Journal of **Dental Sciences**

Original Article

Convolutional-neural-network-based radiographs evaluation assisting in early diagnosis of the periodontal bone loss via periapical radiograph

I-Hui Chen [a], Chia-Hua Lin [b], Min-Kang Lee [c], Tsung-En Chen [d], Ting-Hsun Lan [e,f], Chia-Ming Chang [g], Tsai-Yu Tseng [g], Tsaipei Wang [g**], Je-Kang Du [e,f*]

[a] Division of Periodontology, Department of Dentistry, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[b] Department of Dentistry, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[c] Division of Family Dentistry, Department of Dentistry, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[d] Department of Dentistry, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan
[e] Division of Prosthodontics, Department of Dentistry, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[f] School of Dentistry, College of Dental Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[g] Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Abstract *Background/Purpose:* The preciseness of detecting periodontal bone loss is examiners dependent, and this leads to low reliability. The need for automated assistance systems on dental radiographic images has been increased. To the best of our knowledge, no studies have quantitatively and automatically staged periodontitis using dental periapical radiographs. The purpose of this study was to evaluate periodontal bone loss and periodontitis stage on dental periapical radiographs using deep convolutional neural networks (CNNs).
Materials and methods: 336 periapical radiographic images (teeth: 390) between January 2017 and December 2019 were collected and de-identified. All periapical radiographic image datasets were divided into training dataset (n = 82, teeth: 123) and test dataset (n = 336, teeth: 390). For creating an optimal deep CNN algorithm model, the training datasets were directly

  * Corresponding author. Division of Prosthodontics, Department of Dentistry, Kaohsiung Medical University Hospital, Kaohsiung Medical University No. 100, Shih-Chuan 1st Road, Kaohsiung 807, Taiwan.
  ** Corresponding author. Department of Computer Science, National Yang Ming Chiao Tung University, No. 1001, Daxue Rd. East Dist., Hsinchu City 300093, Taiwan.
  E-mail addresses: wangts@cs.nctu.edu.tw (T. Wang), dujekang@gmail.com (J.-K. Du).

used for the segmentation and individual tooth detection. To evaluate the diagnostic power, we calculated the degree of alveolar bone loss deviation between our proposed method and ground truth, the Pearson correlation coefficients (PCC), and the diagnostic accuracy of the proposed method in the test datasets.

Results: The periodontal bone loss degree deviation between our proposed method and the ground truth drawn by the three periodontists was 6.5 %. In addition, the overall PCC value of our proposed system and the periodontists' diagnoses was 0.828 ($P < 0.01$). The total diagnostic accuracy of our proposed method was 72.8 %. The diagnostic accuracy was highest for stage III (97.0 %).

Conclusion: This tool helps with diagnosis and prevents omission, and this may be especially helpful for inexperienced younger doctors and doctors in underdeveloped countries. It could also dramatically reduce the workload of clinicians and timely access to periodontist care for people requiring advanced periodontal treatment.

Introduction

Periodontitis involves the progressive loss of the alveolar bone, gingiva, and periodontal ligaments around the tooth. Periodontitis is an inflammatory disease if it not diagnosed and treated properly could lead to an ultimately loss of teeth.[1] Chronic systematic inflammation caused by periodontal pathogens is a risk factor or risk indicator for diabetes mellitus, cardiovascular disease, osteoporosis, obesity, cancer, and erectile dysfunction have shown in many epidemiological and experimental studies.[2,3] Thus, for detecting the health and clinical outcome of periodontal disease, early detection and management of periodontal bone loss plays an important role.

Clinically, the diagnosis and staging periodontitis can be established from probing pocket depths and gingival recession, resulting in clinical attachment loss measurements, or evaluating the patient's radiographs to determine the amount of alveolar bone loss around the teeth. Radiological examinations, including bitewing and periapical films, have been widely used in clinical practice to acquire valuable information for the diagnosis and treatment of periodontal bone loss. In the last thirty years, the emerging scientific proof has been modified by the classification of periodontitis again and again.[4] A new classification framework and definition for periodontitis was developed based on a multidimensional staging and grading system offered by the American Academy of Periodontology and the European Federation of Periodontology in 2017.[5] The staging is related to the severity and extent of periodontitis at present. The grading allows the rate of progression to be calculated.[5] Clinically, periodontal health can be evaluated by measuring the clinical attachment loss (CAL) by probing pocket depths and gingival recession. However, the limits of this method's reliability are related to the probing force of each dentist, angulation, and tip diameter.[6–8] Hence, radiographic bone loss (RBL) can be used as a reference tool if CAL is not available.[5]

The preciseness of detecting periodontal bone loss is examiners dependent, and this leads to low reliability.[9] This was demonstrated by an extensive range of studies.[10]

Therefore, human inspection of such radiographs tends to be subjective or inconsistent because some dentists may not have enough specialized training or be loaded with too much work to concentrate enough when interpreting the radiographs. Hence, the need for automated assistance systems on dental radiographic images has been increased.[11]

Convolutional neural networks (CNNs) have been widely studied since roughly 2010.[12] Radiological and pathological research has produced impressive results in terms of diagnosis and prediction in this rapidly growing region. Hence, the most recently reported artificial intelligence performance has been developed mainly for medical image classification. However, it is challenging for high precision performance of the automated landmark detection. In latest years, efforts have been made to develop computerized dental X-ray image analysis systems for clinical use, such as landmark identification,[13] image segmentation, and treatment.[14–16]

In the periodontal research literature, validating a dental image analyzer tool to survey the alveolar bone loss are available in two reports.[17,18] However, in these two reports,[17,18] the critical points/positions for either horizontal or vertical measurement are manually marked by dentists. In another study,[19] textures of periapical radiographs were used as a tool for dental implant treatment planning.

