

Barbara Kimbauer, DDS, PhD,\*  
Arneta Hadzic, MSc,<sup>†</sup>  
Norbert Jakse, MD, DDS, PhD,\*  
Horst Bischof, MS, PhD,<sup>†</sup> and  
Darko Stern, PhD<sup>†</sup>

## BASIC RESEARCH – TECHNOLOGY

# Automatic Detection of Periapical Osteolytic Lesions on Cone-beam Computed Tomography Using Deep Convolutional Neuronal Networks



### SIGNIFICANCE

Deep learning is becoming an increasingly powerful instrument in dentistry, intending to reduce the professional's workload and also improve the efficiency and precision concerning the interpretation of findings.

From the \*Division of Oral Surgery and Orthodontics, Department of Dental Medicine and Oral Health, Medical University of Graz; and <sup>†</sup>Institute for Computer Vision and Graphics, Graz University of Technology, Graz, Austria

Address requests for reprints to Dr Barbara Kimbauer, Division of Oral Surgery and Orthodontics, Department of Dental Medicine and Oral Health, Medical University of Graz, Billrothgasse 4, 8010 Graz, Austria.  
E-mail address: [barbara.kimbauer@medunigraz.at](mailto:barbara.kimbauer@medunigraz.at)  
0099-2399

Copyright © 2022 The Authors. Published by Elsevier Inc. on behalf of American Association of Endodontists. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).  
<https://doi.org/10.1016/j.joen.2022.07.013>

### ABSTRACT

**Introduction:** Cone-beam computed tomography (CBCT) is an essential diagnostic tool in oral radiology. Radiolucent periapical lesions (PALs) represent the most frequent jaw lesions. However, the description, interpretation, and documentation of radiological findings, especially incidental findings, are time-consuming and resource-intensive, requiring a high degree of expertise. To improve quality, dentists may use artificial intelligence in the form of deep learning tools. This study was conducted to develop and validate a deep convolutional neuronal network for the automated detection of osteolytic PALs in CBCT data sets.

**Methods:** CBCT data sets from routine clinical operations (maxilla, mandible, or both) performed from January to October 2020 were retrospectively screened and selected. A 2-step approach was used for automatic PAL detection. First, tooth localization and identification were performed using the SpatialConfiguration-Net based on heatmap regression. Second, binary segmentation of lesions was performed using a modified U-Net architecture. A total of 144 CBCT images were used to train and test the networks. The method was evaluated using the 4-fold cross-validation technique. **Results:** The success detection rate of the tooth localization network ranged between 72.6% and 97.3%, whereas the sensitivity and specificity values of lesion detection were 97.1% and 88.0%, respectively.

**Conclusions:** Although PALs showed variations in appearance, size, and shape in the CBCT data set and a high imbalance existed between teeth with and without PALs, the proposed fully automated method provided excellent results compared with related literature. (*J Endod* 2022;48:1434–1440.)

### KEY WORDS

Artificial intelligence; deep learning/machine learning; digital imaging/radiology; inflammation; oral diagnosis

Cone-beam computed tomography (CBCT) is widely used in dentists' daily routine for various indications, and its use continues to expand<sup>1</sup>. CBCT enables a significantly more precise radiological analysis. Studies have reported that in the case of periapical lesions (PALs), CBCT could detect 60.9% and 91.3% of lesions compared to 39.5% and 69.5% by conventional periapical radiography, respectively<sup>2,3</sup>. However, the interpretation and documentation of findings, especially incidental findings, are time-consuming and resource-intensive and require a high degree of expertise<sup>1,4</sup>. Furthermore, in a study by Leonardi Dutra et al<sup>5</sup>, rather varying intraobserver and interobserver agreements were reported based on kappa values of 0.23–0.72 and 0.38–0.86, respectively. In the case of misinterpretation, both local and systemic effects may substantially influence patients' oral and general state of health<sup>6</sup>. To improve treatment quality and patient safety, artificial intelligence, more precisely deep learning tools with the application of different types of neuronal networks, has been used to improve various medical fields<sup>7,8</sup>. Deep learning is a tool inspired by the neuronal meshes of the human brain and is the most

