the large-scale clinical performance test.

**SUPPLEMENTARY FIGURE LEGENDS**

**Figure S1: Segmentations for cases with complicated dental arch morphologies.** Five cases where the baseline model usually failed to distinguish a large part of the teeth from gingiva. DC-Net and ground truth never make such mistakes.

**Figure S2:** **Segmentations for cases with complicated teeth-gingiva or tooth-tooth boundary.** Five caseswherethebaseline model failed to identify the teeth-gingiva or tooth-tooth boundaries. DC-Net rarely made such mistakes.

**Figure S3:** **Segmentations for cases with second or third molars.** Five cases where the baseline model usually failed to recognize the molars and mislabelled lots of teeth. DC-Net might make similar mistakes as in case 11, but it generates correct segmentations in most cases and rarely mislabels teeth.

|  |
| --- |
| **Table S1: Statistical analysis of the segmentation results for individual teeth and gingiva with DLBS and the Baseline (Maxillary, tested on n=100 patients).**  |
| **Class** | **n\*** | **SENS (%)** | **SPEC (%)** |
| **DLBS** | **Baseline** | **DLBS** | **Baseline** |
| Gingiva | 100 | 97.23 (96.93, 97.53) | **97.91 (97.37, 98.45)** | **98.11 (97.56, 98.65)** | **96.73 (95.96, 97.51)** |
| Tooth 11 | 100 | 98.09 (97.89, 98.29) | 97.22 (95.15, 99.29) | 99.88 (99.80, 99.95) | 99.88 (99.78, 99.98) |
| Tooth 12 | 99 | 97.50 (95.52, 99.48) | 96.59 (93.75, 99.42) | 99.88 (99.83, 99.94) | 99.84 (99.73, 99.95) |
| Tooth 13 | 100 | 97.72 (95.76, 99.69) | 95.03 (91.15, 98.91) | 99.86 (99.80, 99.93) | **99.91 (99.82, 99.99)** |
| Tooth 14 | 96 | 96.57 (93.70, 99.45) | 91.91 (86.65, 97.17) | 99.89 (99.83, 99.95) | 99.89 (99.79, 99.99) |
| Tooth 15 | 98 | 96.20 (93.27, 99.13) | 91.35 (85.86, 96.84) | 99.82 (99.70, 99.93) | 99.41 (99.04, 99.78) |
| Tooth 16 | 100 | 96.14 (93.09, 99.19) | 91.10 (85.67, 96.53) | 99.76 (99.59, 99.94) | 99.55 (99.27, 99.83) |
| Tooth 17 | 99 | 96.86 (94.09, 99.63) | 84.93 (78.51, 91.35) | 99.75 (99.60, 99.91) | 99.59 (99.32, 99.85) |
| Tooth 18 | **35** | **57.28 (40.92, 73.65)** | **37.35 (21.40, 53.30)** | 99.81 (99.70, 99.92) | 99.82 (99.72, 99.93) |
| Tooth 21 | 100 | 98.42 (98.22, 98.62) | 95.61 (92.68, 98.54) | 99.90 (99.89, 99.91) | 99.91 (99.83, 99.99) |
| Tooth 22 | 99 | 98.43 (98.23, 98.63) | 96.49 (93.67, 99.32) | 99.92 (99.90, 99.93) | 99.86 (99.75, 99.97) |
| Tooth 23 | 100 | 97.46 (95.50, 99.43) | 94.39 (90.07, 98.72) | **99.92 (99.91, 99.93)** | 99.84 (99.72, 99.97) |
| Tooth 24 | 98 | 97.54 (95.53, 99.54) | 93.08 (88.29, 97.87) | 99.91 (99.89, 99.92) | 99.87 (99.77, 99.98) |
| Tooth 25 | 98 | 98.34 (98.11, 98.57) | 94.38 (89.97, 98.79) | 99.91 (99.90, 99.92) | 99.70 (99.47, 99.92) |
| Tooth 26 | 100 | **98.43 (97.57, 99.30)** | 95.52 (91.64, 99.40) | 99.87 (99.86, 99.89) | 99.76 (99.57, 99.94) |
| Tooth 27 | 98 | 98.21 (96.86, 99.55) | 89.56 (84.20, 94.92) | 99.85 (99.84, 99.87) | 99.75 (99.57, 99.94) |
| Tooth 28 | **34** | **61.89 (45.63, 78.15)** | **52.12 (35.34, 68.89)** | 99.80 (99.71, 99.90) | 99.78 (99.64, 99.92) |
| **Mean** | 93.08 | 87.92 | 99.76 | 99.59 |
| n\*: the number of patients with the corresponding tooth in the 100 patients. SENS: sensitivity; SPEC: specificity. The 95% confidence intervals are reported in the braces. The highest and lowest (or second lowest) scores are highlighted with bold black or bold green, respectively.  |