In another study,[20] they detected the periodontal bone loss (PBL) on panoramic dental radiographs by applied deep convolutional neural networks (CNNs). A deep feed-CNN was trained and validated 10-times via repeated group shuffling. The results showed that the accuracy of the CNN and dentist were 0.81 and 0.76, respectively, but the CNN was not statistically significantly superior compared to the dentists. The proposed method by Krois et al.[20] showed some limitations: their deep convolutional neural networks (DCNNs) are trained using manually cropped teeth patches, the dataset used in their study was small (n = 85), and their DCNN architecture was shallow. In another study,[21] transfer learning and lesion correlation from prior information to detect the periodontal bone loss (PBL) in

panoramic dental radiographs was developed. The study[21] used 12,179 panoramic dental radiographs and the results showed that the method achieved the F1 score of 0.75 on the test set, whereas the average performance of dental clinicians was 0.69. To detect the loss of periodontal bone, it used panoramic dental radiographs in these two studies.[20,21] However, both results did not show that the CNN was superior to the dentists. Panoramic dental radiographs capture a wide field of view, which results in low resolution for each individual tooth. This hinders the detection of local morphological changes in PBL, and consequently, the overall sensitivity performance of both dental clinicians and deep convolutional neural networks (CNNs) for PBL detection is limited. For diagnosing periodontitis or periapical lesions such as abscesses or cysts, a periapical radiograph is the best choice, as both types of anomalies usually occur around tooth boundaries and can only be significant enough to be detected in close-up views.[22] Hence, in this paper, we propose a CNN-based algorithm to segment each tooth and evaluate the PBL on the periapical radiographs.

Khan et al.[23] used off-the-shelf networks (specifically U-Net and DenseNet) to segment periapical radiographs and identify their key features, one of which is the area of PBL. Lin et al.[11] used classical CAD and image processing methods, and the bone loss measurement system first adopted the methods TSLS[22] and ABLifBm[24] to extract tooth contours and bone loss areas in periapical radiographs. The proposed method is then applied to locate the positions of the CEJ, alveolar crest (ALC), and apex of the tooth root (APEX). This method[11] solely measured horizontal PBL but did not compare severity grades to clinical estimates, and was only tested on 18 individual teeth from 12 periapical radiographs, which is an extremely limited dataset. However, there is still room for improvement in the algorithms of tooth segmentation and alveolar bone loss area localization for more accurate locations of APEX and ALC, respectively. Tiulpin et al.[25] successfully utilized deep learning for localizing medical landmarks. Their work[25] utilized a single hourglass network with hierarchical multi-scale parallel (HMP) residual blocks, MixUp data augmentation, and transfer learning from low-budget annotations for network training. The low-budget training can establish the region-of-interest (ROI) within the radiograph first, before the high-budget annotations fully process the exact landmark localization, thereby contributing to improved performance. Danks et al.[26] proposed an end-to-end system that includes a deep neural network with hourglass architecture to predict dental landmarks in single, double, and triple rooted teeth using periapical radiographs. They also introduced a novel adaptation of the MixUp data augmentation that improves landmark localization. The average PBL error was 10.69 %, with a severity stage accuracy of 58 % compared to the clinicians' visual evaluations. However, some limitations of the study[26] include the lack of labels allowing for assessment of PBL between roots (hindering extreme PBL classification), and a scarcity of diverse data.

To the best of our knowledge, no studies have quantitatively and automatically staged periodontitis using dental periapical radiographs. Hence, diagnosing periodontal bone loss and periodontitis stage on dental periapical radiographs was the purpose of the current study, based on the new criteria proposed at the 2017 World Workshop[5] by using deep convolutional neural networks (CNNs).

Materials and methods

Dataset collection

The study was reviewed and approved by the ethics committee of Kaohsiung Medical University Hospital (KMUHIRB-E(I)-20190,259). Three hundred and thirty-six periapical radiographic images (teeth: 390) between January 2017 and December 2019 were collected, and de-identified. All images were obtained from Kaohsiung Medical University Chung-Ho Memorial Hospital in Kaohsiung, Taiwan. Periapical radiographs of patients aged 12 years or younger, as well as images with severe noise or haziness or showing teeth that were partially present or severely distorted were excluded. Teeth with more than four roots, those that had undergone root resection surgery (e.g., hemi-section, separation, and root amputation), those who had undergone apical surgery with root resection, those with caries, those with a full restorative crown, and teeth with a shape that deviated from normal anatomical structures were also excluded.

Overview of processing methods

The processing steps of the proposed system are depicted in Fig. 1. Part of the radiographs (82 periapical radiographs) with dentist-annotated teeth masks and tissue types were used to train two neural network models: a U-net[27] based model for pixel-level tissue type classification, and a Mask-RCNN[27] model for extracting the masks of the individual teeth. We use the two models because Mask-RCNN facilitates the identification and separation of individual teeth, while U-net provides more detailed information about tissues shown in the radiographs. In addition, we can combine the results for more accurate keypoint localization. The details of these components are explained below.