promising technology in the field of artificial intelligence currently<sup>9</sup>. Moreover, different artificial convolutional neuronal networks (CNNs) have been developed, among which AlexNet<sup>10</sup>, GoogLeNet<sup>11</sup>, and U-Net<sup>12</sup> are the most applied networks, which vary in technical aspects, potential applications, and performance. In comparison, Spatial Configuration-Net (SCN)<sup>13,14</sup> is less widespread. In general, such networks consist of artificial neurons representing a mathematical function of biological neurons. CNNs were specially constructed for spatially structured images, as required for radiological diagnostics. Because of the presence of more than 1 hidden layer, they resemble the so-called deep CNNs<sup>10</sup>. CNNs are generally trained in a supervised manner, where the training data consist of training instances and the corresponding ground truth labels with the target to obtain accurate predictions on new, previously unseen data<sup>9</sup>. Deep learning is becoming an increasingly more consequential instrument in dentistry<sup>15</sup>, aimed at reducing the workload of professionals in a time of constantly growing amounts of data, as well as the rise of efficiency concerning the reading and reporting of those data and an increase in the precision concerning the interpretation of findings<sup>16</sup>. Various dental applications with promising results have been documented in cariology<sup>17</sup>, endodontology<sup>18</sup>, periodontology<sup>19</sup>, orthodontics<sup>20</sup>, or forensic dentistry<sup>21</sup>. Although enface or intraoral photographs are used for gingivitis detection<sup>22</sup> or facial feature classification<sup>23</sup>, deep learning in dentistry is primarily related to radiological imaging<sup>15,16</sup>. To date, most research has been focusing on image segmentation and anatomic landmark detection<sup>15,24</sup>. For instance, Wang et al<sup>24</sup> summarized a 2-mm success detection rate of 74.84% for 100 cephalometric 2-dimensional images. Meanwhile, Payer et al<sup>14</sup> and Thaler et al<sup>25</sup> reported improved success detection rates (2-mm success detection rate, 76.00%–89.76%).

In the fields of oral radiology and pathology, literature concerning the detection of jaw lesions is rather modest. There are some reports on the automatic detection and differentiation of cystic jaw lesions<sup>26,27</sup>. For instance, Lee et al<sup>26</sup> included CBCT volumes in addition to panoramic radiographs for evaluating the automated detection of radicular, kerato, and dentigerous cysts in their research. Regarding the detection of PALs in CBCT data sets, to the best of our knowledge, only Setzer et al<sup>28</sup>, Zheng et al<sup>29</sup>, and Orhan et al<sup>30</sup> developed the CNN approach concerning this issue. Despite promising

methods, certain limitations can be found in each of them.

Therefore, the aim of this study was to develop and validate an optimized fully automated deep CNN for the automated detection and segmentation of PALs in routine 3-dimensional (3D) CBCT data sets.

## MATERIALS AND METHODS

Ethical approval was provided by the local commission (review board number 33-048 ex 20/21).

### CBCT Data Set Selection

CBCT volumes from routine clinical operations performed for different diagnostic indications (ie, implant planning, radiological assessment of impacted teeth, assessment of odontogenic tumors or other lesions, and orthodontic reasons) from January to October 2020 were retrospectively screened and selected according to the following criteria with subsequent anonymization:

1. Field of view with a representation of the entire dental arch (upper jaw, lower jaw, or both)
2. No edentulism
3. Completed root development
4. An acceptable degree of scatter and/or artifacts (exclusion of clinically insufficient interpretable data sets, ie, severe metal artifacts inhibiting individual crown visualization, and ghost effects/double images due to long-motion artifacts)

All CBCT investigations were conducted using Planmeca ProMax 3D Max (Planmeca, Helsinki, Finland) with a field of view of 10.0 × 5.9 cm or 10.0 × 9.3 cm, covering at least 1 complete dental arch, with a 200- $\mu$ m voxel size (96 kV, 5.6–9.0 mA, 12 s), which is labeled as “normal” mode by the manufacturer. An initial data set screening was performed on an MDNC-2221 monitor (resolution 1600 × 1200; size 432 × 324 mm; 59.9 Hz; Barco Control Rooms GmbH, Karlsruhe, Germany) using the Planmeca Romexis software (Planmeca, Helsinki, Finland).