|  |
| --- |
| **Table S2: Statistical analysis of the segmentation results for individual teeth and gingiva with DLBS and the Baseline (Maxillary, tested on n=100 patients).**  |
| **Class** | **n\*** | **PPV (%)** | **NPV (%)** |
| **DLBS** | **Baseline** | **DLBS** | **Baseline** |
| Gingiva | 100 | 97.45 (96.80, 98.11) | 95.45 (94.33, 96.56) | **98.13 (97.95, 98.31)** | **98.55 (98.18, 98.92)** |
| Tooth 11 | 100 | 97.26 (96.13, 98.38) | **97.28 (95.03, 99.53)** | 99.93 (99.92, 99.94) | 99.89 (99.81, 99.97) |
| Tooth 12 | 99 | 96.10 (94.13, 98.07) | 95.31 (92.16, 98.46) | 99.92 (99.84, 99.99) | 99.88 (99.78, 99.98) |
| Tooth 13 | 100 | 95.74 (93.56, 97.92) | 97.11 (94.27, 99.94) | 99.93 (99.88, 99.99) | 99.85 (99.73, 99.97) |
| Tooth 14 | 96 | 97.00 (94.94, 99.06) | 96.89 (93.84, 99.94) | 99.87 (99.76, 99.98) | 99.69 (99.48, 99.90) |
| Tooth 15 | 98 | 95.38 (92.58, 98.18) | 90.28 (84.98, 95.59) | 99.86 (99.75, 99.96) | 99.65 (99.42, 99.88) |
| Tooth 16 | 100 | 96.47 (93.79, 99.15) | 92.48 (87.63, 97.33) | 99.78 (99.62, 99.94) | 99.45 (99.12, 99.79) |
| Tooth 17 | 99 | 95.86 (93.11, 98.60) | 92.74 (88.06, 97.43) | 99.80 (99.62, 99.99) | 99.15 (98.79, 99.51) |
| Tooth 18 | **35** | **90.29 (85.99, 94.58)** | **83.43 (70.59, 96.28)** | 98.84 (98.32, 99.37) | 98.32 (97.79, 98.86) |
| Tooth 21 | 100 | 97.38 (97.10, 97.65) | 97.62 (95.62, 99.63) | 99.94 (99.93, 99.95) | 99.83 (99.73, 99.94) |
| Tooth 22 | 99 | 97.21 (96.90, 97.53) | 96.54 (94.02, 99.06) | 99.95 (99.95, 99.96) | **99.89 (99.81, 99.98)** |
| Tooth 23 | 100 | 97.51 (97.26, 97.77) | 95.05 (91.14, 98.97) | 99.92 (99.86, 99.98) | 99.83 (99.71, 99.96) |
| Tooth 24 | 98 | 97.65 (97.36, 97.93) | 96.48 (93.41, 99.56) | 99.89 (99.79, 99.99) | 99.68 (99.46, 99.91) |
| Tooth 25 | 98 | 97.62 (97.37, 97.87) | 93.92 (89.81, 98.03) | **99.94 (99.93, 99.95)** | 99.79 (99.62, 99.96) |
| Tooth 26 | 100 | **98.04 (97.78, 98.31)** | 95.12 (91.26, 98.98) | 99.90 (99.84, 99.95) | 99.71 (99.46, 99.97) |
| Tooth 27 | 98 | 97.51 (97.20, 97.82) | 95.36 (91.80, 98.93) | 99.90 (99.82, 99.98) | 99.44 (99.16, 99.72) |
| Tooth 28 | **34** | **90.25 (86.23, 94.28)** | **88.78 (83.54, 94.03)** | 99.16 (98.77, 99.54) | 98.99 (98.62, 99.37) |
| **Mean** | 96.16 | 94.11 | 99.69 | 99.51 |
| n\*: the number of patients with the corresponding tooth in the 100 patients. PPV: positive predictive value; NPV: negative predictive value. The 95% confidence intervals are reported in the braces. The highest and lowest (or second lowest) scores are highlighted with bold black or bold green, respectively.  |