U-net model and tissue type classification

U-net is a U-shaped deep convolutional network that has been used for pixel-level semantic segmentation in numerous medical imaging applications, including dental X-ray images.[27] The architecture consists of a series of down-sampling operations followed by a series of up-sampling operations until the output has the same size as the input. Skip connections are used to connect the down- and up-sampling stages of the same size to maintain local details, which are concatenated with information from lower-resolution layers that provide context. Overall, these allow multi-scale information to be fused together for classification. Our U-net has four stages of down-sampling and up-sampling. The final output layer consists of six channels, which represent these six tissue types: enamel, dentin, pulp, artificial substances, background, and bone. A pixel-wise softmax operation across the six channels was used to assign the tissue type of each pixel.
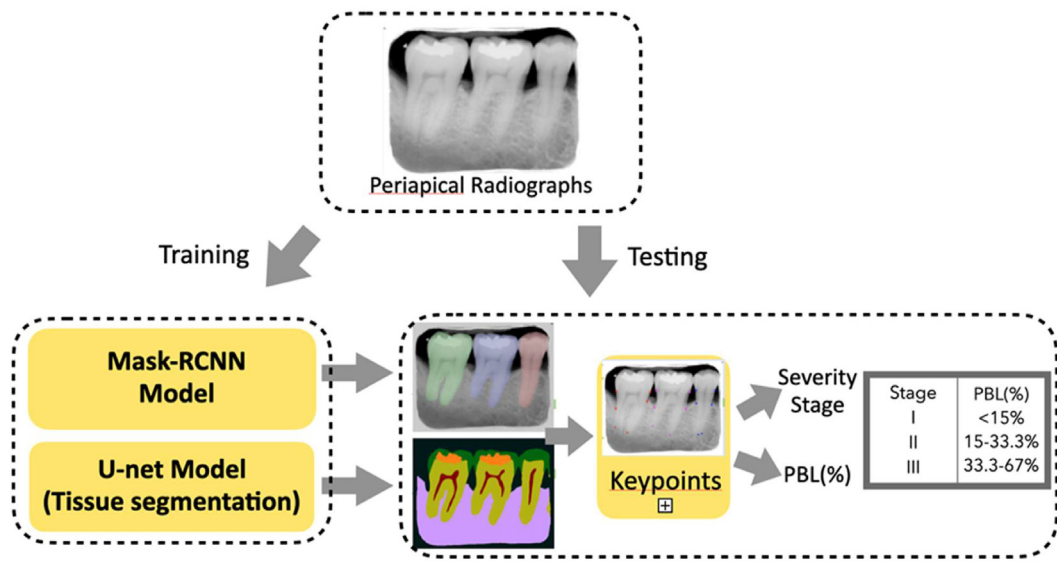
Figure 1 The block diagram of estimating bone-loss ratio from periapical radiographs. PBL: periodontal bone loss.

To increase the variety of the training data, we applied the following data augmentation approaches: random horizontal flipping, random cropping and translation, random intensity shift, random small-angle rotation (within 10°), and random deformation.

Some example results of tissue segmentation are shown in the second row of Fig. 2, whereas the top row contains the original X-ray images. We can see that the results are mostly accurate.

Mask-RCNN model and tooth mask segmentation

Mask-RCNN[28] is an extension of the well-known Faster-RCNN model.[29] In addition to existing output branches for object classification and bounding box regression, it adds an output branch that predicts the object's mask within the bounding box. The training procedure is similar to that of the original Mask-RCNN, except that the anchor box aspect ratios were changed to 1.0, 1.5, and 2.0 to better fit the vertical orientation of the teeth in the images.

We use test-time augmentation to obtain more robust results with Mask-RCNN, given the limited size of the training set. Specifically, for each input image, we applied Mask-RCNN to five copies of the source image rotated in 15° steps. The Hungarian algorithm was used to find the optimal pairing between masks in any two copies, with an IoU threshold of 0.75 to filter out mask pairs with insufficient overlap. This step allows us to obtain multiple mask candidates for each actual tooth, and the mask with the highest confidence score is retained. This procedure allows us to obtain the mask of every tooth that is at least 50 % visible in our dataset.

As there are often small gaps between the mask borders and the boundaries of the tooth and non-tooth tissues in the U-net output, we also apply a post-processing step of the mask to make them more consistent. Specifically, morphological reconstruction is applied to the extracted masks while constrained by the tooth tissue region and excluding regions occupied by adjacent teeth.

Keypoint localization

We localize the three keypoints used in the bone loss ratio estimation using the outputs of the U-net and Mask-RCNN models (Fig. 3). The procedures are as follows.

For a given tooth in the radiograph, we used its mask from Mask-RCNN to identify the tissue segmentation of the tooth. We determined the major axis of the tooth using principal component analysis (PCA) of the pixels within the mask.

1. Keypoint C (cemento-enamel junction (CEJ)) was identified as the lowest point of the enamel region along the tooth boundary. This was done for both sides of the tooth.
2. Keypoint D (the alveolar bone crest (ALC)) is identified as the location where the regions of these three tissue segmentation types: dentin, bone, and background, intersect.
3. At times, Keypoint D cannot be determined with the previous step, most likely because the tooth is so close to its adjacent tooth that no background pixels appear between them. In such cases, we take the mirror position of Keypoint D at the other side of the tooth with respect to the major axis, and find its closest point along the tooth boundary as the estimated Keypoint D. This location is constrained to be below Keypoint C.
4. The first Keypoint A (APEX) of a tooth is given by the lowest point of the mask. From this point, we drew two lines, one on each side, at 45° upward. The second Keypoint A is given by the pixel in the mask that has the largest vertical distance below the lines.

Length-based alveolar bone loss degree measurement

The normal alveolar bone crest was located at a distance of 1—2 mm from the CEJ towards the apex. If there was bone
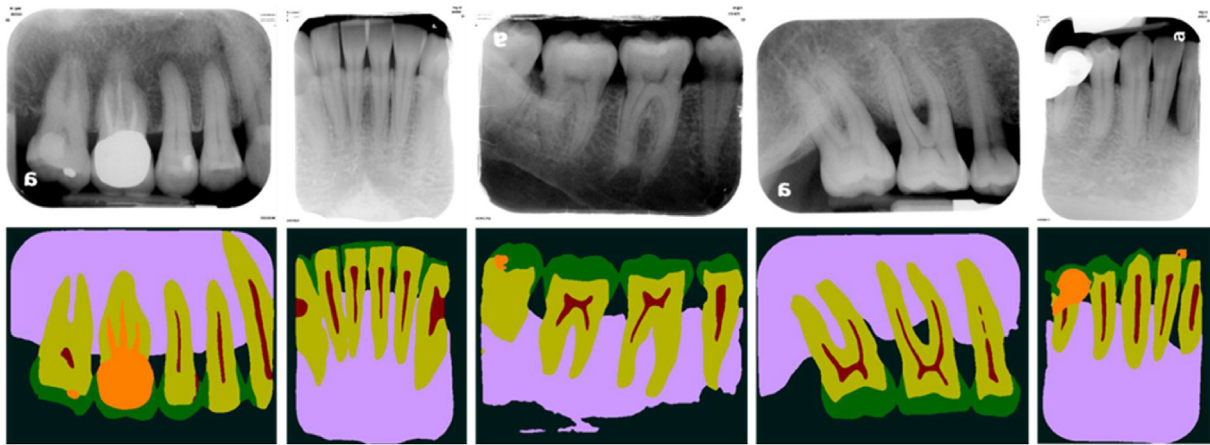
Figure 2 Example results of tissue segmentation and tooth masks. Top row: Original X-ray images. Bottom row: Tissue segmentation results (meaning of the different colors: yellow: dentin; green: enamel; red: pulp; orange: artificial substance; light purple: bone; black: background). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

loss, the alveolar bone crest was located 2 mm apical to the CEJ. The bone crest level was defined as the point along the root where an intact lamina was found.