### CNN Architecture

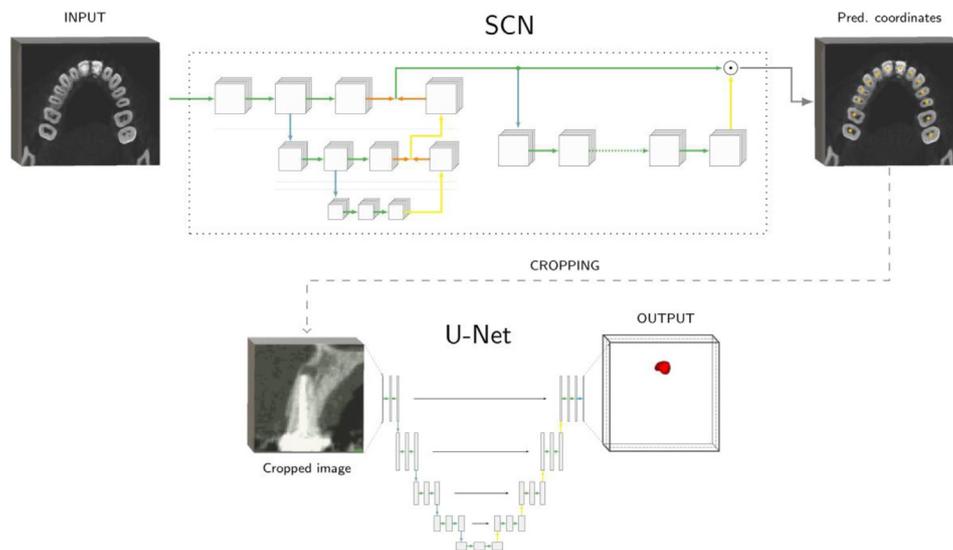
The method for automated PAL detection was developed, trained, and validated in a multistage process using neural networks (Fig. 1). Due to current graphics processing unit memory limitations, 3D CBCT images cannot be directly processed at their original resolution. Moreover, reducing the resolution would lead to the loss of details in images, which would primarily affect the detection of small lesions. Therefore, the SCN<sup>14</sup> was first

trained to automatically detect all teeth present in an input image at a low resolution. Then, based on the detected tooth coordinates, the tooth regions were automatically cropped from the input image at a high resolution and fed into the modified U-Net for segmentation. Finally, the modified U-Net was trained to produce segmentation maps for all extracted tooth regions from the input image in the original, high resolution.

The SCN is a landmark localization network consisting of 2 components—local appearance and spatial configuration—trained to produce heatmap images of teeth in an end-to-end manner. The predicted coordinates were generated directly from the predicted heatmap images, where a predicted coordinate for a specific tooth was defined as a coordinate with the highest intensity value in a predicted heatmap image for that tooth. Next, the ground truth data were generated using the open-source program Medical Imaging Interaction Toolkit, where each missing tooth or a tooth that could not be detected in the CBCT image was annotated with the coordinate (−1, −1, −1). Otherwise, the coordinates corresponded to the tooth locations in the 3D CBCT images.

The semi-automatic weighted total variation<sup>31</sup> framework was used to create the ground truth segmentation of lesions for training the modified U-Net. Several contours in the image region affected by a PAL were marked as an input. Based on these input contours, the total variation software segmented the affected regions automatically over iterations using the energy functional parameters  $\alpha = 50$ ,  $\beta = 15$ , and  $\gamma = 0.55$ . Such segmentation was followed by a manual verification and, if necessary, an adaption in the open-source software program ITK-SNAP<sup>32</sup> by an oral surgeon specializing in oral radiology. The presence of a PAL was defined as an interruption of the lamina dura, with a periapical radiolucency diameter of >0.5 mm, according to the periapical index score described by Estrela et al<sup>3</sup>.

The ground truth segmentation images used for training the modified U-Net were cropped together with the original images for each present tooth and fed into the modified U-Net. Considering an average size of a tooth and a variation in the tooth detection step, each image was resampled to an isotropic voxel size of 0.4 mm using trilinear interpolation and cropped to a size of [64, 64, 64]. The center of a cropped image resembled the coordinate of a particular tooth predicted by the SCN. Moreover, each image was translated in the direction of the root tips by



**FIGURE 1** – Schematic diagram of the method. A 3D cone-beam computed tomography (CBCT) volume of a complete dental arch was used as input to the Spatial Configuration-Net (SCN). The network was trained to predict the coordinates of each tooth in an input volume, represented by the *yellow points*. A *gray-dashed arrow* represents a cropping procedure performed on all detected teeth. For simplicity, only the cropping of one detected tooth is shown in this figure. A cropped region of the detected tooth, where the center of the region corresponds to the predicted coordinates of the tooth, was used as an input to the U-Net. Then, the network was trained to produce a segmented volume, in which the detected lesion was visualized with a *red label*. Details of the SCN architecture have been described by Payer et al<sup>14</sup>, and we modified the U-Net architecture proposed by Ronneberg et al<sup>12</sup> such that it is optimized for 3D segmentation.

fixed factors, ensuring that the entire roots of a tooth are visible in the cropped image.