|  |
| --- |
| **Table S3: Statistical analysis of the segmentation results for individual teeth and gingiva with DLBS and the Baseline (Mandible, tested on n=100 patients).**  |
| **Class** | **n\*** | **SENS (%)** | **SPEC (%)** |
| **DLBS** | **Baseline** | **DLBS** | **Baseline** |
| Gingiva | 100 | 96.65 (96.20, 97.10) | **97.57 (97.07, 98.07)** | **98.57 (98.19, 98.94)** | **92.99 (91.13, 94.84)** |
| Tooth 31 | 99 | 94.55 (91.20, 97.90) | 83.83 (77.20, 90.45) | 99.87 (99.82, 99.91) | 99.70 (99.54, 99.86) |
| Tooth 32 | 97 | 98.02 (97.74, 98.30) | 82.51 (75.30, 89.72) | 99.89 (99.87, 99.91) | 99.72 (99.55, 99.88) |
| Tooth 33 | 100 | 98.46 (98.29, 98.63) | 85.31 (78.69, 91.94) | **99.91 (99.90, 99.92)** | 99.75 (99.58, 99.93) |
| Tooth 34 | 99 | 97.64 (95.65, 99.62) | 89.35 (83.34, 95.35) | 99.91 (99.90, 99.92) | 99.87 (99.75, 99.98) |
| Tooth 35 | 99 | 97.42 (95.43, 99.41) | 95.21 (91.29, 99.14) | 99.85 (99.75, 99.95) | 99.91 (99.79, 99.99) |
| Tooth 36 | 97 | 98.01 (95.99, 99.99) | 98.43 (96.39, 99.99) | 99.78 (99.60, 99.95) | 99.83 (99.67, 99.99) |
| Tooth 37 | 98 | 97.77 (95.76, 99.79) | 85.43 (78.91, 91.94) | 99.79 (99.63, 99.94) | 99.73 (99.52, 99.94) |
| Tooth 38 | **27** | **60.50 (41.69, 79.30)** | **53.04 (34.46, 71.62)** | 99.78 (99.66, 99.91) | 99.54 (99.25, 99.83) |
| Tooth 41 | 98 | 93.42 (89.52, 97.31) | 80.21 (72.77, 87.66) | 99.90 (99.89, 99.91) | 99.81 (99.68, 99.94) |
| Tooth 42 | 98 | 97.81 (97.42, 98.21) | 79.17 (71.54, 86.80) | 99.91 (99.90, 99.92) | 99.68 (99.48, 99.88) |
| Tooth 43 | 100 | 98.48 (98.30, 98.65) | 79.89 (72.38, 87.39) | 99.91 (99.89, 99.92) | 99.73 (99.55, 99.91) |
| Tooth 44 | 99 | 97.76 (95.77, 99.74) | 90.18 (84.45, 95.90) | 99.90 (99.89, 99.91) | 99.98 (99.97, 99.99) |
| Tooth 45 | 100 | 98.47 (98.16, 98.78) | 95.95 (92.46, 99.44) | 99.84 (99.75, 99.93) | **99.93 (99.86, 99.99)** |
| Tooth 46 | 98 | **98.98 (98.83, 99.12)** | 95.09 (91.09, 99.08) | 99.87 (99.85, 99.89) | 99.80 (99.64, 99.97) |
| Tooth 47 | 99 | 97.97 (95.97, 99.96) | 84.04 (77.11, 90.96) | 99.79 (99.66, 99.92) | 99.56 (99.23, 99.90) |
| Tooth 48 | **22** | **64.14 (43.93, 84.36)** | **53.94 (34.73, 73.15)** | 99.84 (99.72, 99.97) | 99.65 (99.41, 99.89) |
| **Mean** | 93.30 | 84.07 | 99.78 | 99.36 |
| n\*: the number of patients with the corresponding tooth in the 100 patients.SENS: sensitivity; SPEC: specificity. The 95% confidence intervals are reported in the braces. The highest and lowest (or second lowest) scores are highlighted with bold black or bold green, respectively.  |