Based on the new criteria proposed at the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions,[5] the length-based alveolar bone loss degree (ABLD) can be measured using two metrics: BL and TR, where BL is the distance between the positions of 2 mm below the cemento-enamel junction (CEJ) and alveolar bone crest (ALC), and *TR* is the distance between the CEJ and apex of the tooth (APEX). The formulae are as follows:

$$ABLD = (BL/TR) \times 100\% = (\|CEJ - ALC - 2\ mm\|/\|CEJ-APEX- 2\ mm\|) \times 100\%$$

The classification criteria for staging the periodontitis based on the ABLD of the tooth were as follows: [1] stage I: ABLD was <15 % (in the coronal third of the root); [2] stage II: the ABLD was between 15 % and 33.3 % (in the coronal third of the root); and [3] stage III: the ABLD was >33.3 % (extending to the middle third of the root and beyond).[5]

The reference test measured radiographic periodontal bone loss (in % of the root length), quantified by three independent calibrated board-certified periodontists. First, the three independent calibrated board-certified periodontists locate the positions of the CEJ, alveolar bone crest (ALC), and apex of the tooth root (APEX). Finally, the system computes the ratio of the distance between the positions of the CEJ and ALC to the distance between the positions of the CEJ and apex as the degree of bone loss for that tooth. Using the percentage rate, we were able to stage the periodontitis of a single tooth.

Statistical analysis

The data were summarized as mean ± standard deviation, and the between-group differences were tested by independent *t*-test. Percentages are summarized as categorical variables. Differences between groups were compared using the chi-square test for categorical variables. The training datasets were directly used for the segmentation of dental x-ray images and individual tooth detection, and then to create optimal model for a deep CNN algorithm. To
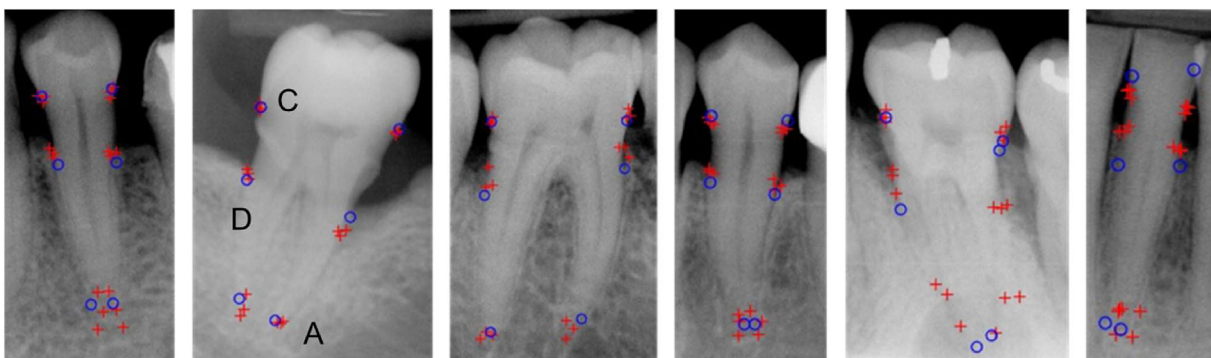
Figure 3 Example of keypoint localization (A, C, D) results (A: apex, C:CEJ, D: alveolar bone level). The red crosses are expert annotations (three per keypoint), and the blue circles are predicted locations. The two right images are examples that show significant deviations between annotated and predicted locations exist. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

evaluate the diagnostic power, we calculated the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), receiver operating characteristic (ROC) curve, area under the ROC curve (AUROC), and confusion matrix in the test datasets. We also calculated Pearson correlation coefficients (PCC) and intraclass correlation coefficients (ICC) between the CNN-based algorithm and periodontists using MATLAB (MathWorks Inc., Natick, Massachusetts, USA). A tooth was assessed as positive or negative according to two cut-off values (0.15 and 0.33). The statistical significance was set at $P < 0.05$.

Results

We tested 390 teeth images segmented from 336 periapical radiographs. All periapical radiographic image datasets were divided into a training dataset (n = 82, teeth: 123), a validation dataset (n = 20), and a test dataset (n = 336, teeth: 390).

Alveolar bone loss degree measurement

Table 1 lists the alveolar bone loss degree deviation between the CNN-based algorithm and the ground truth which is calculated based on the positions drawn by three periodontists and the average deviation between the proposed method and ground truth is 6.5 %. Otherwise, the average deviation of three periodontists were 2.9 %, 4.2 %, and 4.2 %, respectively.

Classification performance of correlations between the CNN-based algorithm and the three periodontists

The Pearson correlation coefficients (PCC) between the CNN-based algorithm and the three periodontists' diagnoses were 0.838 ($P < 0.001$), 0.795 ($P < 0.001$), and 0.85($P < 0.001$), respectively (Table 2). The PCC between the CNN-based algorithm and the three periodontists' diagnoses showed the highest correlation. In addition, the overall PCC value of the CNN-based algorithm and the periodontists' diagnoses was 0.828. This PCC showed a strong correlation between the periodontists and CNN-based algorithm diagnoses. There was also a strong correlation between the three periodontists (PCC = 0.834). The intraclass correlation coefficient (ICC) values of the

CNN-based algorithm and the three periodontists' diagnoses were 0.805 ($P = 0.000$), 0.658 ($P = 0.000$), and 0.832 ($P = 0.000$), respectively (Table 3). The overall ICC between the CNN-based algorithm and the three periodontists' diagnoses was 0.765, which showed a high correlation. This ICC value also indicates the excellent reliability of the CNN-based algorithm for the detection of periodontal bone loss and staging periodontitis.