The modified U-Net architecture consisted of 5 levels. Each convolutional layer had a kernel size of [3, 3, 3] and 16 filter outputs. Two convolutional operations followed by a downsampling layer were applied in the contracting path, where average pooling was performed. In the expansive path, each upsampling layer performed trilinear interpolation followed by 2 convolutional operations. A dropout of 0.3 after each convolutional layer was included. All layers had a ReLU activation function, except the last one, where the sigmoid activation function was removed to obtain logits instead of probabilities as the output of the network. The voxels in each output image lay in the range  $[-\infty, +\infty]$ . A threshold value of 0 was applied to an output image to reproduce a final prediction segmentation map. All voxels with a negative value in an output image belonged to the background, whereas all voxels with a nonnegative value belonged to the foreground in the final prediction segmentation map. In addition, bounding cuboids were used to perform tooth- or lesion-based evaluations. The network was trained with the focal loss function<sup>33</sup>, whereas focal Tversky loss<sup>34</sup> and combo loss function<sup>35</sup> were also applied for comparison purposes. All 3 loss functions were specifically designed to deal with class imbalance problems. The parameters of the loss functions were set as follows: as the lesion

voxels in each cropped image occupied <10% of a volume, causing a high imbalance between positive and negative voxels, the weighting factors  $\alpha$  and  $\beta$  were set to 0.9 in the corresponding loss functions to control the class weights. The parameter  $\gamma$ , which controls the downweighting of examples of the majority class, was treated as a hyperparameter and set to 2.0 in the focal loss and focal Tversky loss functions. In the combo loss function,  $\gamma$  was set to 0.5 to control the contribution of separate loss functions.

### Training and Validation

A total of 144 CBCT volumes were selected, comprising 54 scans of only the lower jaw, 74 of only the upper jaw, and 16 of both jaws, which resulted in a total of 2128 teeth (206 with the presence of PALs and 1922 without the presence of PALs) that were used for training and testing procedures.

Both networks were trained and tested using the 4-fold cross-validation technique. For tooth localization using SCN, 144 3D CBCT scans were divided into 4 folds with 36 volumes each. Thus, in each run, a different fold with 36 volumes was used for testing, and the 3 remaining folds were used for training the SCN. Concerning the detection of PALs using the U-Net, the 2128 cropped tooth scans were divided into 4 folds, over which 206 osteolytic lesions were distributed uniformly. In each run, a different fold with 532 tooth volumes was used for testing, and the 3 remaining folds

were used for training the U-Net.

Approximately 10% of teeth were affected by a PAL in each fold. The training and testing of the model were performed using an Intel(R) CPU 920 with an NVIDIA GeForce GTX TITAN X. The CNNs were run on the operating system Ubuntu 20.04 with Python 3.7 and TensorFlow 1.15.0. The inference time of the tooth localization network for a CBCT volume was approximately 15 s, whereas the inference time of the segmentation network, including the cropping procedure, was approximately 2–3 minutes.

### Evaluation Metrics and Statistical Analysis

Regarding tooth detection, the point-to-point error metric (PE) was used to calculate the Euclidean distance between ground truth and predicted landmarks. In addition, the image-to-point error metric, defined as the mean value of the point-to-point errors of all landmarks in a scan, was calculated. The accuracy of the tooth detection network was defined as the percentage of correctly identified landmarks over all predicted landmarks<sup>14,36</sup>. A landmark was identified correctly if the closest ground truth landmark was the correct one and the PE to the ground truth landmark was less than a specified radius  $r$ .

The overlap between the ground truth and predicted segmentation maps was

evaluated using the dice score (DSC), true positive rate (TPR), and false negative rate (FNR). The TPR was defined as the overlap only in the area of the ground truth lesion, and the FNR was defined as  $1 - \text{TPR}$ . Concerning the precision of lesion detection, the sensitivity and specificity metrics were defined as True Positive (TP)/(TP + False Negative [FN]) and True Negative (TN)/(TN + False Positive [FP]), respectively, to calculate the number of teeth with/without lesions that were correctly predicted. A predicted segmentation map for a tooth with a lesion was classified as a TP if the DSC between the ground truth and predicted segmentation map was  $>0$ ; otherwise, the predicted segmentation map was classified as an FN. A predicted segmentation map for a tooth without a lesion was classified as a TN if all voxels in the predicted segmentation map were 0; otherwise, the predicted segmentation map was classified as an FP.

## RESULTS

Both the SCN for automatic tooth localization and the modified U-Net for lesion detection/segmentation networks were evaluated over 4 cross-validations.

Using a radius of 6 mm for defining the correctly identified landmarks, the SCN achieved an average accuracy of 97.3%. More specifically, for 97.3% of predicted teeth, the closest ground truth landmark was the correct one, and the PE to the ground truth landmark was  $<6$  mm. For radii of 4, 3, and 2 mm, the achieved average accuracy values were 94.7%, 89.1%, and 72.6%, respectively. In addition, a mean PE of  $1.74 \pm 1.44$  mm and a mean image-to-point error metric of  $0.81 \pm 0.44$  mm were achieved. For the modified U-Net trained with focal loss, the obtained mean values of sensitivity and specificity metrics were  $0.97 \pm 0.03$  and  $0.88 \pm 0.04$ , respectively. Specifically, 97.1% of lesions were detected (TP), 2.9% of lesions were missed (FN), 88% of teeth without lesions were predicted as TNs, and 12% of teeth without lesions were predicted as FPs. For segmentation, the corresponding mean values obtained for the overlapping metrics TPR, FNR, and DSC were  $80.9\% \pm 3.21\%$ ,  $19.1\% \pm 3.21\%$ , and  $66.5\% \pm 3.28\%$ , respectively. Figure 2 shows examples for the segmentation results of different lesions.