|  |
| --- |
| **Table S4: Statistical analysis of the segmentation results for individual teeth and gingiva with DLBS and the Baseline (Mandible, tested on n=100 patients).**  |
| **Class** | **n\*** | **PPV (%)** | **NPV (%)** |
| **DLBS** | **Baseline** | **DLBS** | **Baseline** |
| Gingiva | 100 | **97.69 (97.09, 98.30)** | 90.48 (88.25, 92.72) | **97.96 (97.70, 98.23)** | **98.35 (97.94, 98.75)** |
| Tooth 31 | 99 | 94.98 (92.98, 96.98) | 89.08 (83.39, 94.77) | 99.87 (99.80, 99.94) | 99.56 (99.37, 99.74) |
| Tooth 32 | 97 | 96.39 (95.94, 96.84) | 89.65 (83.74, 95.57) | 99.94 (99.93, 99.95) | 99.50 (99.30, 99.71) |
| Tooth 33 | 100 | 97.45 (97.15, 97.74) | 92.88 (87.82, 97.93) | 99.95 (99.94, 99.95) | 99.50 (99.27, 99.73) |
| Tooth 34 | 99 | 97.74 (97.48, 98.01) | 95.98 (92.30, 99.65) | 99.90 (99.81, 99.99) | 99.59 (99.35, 99.82) |
| Tooth 35 | 99 | 96.62 (94.63, 98.61) | 98.18 (96.12, 99.99) | 99.85 (99.69, 99.99) | 99.82 (99.67, 99.97) |
| Tooth 36 | 97 | 97.30 (95.28, 99.31) | 97.86 (95.61, 99.99) | 99.84 (99.68, 99.99) | 99.90 (99.78, 99.99) |
| Tooth 37 | 98 | 97.01 (95.01, 99.01) | 95.08 (91.27, 98.89) | 99.85 (99.72, 99.99) | 99.03 (98.59, 99.47) |
| Tooth 38 | **27** | **90.40 (85.77, 95.04)** | **74.97 (59.39, 90.54)** | 99.17 (98.71, 99.63) | 98.71 (98.09, 99.33) |
| Tooth 41 | 98 | 96.27 (95.88, 96.65) | 92.50 (87.57, 97.42) | 99.82 (99.71, 99.93) | 99.47 (99.27, 99.67) |
| Tooth 42 | 98 | 96.85 (96.52, 97.18) | 90.01 (83.95, 96.07) | 99.94 (99.93, 99.95) | 99.38 (99.15, 99.61) |
| Tooth 43 | 100 | 97.40 (97.09, 97.70) | 91.49 (85.82, 97.15) | **99.95 (99.94, 99.95)** | 99.29 (99.02, 99.56) |
| Tooth 44 | 99 | 97.56 (97.28, 97.84) | **99.46 (99.28, 99.63)** | 99.91 (99.82, 99.99) | 99.64 (99.42, 99.85) |
| Tooth 45 | 100 | 97.00 (95.97, 98.04) | 98.08 (96.06, 99.99) | 99.94 (99.93, 99.95) | 99.81 (99.64, 99.98) |
| Tooth 46 | 98 | **98.34 (98.14, 98.54)** | 97.32 (94.86, 99.77) | 99.92 (99.91, 99.93) | **99.62 (99.30, 99.94)** |
| Tooth 47 | 99 | 97.09 (96.14, 98.03) | 94.00 (89.64, 98.36) | 99.92 (99.88, 99.95) | 99.12 (98.72, 99.53) |
| Tooth 48 | **22** | **94.21 (90.13, 98.28)** | **85.01 (77.90, 92.13)** | **98.71 (97.79, 99.63)** | **98.16 (97.18, 99.15)** |
| **Mean** | 96.49 | 92.47 | 99.67 | 99.32 |
| n\*: the number of patients with the corresponding tooth in the 100 patients. PPV: positive predictive value; NPV: negative predictive value. The 95% confidence intervals are reported in the braces. The highest and lowest (or second lowest) scores are highlighted with bold black or bold green, respectively.  |