Assessment of the diagnostic accuracy

As shown in Fig. 4, the confusion matrix displays the results, where the diagonal elements are the numbers where the predicted diagnosis was the same as the periodontists' diagnoses, while the CNN-based algorithm misinterpreted the non-diagonal elements. The higher the classification value and the darker the shade of blue, the more accurate the diagnosis. The overall periodontists' performance was calculated by three periodontists' decisions (classification of the target tooth by the radiographic alveolar bone loss degree) on the target tooth with the majority voting among three periodontists, which served as the ground truth. The total diagnostic accuracy was 72.8 %. Otherwise, the diagnostic accuracy was the highest for stage III (94.0 %), and the diagnostic accuracies of stage I and stage II were 64.2 % and 74.3 %, respectively.

Evaluation of detection and classification performance

The performance of the CNN-based algorithm and three periodontists in periodontal bone loss detection on the test dataset are summarized in Table 3, and the receiver operating characteristic (ROC) curves are shown in Fig. 5. The selected cut-off values (threshold) of radiographic bone loss degree in ground tooth were 0.15, and 0.33. The stage of the target tooth was classified according to the degree of radiographic alveolar bone loss, which was calculated by three periodontists' annotations on the target tooth, and the stage of each target tooth with the majority voting among the three periodontists served as the ground truth. Otherwise, we measured the area under the receiver operating characteristics curve (AUROC), F1 score, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) at the cut-off values (threshold) of 0.15 and 0.33.

Comparing the accuracy of diagnosing between stage I and stage II & III (cut-off point (threshold) = 0.15 alveolar bone loss degree), the average performance of CNN-based algorithm were 0.841, 0.97, 0.638, 0.742, and 0.952 in the F1 score, sensitivity, specificity, in PPV, and NPV, respectively. With the cut-off value of stage three periodontitis (cut-off point (threshold) = 0.33 alveolar bone loss degree), the diagnostic performance of the CNN-based algorithm for the diagnosis of stage I & II periodontitis and stage III were 0.764, 0.952, 0.896, 0.636, and 0.990 for F1 score, sensitivity, specificity, PPV, and NPV, respectively.

As shown in Fig. 5, the selected cut-off values (threshold) of radiographic bone loss degree in ground tooth were 0.15, and 0.33, respectively. With a cut-off value of 0.15, the area under the receiver operating

Table 1 The alveolar bone loss degree deviation between the proposed system and ground truth calculated based on the positions drawn by the three periodontists.

	The alveolar bone loss degree deviation (%)
periodontist 1 — annotation consensus	2.9
periodontist 2 — annotation consensus	4.2
periodontist 3 — annotation consensus	4.2
automated output — annotation consensus	6.5

Table 2 The Pearson correlation coefficient (PCC) and the intraclass correlation coefficient (ICC) between stages were obtained using the automatic method and those diagnosed by the periodontists. (*P < 0.001).

The Pearson correlation coefficients (PCC)	Automatic output	Periodontist 1	Periodontist 2	Periodontist 3
Automatic output	1	0.838*	0.795*	0.85*
Periodontist 1	0.838*	1	0.813*	0.887*
Periodontist 2	0.795*	0.813*	1	0.801*
Periodontist 3	0.85*	0.887*	0.801*	1
The intraclass correlation coefficient (ICC)	**Automatic output**	**Periodontist 1**	**Periodontist 2**	**Periodontist 3**
Automatic output	1	0.805*	0.658*	0.832*
Periodontist 1	0.805*	1	0.774*	0.865*
Periodontist 2	0.658*	0.774*	1	0.700*
Periodontist 3	0.832*	0.865*	0.700*	1

Table 3 Performance comparison of the CNN-based algorithm and three periodontists on the test dataset. AUROC is the area under the receiver operating characteristic curve, PPV is the positive predictive value, and NPV is the negative predictive value.

Threshold = 0.15	AUROC	F1 score	Sensitivity	Specificity	PPV	NPV
Periodontist 1	0.966	0.915	0.936	0.883	0.896	0.927
Periodontist 2	0.945	0.825	0.723	0.968	0.961	0.765
Periodontist 3	0.980	0.932	0.98	0.867	0.888	0.976
Periodontists average	0.964	0.891	0.88	0.906	0.915	0.889
CNN-based algorithm	0.946	0.841	0.97	0.638	0.742	0.952
Threshold=0.33	**AUROC**	**F1 score**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Periodontist 1	0.992	0.902	0.952	0.969	0.857	0.991
Periodontist 2	0.971	0.778	0.667	0.991	o.933	0.939
Periodontist 3	0.992	0.803	1	0.905	0.67	1
Periodontists average	0.985	0.828	0.873	0.955	0.82	0.977
CNN-based algorithm	0.968	0.764	0.952	0.896	0.638	0.990

characteristic (AUROC) was 0.946 for the CNN-based algorithm, 0.966 for Periodontist 1, 0.945 for Periodontist 2, and 0.980 for Periodontist 3. With a cut-off value of 0.33, the AUROC was 0.968 for the CNN-based algorithm, 0.992 for periodontitis 1, 0.971 for periodontitis 2, and 0.992 for periodontitis 3.

Evaluation of classification performance

A pairwise comparison between the predicted diagnosis and periodontists' diagnoses for classification is summarized in Table 4. Compared to Periodontist 1, the number of the classification made by the CNN-based algorithm which is more severe than the periodontists' diagnoses were 91 teeth, which is more than the number of under-estimation; compared to Periodontist 2, the number of the classification made by the CNN-based algorithm which is more severe than the periodontists' diagnoses were 155 teeth, which is more than the number of under-estimation; and compared to Periodontist 3, the number of the classification made by the CNN-based algorithm which is more severe than the periodontists' diagnoses were 73 teeth, which is also more than the number of under-estimation. Therefore, the classification made by the CNN-based algorithm is usually more severe than the periodontists' diagnoses.