For comparison, when focal Tversky loss was used instead of focal loss, the sensitivity and specificity of the detection results were  $92.2\% \pm 5.06\%$  and  $88.5\% \pm 0.79\%$ , respectively. The segmentation results of focal Tversky loss in terms of TPR, FNR, and DSC were  $82.6\% \pm 7.02\%$ ,  $17.4\% \pm 7.02\%$ , and

$70\% \pm 3.56\%$ , respectively. For combo loss, the sensitivity and specificity of the detection results were  $85.4\% \pm 3.87\%$  and  $93.7\% \pm 2.09\%$ , respectively. The segmentation results of combo loss in terms of TPR, FNR, and DSC were  $70\% \pm 3.43\%$ ,  $30\% \pm 3.43\%$ , and  $70\% \pm 3.56\%$ , respectively.

## DISCUSSION

This study evaluated a deep CNN algorithm for the automated detection and segmentation of PALs in CBCT volumes. In our opinion, the application of deep learning for detecting periapical radiolucent lesions is indeed of high clinical relevance. Moreover, we believe that it could essentially support dentists with their reporting work to save time, improve early detection, prevent overlooking lesions as incidental findings within large CBCT volumes, and increase safety for both dentists and patients. We are also convinced that both identification and segmentation of volumes are beneficial in dental surgery, for example, for surgery planning or healing assessment of lesions. This, in turn, could help with treatment planning decisions, for example, regarding nonsurgical versus surgical retreatment versus extraction. In addition, when the accuracy of artificial intelligence-based techniques is confirmed to be reliable, they can be used in education and training. Thus, recent graduates and inexperienced dentists could obtain immediate feedback from the deep CNN without the need of direct supervision by senior staff members.

When we compare related literature<sup>26,28–30</sup> with our work, the study conducted by Lee et al<sup>26</sup> is least assimilable and rather not representative of our research question regarding PAL detection. The authors<sup>26</sup> included a large number of 986 CBCT scans in addition to 1140 panoramic radiographs from 247 patients, but their focus was on different odontogenic jaw cysts. As is the nature of such pathologies, radiolucencies were broadly extended, which is different from PALs and due to this reason leads to incommensurable results. Moreover, in their study, all images were cropped and resized and finally fed as 2-dimensional slices into GoogLeNet Inception-v3, resulting in 96.1% sensitivity and 77.1% specificity<sup>26</sup>.

However, consistent with our approach are the studies conducted by Setzer et al<sup>28</sup>, Zheng et al<sup>29</sup>, and Orhan et al<sup>30</sup>. Both Setzer et al<sup>28</sup> and Zheng et al<sup>29</sup> used 20 selected 3D CBCT images, where each image resembled a limited field of view of 40 mm in diameter and height with a voxel size of 125  $\mu\text{m}$ , containing roots with at least one PAL. Orhan et al<sup>30</sup> used