Discussion

Observing the results of the periodontal bone loss degree deviation between our proposed method and the ground truth is 6.5 %. In comparison with another study,[11] it shows that the average deviation is 9.5 %. In another study,[26] the average PBL error was 10.69 %. Thus, our results achieved better outcome. Furthermore, the accuracy of this predicted model may be further improved when additional imagery data (e.g., cone-beam computerized tomography) or data sources (clinical records [e.g., clinical attachment level], systematic diseases [e.g., diabetes mellitus], smoking habits) are integrated into the analytic pipeline. Further assessment of the periodontal bone loss morphology (horizontal or vertical bone loss), rate of periodontal loss (% bone loss/age), and other radiographically assessable items (root morphology and furcation involvement) may be helpful. The integration of these data could provide more information to dentists in clinical diagnoses.[30]

The overall Pearson correlation coefficient (PCC) between the CNN-based algorithm and periodontists was 0.828. The overall intra-class correlation coefficient (ICC) was 0.765. Thus, the automatic classification of periodontal bone loss for staging periodontitis in the CNN-based
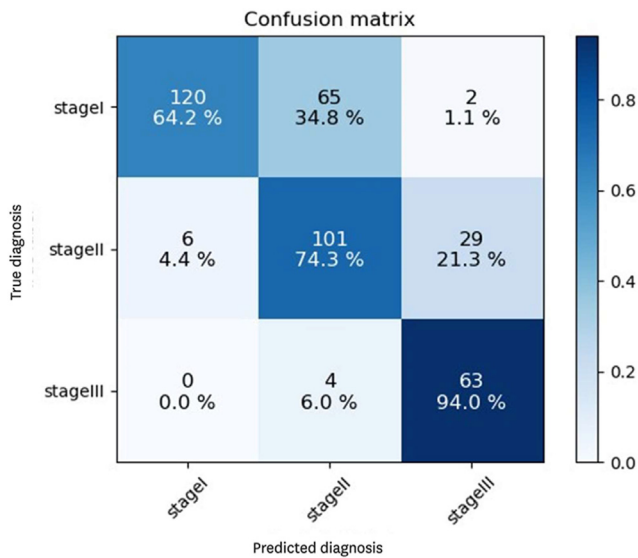
Figure 4 Multi-class classification confusion matrix among teeth in the test dataset. In the confusion matrix, each column of the matrix represents the instances of a predicted class, while each row represents the instances of the ground truth class. The averaged diagonal of a confusion matrix represents the success classification rate. The diagonal elements are the numbers where the predicted diagnosis was the same as the periodontists' diagnoses. The higher the diagonal value and the darker the shade of blue, the more accurate the diagnosis between estimation and ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

algorithm has high accuracy and excellent diagnostic reliability. Some studies have explored factors associated with the CNN modelperformance. When the datasets for training the CNN are insufficient, the CNN model learns statistical regularity that is specific to the training set, and shows low accuracy for a new dataset.[31] Thus, a larger dataset with exact annotated by specialists in medical images may be helpful for improving the detection performance of the CNN model.

In comparison with another study,[32] the diagnostic accuracy for periodontally compromised teeth was 81.0 % for premolars and 76.7 % for molars. Using 64 premolars and 64 M that were clinically diagnosed as severe periodontally compromised teeth, the accuracy of predicting extraction was 82.8 % (95 % CI, 70.1%—91.2 %) for premolars and 73.4 % (95 % CI, 59.9%—84.0 %) for molars.[32] In comparison with another study,[26] the accuracy of a severity stage was 58 % when compared to clinicians' visual evaluations of full radiographs. In the present study, the total diagnostic accuracy of staging periodontitis was 72.8 %; otherwise, the diagnostic accuracy was the highest for stage III periodontitis (97.0 %), and the diagnostic accuracy of stage I and stage II were 64.2 % and 74.3 %, respectively (Fig. 4). The present study had a higher diagnostic accuracy for identifying stage III periodontitis than stage I and stage II. The periodontal bone loss of the proposed method tended to be more severe than that in periodontists (Table 4). Therefore, in judging teeth with poor or questionable prognosis, the accuracy was better in our study. If applied to the routine clinical practice, this trained deep CNN algorithm can serve as a valuable tool for dentists in periodontitis detection, especially for teeth with advanced bone destruction
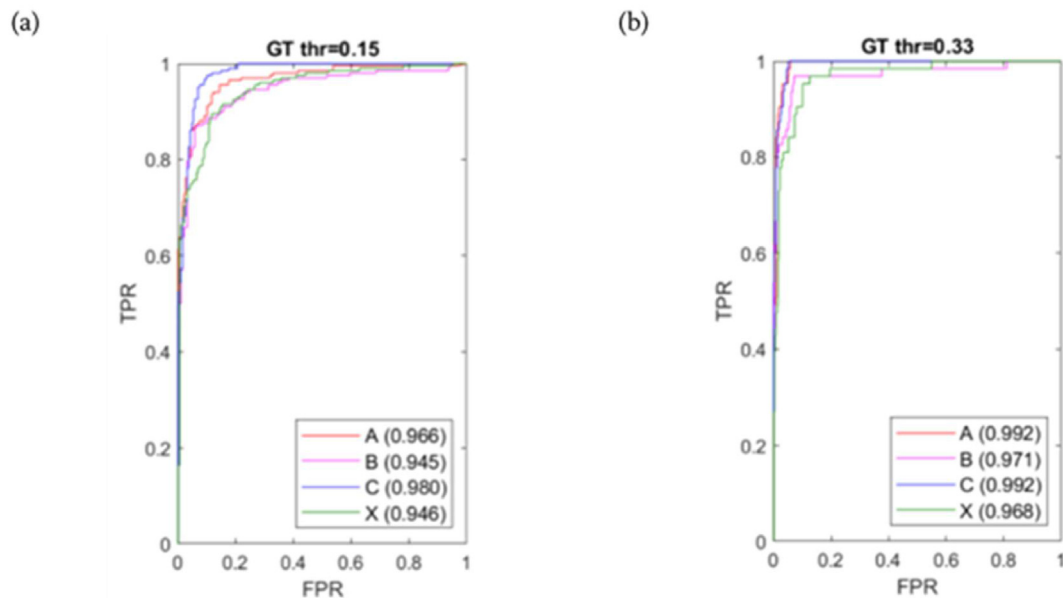

Figure 5 The performance of the CNN-based algorithm and three periodontists in the staging periodontitis, showing the true positive rate (TPR) against the false positive rate (FPR) at the cut-off values (threshold) of 0.15 and 0.33. ROC curves of the CNN-based algorithm (X, green curve) and three periodontists (A: Periodontist 1, orange curve, B: Periodontist 2, pink curve, C: Periodontist 3, purple curve) are shown, and the area under the receiver operating characteristic (AUROC) is listed within parentheses. (a) With a cut-off value of 0.15; (b) with a cut-off value of 0.33. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4 The number of the pairwise comparison between the predicted diagnosis and the diagnosis of the three periodontists.

Number(n)	More severe in Predicted diagnosis	The same classification	More severe in True diagnosis
Periodontist 1	91	282	17
Periodontist 2	155	230	5
Periodontist 3	73	286	31

(>33.3 % periodontal bone loss), irrespective of their individual level of experience and training. In other words, if a patient had advanced periodontitis and was detected by the CNN-based algorithm in a local dental clinic, the dentist can refer the patient to a specialist for further treatment and prevent medical negligence. Otherwise, this trained deep CNN algorithm can also assess the status of alveolar bone following various types of nonsurgical and surgical therapies. The application of this trained deep CNN algorithm seems promising for assisting dentists with dental imagery diagnostics.

As shown in Table 3, the selected cut-off values of radiographic bone loss to predict patients with periodontitis were 0.15 and 0.33, respectively. The negative predicted values (NPVs) of the CNN-based algorithm were 95.2 % and 99 %, respectively. The positive predicted values (PPVs) of the CNN-based algorithm were 74.2 % and 63.8 %, respectively. The sensitivities of the CNN-based algorithm were 97 % and 95.2 %, respectively. The results showed high sensitivity and negative predictive value (NPV), regardless of the cut-off definition of periodontal bone loss. This means that the CNN-based algorithm has high accuracy in diagnosing subjects that truly do not have periodontitis, and the probability of misdiagnosis is low.

As shown in Fig. 5, the AUROC of the CNN-based algorithm was similar to that of the three periodontists. Compared to another study by Kim et al.,[21] the deep neural transfer network achieved the AUROC of 0.95 in the panoramic dental radiographs. In another study,[20] the mean (SD) sensitivity and specificity of the CNN were 0.81 (0.04) and 0.81 (0.05), respectively. The mean (SD) sensitivity and specificity of the dentists were 0.92 (0.02) and 0.63 (0.14), respectively. The study[20] showed lower sensitivity than individual examiners, especially when a higher cut-off definition for periodontal bone loss was used. In our study, the CNN-based algorithm on radiographic periapical images showed a similar discrimination ability as dentists for assessing periodontal bone loss. Thus, the application of this trained deep CNN algorithm seems promising for assisting dentists in dental imagery diagnostics.

With the help of this AI imaging diagnostic tools, it may improve the convenience of learning resources to students' strengths and needs. Inexperienced dentist or students can study online at any time they need and are not restricted to having a teacher present. Otherwise, AI could provide university teachers greater support and enable

achieving educational priorities in better ways, at scale, and with lower costs. During clinical consultation, it can also provide timely assistance to inexperienced doctors, helping to diagnose periodontal disease earlier and prevent misdiagnosis.

The limitation of this study was that relatively few images of the full segmentation of dental X-ray images into the classes: enamel, dentin, pulp, crown, and restoration by calibrated board-certified periodontists were used. To achieve superior artificial intelligence performance with deep learning, the design of the deep CNN algorithm itself is important, but it is also important to have a high-quality annotated dataset. For example, 130,000 fundus photographs annotated by 54 ophthalmologists were used in the study for the detection of diabetic retinopathy.[33] In addition, anatomic structures (e.g., zygomatic arch, tuberosity, or neuro-vascular canal) can distract from key features. However, if a CNN is trained on a large enough cohort, it may learn to deal with potential artifacts. Nonetheless, unbiased lesion selection and recent advances in three-dimensional (3D) imaging would invariably improve the accuracy of the algorithm performance.[34]

Another limitation of this study is that it is impossible to make a complete diagnosis of periodontitis using only 2-dimensional periapical radiographs. For a more accurate diagnosis and prediction of periodontal bone loss, a comprehensively review over the radiographic and clinical data, such as the patient's medical history, clinical probing depth, clinical attachment level, bleeding on probing, mobility, percussion, electric pulp test, and the progression rate of periodontitis was essential. In the future, a deep CNN algorithm based on 3D image will be even more helpful for comprehensive and effective diagnosis. Lately, several studies based on artificial intelligence have investigated the potential usefulness of clinical photographs, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography scans (PET) for interpreting medical images. In particular, the deep CNN algorithm has found widespread use and has yielded promising results.[33,35,36]

In conclusion, the proposed method can help dentists diagnose and monitor periodontitis progress on periapical radiographs. Hence, dentists with different levels of training and experience may benefit from the use of CNN's image classification.

Declaration of competing interest

The authors declare no conflicts of interest with respective to the research and publication of this article.