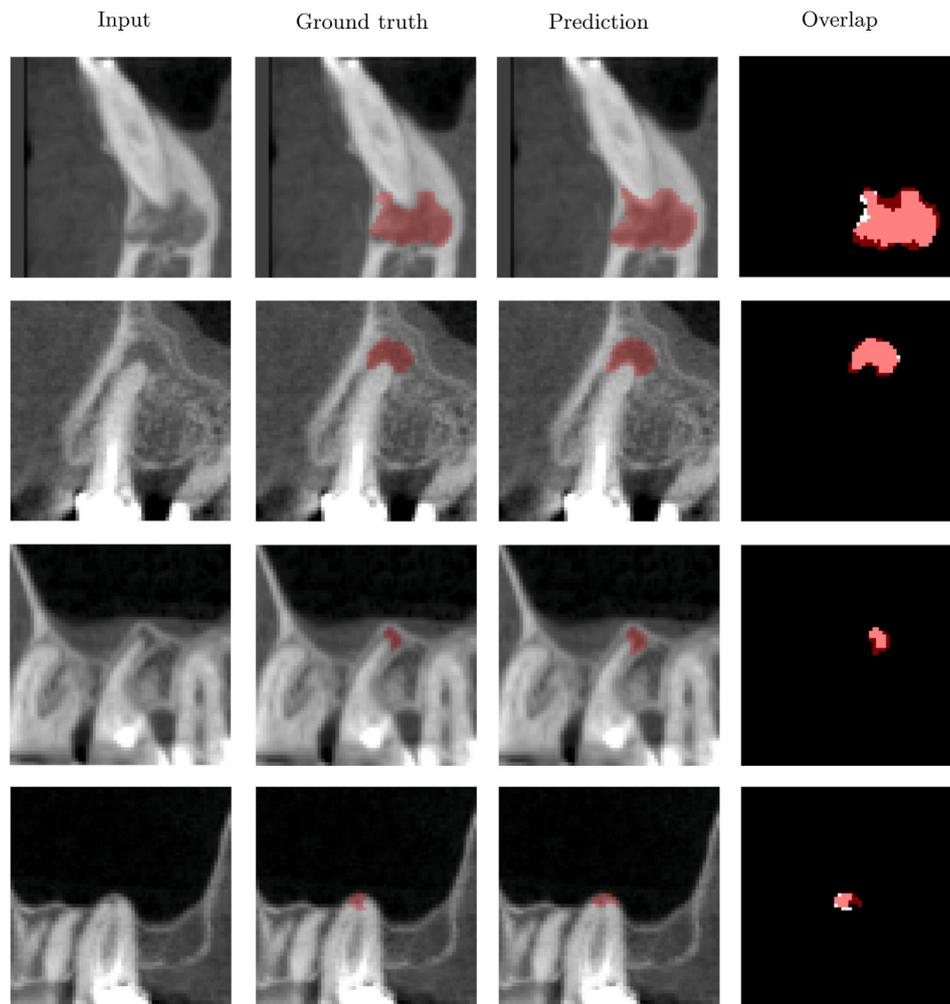
volumes from 109 patients, including 153 PAL images, with varying field of views from 50 to 100 mm in diameter and a voxel size ranging from 75 to 200  $\mu\text{m}$ . Nevertheless, we obtained our CBCT volumes from 144 patients, including a cross section of 206 PALs different in shape and size. CBCTs were taken for different indications, and not only endodontic reasons, including at least 1 jaw, and had a resolution of 200  $\mu\text{m}$ , which is, concerning the labeling mode and case number, rather comparable with the study of Orhan et al<sup>30</sup> and quite different from those of Setzer et al<sup>28</sup> and Zheng et al<sup>29</sup>.

Regarding the technical approach of Setzer et al<sup>28</sup> and Zheng et al<sup>29</sup>, as ground truth, PALs, and surrounding structures were semiautomatically segmented, followed by manual revision.

Finally, 5 representative slices rather than the actual 3D image were used to create a volume sent as input data to feed a 3D U-Net-based network. Moreover, to address the technical challenge of utilizing 3D images in a neural network, Setzer et al<sup>28</sup> performed 5-fold cross-validation with four images in each of the 1 testing and 4 training groups, whereas 4-fold cross-validation with 5 images in 1 testing and 3 training groups was performed by Zheng et al<sup>29</sup>. Thus, Setzer et al<sup>28</sup> reported 93% sensitivity and 88% specificity concerning lesion detection, whereas Zheng et al<sup>29</sup> reported 84% sensitivity.

Conceptually, the approach in this study is most similar to that reported by Orhan et al<sup>30</sup>. In their test assembly, CBCT images were uploaded into a deep CNN following a 2-step approach with first tooth detection and second lesion detection, resulting in 92.8% sensitivity. However, in their work, the imbalanced ratio between teeth with and without PALs, as well as the variations in the volume of PALs and the surrounding region, was not considered, which may have hampered their results.

In the present study, we developed and validated an optimized, fully automated deep CNN for the automated detection and segmentation of PALs with a sensitivity of 97.1%, which is an increased sensitivity in comparison with related works<sup>28–30</sup>, and a specificity of 88%, which is equal to that obtained by Setzer et al<sup>28</sup> who used different data sets than those used in our study. To detect PALs in CBCT scans with a large field of view in a fully automated manner, our end-to-end approach takes a 3D CBCT volume as an input and provides a detected and segmented PAL region as an output. This was essential for handling the small PAL volumes as the first step can be conducted on a low image resolution, whereas in the second step, when



**FIGURE 2** – Visualization of lesion detection/segmentation results for a selected cross-section. The first column shows four examples of test images with lesions. The last column shows the overlap between the corresponding ground truth and prediction segmentation maps. Dice score (DSC) values can range between 0, which represents no overlap, and 1, which represents a perfect overlap. High DSC values of 86.7% and 79.2% were obtained for the test images in the first and second rows, respectively. Lower DSC values of 53.7% and 46.9% were obtained for the test images in the third and fourth rows, respectively. For images with smaller lesions, the differences between predicted and ground truth segmentation maps had a larger impact on the DSC value.