Acknowledgments

study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

References

1. Papapanou PN, Sanz M, Buduneli N, et al. Periodontitis: consensus report of workgroup 2 of the 2017 world workshop on the classification of periodontal and peri-implant diseases and conditions. *J Periodontol* 2018;89:S173−82.
2. Choi JK, Kim YT, Kweon HI, Park EC, Choi SH, Lee JH. Effect of periodontitis on the development of osteoporosis: results from a nationwide population-based cohort study (2003−2013). *BMC Wom Health* 2017;17:77−84.
3. Lee JH, Kweon HH, Choi JK, Kim YT, Choi SH. Association between periodontal disease and prostate cancer: results of a 12-year longitudinal cohort study in South Korea. *J Cancer* 2017;8: 2959−65.
4. Caton JG, Armitage G, Berglundh T, et al. A new classification scheme for periodontal and peri-implant diseases and conditions − introduction and key changes from the 1999 classification. *J Clin Periodontol* 2018;45:S1−8.
5. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. *J Periodontol* 2018;89:S159−72.
6. Keagle JG, Garnick JJ, Searle JR, King GE, Morse PK. Gingival resistance to probing forces. I. determination of optimal probe diameter. *J Periodontol* 1989;60:167−71.
7. Garnick JJ, Silverstein L. Periodontal probing: probe tip diameter. *J Periodontol* 2000;71:96−103.
8. Trombelli L, Farina R, Silva CO, Tatakis DN. Plaque-induced gingivitis: case definition and diagnostic considerations. *J Periodontol* 2018;89:S46−73.
9. Akesson L, Håkansson J, Rohlin M. Comparison of panoramic and intraoral radiography and pocket probing for the measurement of the marginal bone level. *J Clin Periodontol* 1992; 19:326−32.
10. Choi IGG, Cortes ARG, Arita ES, Georgetti MAP. Comparison of conventional imaging techniques and CBCT for periodontal evaluation: a systematic review. *Imaging Sci Dent* 2018;48: 79−86.
11. Lin PL, Huang PY, Huang PW. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. *Comput Methods Progr Biomed* 2017;148:1−11.
12. Sklan JE, Plassard AJ, Fabbri D, Landman BA. Toward content based image retrieval with deep convolutional neural networks. *Proc SPIE-Int Soc Opt Eng* 2015;19:9417.
13. Nikneshan S, Mohseni S, Nouri M, Hadian H, Kharazifard MJ. The effect of emboss enhancement on reliability of landmark identification in digital lateral cephalometric images. *Iran J Radiol* 2015;12:e19302.
14. Nakamoto T, Taguchi A, Ohtsuka M, et al. A computer-aided diagnosis system to screen for osteoporosis using dental panoramic radiographs. *Dento Maxillo Fac Radiol* 2008;37: 274−81.
15. López-López J, Alvarez-López JM, Jané-Salas E, et al. Computer-aided system for morphometric mandibular index computation (Using dental panoramic radiographs). *Med Oral Patol Oral Cir Bucal* 2012;17:e624−32.
16. Wriedt S, Jaklin J, Al-Nawas B, Wehrbein H. Impacted upper canines: examination and treatment proposal based on 3D versus 2D diagnosis. *J Orofac Orthop* 2012;73:28−40.
17. Teeuw WJ, Coelho L, Silva A, et al. Validation of a dental image analyzer tool to measure alveolar bone loss in periodontitis patients. *J Periodontal Res* 2009;44:94−102.
18. Moutinho RP, Coelho L, Silva A, Lobo Pereira JA, Pinto M, Baptista IP. Validation of a dental image-analyzer tool to measure the radiographic defect angle of the intrabony defect in periodontitis patients. *J Periodontal Res* 2012;47:695−700.
19. Mundim MBV, Dias DR, Costa RM, Leles CR, Azevedo-Marques PM, Ribeiro-Rotta RF. Intraoral radiographs texture analysis for dental implant planning. *Comput Methods Progr Biomed* 2016;136:89−96.
20. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep* 2019; 9:8495−501.
21. Kim J, Lee HS, Song IS, DeNTNet Jung KH. Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci Rep* 2019;9:17615−23.
22. Lin PL, Huang PY, Huang PW, Hsu HC, Chen CC. Teeth segmentation of dental periapical radiographs based on local singularity analysis. *Comput Methods Progr Biomed* 2014;113: 433−45.
23. Khan HA, Haider MA, Ansari HA, et al. Automated feature detection in dental periapical radiographs by using deep learning. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2021;131: 711−20.
24. Lin PL, Huang PW, Huang PY, Hsu HC. Alveolar bone-loss area localization in periodontitis radiographs based on threshold segmentation with a hybrid feature fused of intensity and the H-value of fractional Brownian motion model. *Comput Methods Progr Biomed* 2015;121:117−26.
25. Tiulpin A, Klein S, Bierma-Zeinstra SMA. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep* 2019;9:20038−48.
26. Danks RP, Bano S, Orishko A, et al. Automating periodontal bone loss measurement via dental landmark localisation. *Int J Comput Assist Radiol Surg* 2021;16:1189−99.
27. Olaf Ronneberger PF, Brox T. Dental X-ray image segmentation using a U-shaped deep convolutional network. *International Symposium on Biomedical Imaging* 2015;1:1−13.
28. He K, Gkioxari G, Dollar P, Girshick R, Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell* 2020;42:386−97.
29. Ren S, He K, Girshick R, Sun JFaster R-CNN. Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2016;39:1137−49.
30. Schwendicke F. Tailored dentistry: from "one size fits all" to precision dental medicine. *Oper Dent* 2018;43:451−9.
31. Chang HJ, Lee SJ, Yong TH, et al. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep* 2020;10:7531−8.
32. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci* 2018;48:114−23.
33. Gulshan V, Peng L, Coram Mstumpe MC, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2023; 316:2402−10.
34. Rayner JE, Laino AM, Nufer KL, et al. Clinical perspective of 3D total body photography for early detection and screening of melanoma. *Front Med* 2018;5:152−7.
35. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 2015;175:1828−37.
36. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284: 574−82.