PALs are detected and segmented, the region with high resolution is cropped around detected teeth.

In our work, special attention was given to the amortization of class imbalances because only a small portion of 3D CBCT images was occupied by lesions. Thus, class imbalances represent a common problem in medical image detection/segmentation tasks, which are not considered by conventional loss functions, such as cross-entropy loss, dice loss, and mean squared error<sup>37,38</sup>. The imbalance between 10% of teeth affected by PALs and 90% of unaffected teeth in the training and testing sets had to be addressed. Moreover, the class imbalance owing to a PAL occupying only a small image region in the cropped CBCT image of a tooth was presented. When the parameters of 3 different loss functions (focal loss, focal Tversky loss,

and combo loss) were employed, examples of the minority class, considered difficult to detect, and examples of the majority class, considered easy to detect, could be reweighed. Respectively, it was demonstrated that the focal loss function dealt best with the class imbalance problem of the data set, achieving excellent results in PAL detection in CBCT scans. The segmentation results of our method are consistent with those reported by Setzer et al<sup>28</sup>. However, the test data sets differ in resolution and volume compared with the presented work. By observing the predictions for individual lesions, we noticed that the mis-segmentation was primarily related to PALs with a very small or large volume. In images with very small lesions, minor differences between predicted and ground truth segmentation maps had a large impact on the DSC value.

Moreover, there were a few images in which the network unexpectedly did not predict a lesion. This may be due to the poor contrast in CBCT images and the varying shape and appearance of lesions, which are considered as challenging factors in general.

Regarding the limitations of our data-driven approach, the modest case number of 144 volumes has to be taken under consideration. In contrast, owing to the extended field of view, we included 2128 teeth, which is an overall remarkable number and far exceeds the case number mentioned in related studies<sup>28,29</sup>. Another limitation of this study was the usage of the supervised learning approach<sup>16</sup>, which made ground truth annotation of PALs necessary. This was done semiautomatically with an optional manual adaption by a single expert. Regarding the

advanced knowledge of this person in oral radiology and structure segmentation, a multieye principle, in contrast to Setzer et al<sup>28</sup>, or a second check from the same investigator, in contrast to Orhan et al<sup>30</sup>, was not applied. Moreover, similar to previous studies<sup>28–30</sup>, exclusively adult teeth with finished apex development were included in our study. Healthy immature teeth showing a radicular radiolucency due to the presence of the root developing apical papilla may not have been successfully differentiated from diseased teeth with PALs. The lesion segmentation in our method depends on tooth localization. In the case of tooth misdetection, the follow-up segmentation of the lesion is not feasible. However, to minimize the risk of misdetection, we used a highly accurate deep learning

approach for landmark localization<sup>14</sup> with a success rate of 97%.

In conclusion, the proposed fully automated method provided excellent results, although we included larger field of view volumes with a different resolution compared with related studies<sup>28–30</sup>. Moreover, PALs showed variations in appearance, size, and shape, and there was a high imbalance between positive and negative class examples. Therefore, we cautiously conclude that our method may be acceptable for testing it under clinical conditions. This should hopefully lead soon to the next step of applying the developed approach within a prospective study. Unfortunately, in the present situation, direct comparability to the related literature is not possible because other discussed

approaches<sup>28–30</sup>, and our methods were not evaluated on the same database.

Nevertheless, we strongly believe that our results will provide a sufficient basis for future developments, such as those concerning the inclusion of immature teeth.

## ACKNOWLEDGMENTS

*All authors gave their final approval and agree to be accountable for all aspects of the work.*

*The authors would like to thank Enago ([www.enago.com/](http://www.enago.com/)) for the English language review.*

*This study did not receive any funding.*

*The authors deny any conflicts of interest related to this study*

## REFERENCES

1. Jacobs R, Salmon B, Codari M, et al. Cone beam computed tomography in implant dentistry: recommendations for clinical use. *BMC Oral Health* 2018;18:88.
2. Lofthag-Hansen S, Huuonen S, Gröndahl K, et al. Limited cone-beam CT and intraoral radiography for the diagnosis of periapical pathology. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2007;103:114–9.
3. Estrela C, Bueno MR, Azevedo BC, et al. A new periapical index based on cone beam computed tomography. *J Endod* 2008;34:1325–31.
4. Harvey S, Patel S. Guidelines and template for reporting on CBCT scans. *Br Dent J* 2020;228:15–8.
5. Leonardi Dutra K, Haas L, Porporatti AL, et al. Diagnostic accuracy of cone-beam computed tomography and conventional radiography on apical periodontitis: a systematic review and meta-analysis. *J Endod* 2016;42:356–64.
6. Sasaki H, Hirai K, Martins CM, et al. Interrelationship between periapical lesion and systemic metabolic disorders. *Curr Pharm Des* 2016;22:2204–15.
7. Ebrahimighahnavieh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. *Comput Methods Programs Biomed* 2020;187:105242.
8. Stern D, Payer C, Urschler M. Automated age estimation from MRI volumes of the hand. *Med Image Anal* 2019;58:101538.
9. Trask AW. *Grokking deep learning*. Shelter Island, NY: Manning Publications; 2019.
10. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. In: Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12). 1. Curran Associates Inc. Red Hook; 2012. p. 1097–105.
11. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015 Jun 7–12. Boston, MA. New York: IEEE; 2015. p. 1–9.
12. Ronneberger O, Fischer P, Brox T. U-net: Convolutional Networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). 2015 Nov. 9351. Cham: Springer International Publishing; 2015. p. 234–41.
13. Payer C, Stern D, Bischof H, et al. Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselein S, Joskowicz L, Sabuncu MR, et al., editors. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016). 2016 Oct 17–21. Athens, Greek. 9901. Cham: Springer International Publishing; 2016. p. 230–8.

14. Payer C, Stern D, Bischof H, et al. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Med Image Anal* 2019;54:207–19.
15. Hung K, Montalvao C, Tanaka R, et al. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. *Dento Maxillo Fac Radiol* 2020;49:20190107.
16. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 2020;99:769–74.
17. Zheng L, Wang H, Mei L, et al. Artificial intelligence in digital cariology: a new tool for the diagnosis of deep caries and pulpitis using convolutional neural networks. *Ann Transl Med* 2021;9:763.
18. Umer F, Habib S. Critical analysis of artificial intelligence in endodontics: a scoping review. *J Endod* 2022;48:152–60.
19. Chang HJ, Lee SJ, Yong TH, et al. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep* 2020;10:7531.
20. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, et al. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making—a systematic review. *J Dent Sci* 2021;16:482–92.
21. Khanagar SB, Vishwanathaiah S, Naik S, et al. Application and performance of artificial intelligence technology in forensic odontology—a systematic review. *Leg Med (Tokyo)* 2021;48:101826.
22. Alalharith DM, Alharthi HM, Alghamdi WM, et al. A deep learning-based approach for the detection of early signs of gingivitis in orthodontic patients using faster region-based convolutional neural networks. *Int J Environ Res Public Health* 2020;17:8447.
23. Schwendicke F, Golla T, Dreher M, et al. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent* 2019;91:103226.
24. Wang CW, Huang CT, Lee JH, et al. A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal* 2016;31:63–76.
25. Thaler F, Payer C, Urschler M, et al. Modeling annotation uncertainty with Gaussian heatmaps in landmark localization. *J Mach Learn Biomed Imaging* 2021;14:1–27.
26. Lee JH, Kim DH, Jeong SN. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis* 2020;26:152–8.
27. Abdolali F, Zoroofi RA, Otake Y, et al. Automatic segmentation of maxillofacial cysts in cone beam CT images. *Comput Biol Med* 2016;72:108–19.
28. Setzer FC, Shi KJ, Zhang Z, et al. Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. *J Endod* 2020;46:987–93.
29. Zheng Z, Yan H, Setzer FC, et al. Anatomically constrained deep learning for automating dental CBCT segmentation and lesion detection. *IEEE Trans Autom Sci Eng* 2020;18:603–14.
30. Orhan K, Bayrakdar IS, Ezhov M, et al. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int Endod J* 2020;53:680–9.
31. Bredies K, Kunisch K, Pock T. Total generalized variation. *SIAM J Imaging Sci* 2010;3:492–526.
32. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31:1116–28.
33. Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* 2020;42:318–27.
34. Abraham N, Khan NM. A Novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In: *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019 Apr 8–11. Venice, Italy. New York: IEEE; 2019. p. 683–7.
35. Taghanaki SA, Zheng Y, Zhou SK, et al. Combo loss: handling input and output imbalance in multi-organ segmentation. *Comput Med Imaging Graph* 2019;75:24–33.
36. Glocker B, Zikic D, Konukoglu E, et al. Vertebrae localization in pathological spine CT via dense classification from sparse annotations. *Med Image Comput Assist Interv* 2013;16:262–70.
37. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data* 2019;6:1–54.
38. Yeung M, Sala E, Schoenlieb CB, et al. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalances medical image segmentation. *Comput Med Imaging Graph* 2022;95:102026.