Original Article

Measuring the binary thickness of buccal bone of anterior maxilla in low-resolution cone-beam computed tomography via a bilinear convolutional neural network

Zhuohong Gong[1#], Xiaohui Li[2#], Mengru Shi[1#], Gengbin Cai[1], Shijie Chen[1], Zejun Ye[2], Xuejing Gan[1], Ruihan Yang[1], Ruixuan Wang[2], Zetao Chen[1]

[1]Hospital of Stomatology, Guanghua School of Stomatology, Guangdong Provincial Key Laboratory of Stomatology, Sun Yat-sen University, Guangzhou, China; [2]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China.

Contributions: (I) Conception and design: Z Chen, R Wang; (II) Administrative support: Z Chen, R Wang, M Shi; (III) Provision of study materials or patients: Z Chen, Z Gong, M Shi; (IV) Collection and assembly of data: G Cai, S Chen, X Gan, R Yang; (V) Data analysis and interpretation: Z Gong, X Li, Z Ye; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Ruixuan Wang, PhD. School of Computer Science and Engineering, Sun Yat-sen University, No. 132, Outer Ring East Road, Guangzhou 510006, China. Email: wangruix5@mail.sysu.edu.cn; Zetao Chen, PhD. Hospital of Stomatology, Guanghua School of Stomatology, Guangdong Provincial Key Laboratory of Stomatology, Sun Yat-sen University, No. 56, Lingyuan West Road, Guangzhou 510000, China. Email: chenzet3@mail.sysu.edu.cn.

Background: The thickness of the buccal bone of the anterior maxilla is an important aesthetic-determining factor for dental implant, which is divided into the thick (≥1 mm) and thin type (<1 mm). However, as a micro-scale structure that is evaluated through low-resolution cone-beam computed tomography (CBCT), its thickness measurement is error-prone under the circumstance of enormous patients and relatively inexperienced primary dentists. Further, the challenges of deep learning-based analysis of the binary thickness of buccal bone include the substantial real-world variance caused by pixel error, the extraction of fine-grained features, and burdensome annotations.

Methods: This study built bilinear convolutional neural network (BCNN) with 2 convolutional neural network (CNN) backbones and a bilinear pooling module to predict the binary thickness of buccal bone (thick or thin) of the anterior maxilla in an end-to-end manner. The methods of 5-fold cross-validation and model ensemble were adopted at the training and testing stages. The visualization methods of Gradient Weighted Class Activation Mapping (Grad-CAM), Guided Grad-CAM, and layer-wise relevance propagation (LRP) were used for revealing the important features on which the model focused. The performance metrics and efficacy were compared between BCNN, dentists of different clinical experience (i.e., dental student, junior dentist, and senior dentist), and the fusion of BCNN and dentists to investigate the clinical feasibility of BCNN.

Results: Based on the dataset of 4,000 CBCT images from 1,000 patients (aged 36.15±13.09 years), the BCNN with visual geometry group (VGG)16 backbone achieved an accuracy of 0.870 [95% confidence interval (CI): 0.838–0.902] and an area under the receiver operating characteristic (ROC) curve (AUC) of 0.924 (95% CI: 0.896–0.948). Compared with the conventional CNNs, BCNN precisely located the buccal bone wall over irrelevant regions. The BCNN generally outperformed the expert-level dentists. The clinical diagnostic performance of the dentists was improved with the assistance of BCNN.

Conclusions: The application of BCNN to the quantitative analysis of binary buccal bone thickness validated the model's excellent ability of subtle feature extraction and achieved expert-level performance.

This work signals the potential of fine-grained image recognition networks to the precise quantitative analysis of micro-scale structures.

Keywords: Buccal bone wall; deep learning; fine-grained image analysis; low-resolution; micro-scale structure

Submitted May 25, 2023. Accepted for publication Aug 28, 2023. Published online Oct 07, 2023.
doi: 10.21037/qims-23-744
View this article at: https://dx.doi.org/10.21037/qims-23-744

Introduction

The morphology of the buccal bone wall of the anterior maxilla is an important factor contributing to the aesthetic prognosis and decision-making in immediate dental implants (1). According to the thickness, the buccal bone can be classified into the thick type (thickness ≥1 mm) and the thin type (thickness <1 mm) (2). Supporting the overlying soft tissue, a thin buccal bone leads to high aesthetic risk after implant placement (3), whereas a thick buccal bone is understood to be favorable for immediate implant (4). Therefore, the thickness of the buccal bone should be precisely evaluated before implant surgery to prevent incorrect decision-making of dental implants and the accompanied aesthetic complications.

The accurate measurement of the binary thickness of buccal bone of the anterior maxilla is still very challenging in clinics. The buccal bone wall is typically a micro-scale structure that is traditionally measured by dentists through cone-beam computed tomography (CBCT), with a mean thickness of 0.75–1.05 mm (5). Meanwhile, the maximal visual ability of humans (approximately 100 microns) restricts the accurate measurement and even detection of this millimeter-scale structure (6). Another concern is that primary and basic dentists with insufficient background knowledge and clinical experience are prone to misdiagnosing the buccal bone (7). Although there are numerous edentulous patients, even a low error rate may result in extensive misdiagnosis and excessive economic loss. Worse still, with low-resolution CBCT, the buccal bone only consists of a few pixels and may result in loss of fine spatial information. Therefore, it is of great clinical significance to develop a quantitative analysis tool with high accuracy for micro-scale buccal bone even in low-resolution CBCT.

Deep learning has achieved certain success in detecting large-scale physiological and pathological structures (8-10); however, it encounters difficulties in the automatic quantitative analysis of micro-scale structures such as the buccal bone. As aforementioned, the micro-scale buccal bones manifest as pixel-level regions, and minimal errors in the feature derivation (i.e., missing one or several pixels) may result in wide divergence in its quantitative description (11). In addition, the subtle but discriminative inter-class visual features (i.e., the subtle difference in the thickness between the thick and thin buccal bone type) may be overwhelmed by the large but similar intra-class factors (i.e., the large similar background, tooth position, and so on) (12) (*Figure 1*). Further, the traditional supervised learning manner endorses the segmentation-based or key point-based strategies, which are time- and labor-intensive as they require abundant manual annotations of microstructures.

The dilemma of the deep learning-assisted analysis of micro-scale buccal bone calls on a deep learning network that can perform quantitative subtle feature extraction in an end-to-end manner. Previous research has provided some strategies that enhance the ability in subtle feature analysis, including the attention mechanism (13), the capsule neural network (14), and the integration of the external information (11). The attention module is a submodule that attends to the loosely defined fine-grained objects by the weights adjustment (15), although it tends to overfit, especially in small datasets (11). The capsule network uses vector as inputs and delivers the spatial relationship between different objects across the capsules (14). However, it mainly focuses on making full use of the spatial relationship rather than explicating the subtle features. Moreover, the integration of external information such as the multi-modality data requires additional manual annotations (11). Since the high-order feature encoding may produce more orderless translationally invariant texture descriptors that improve the fine-grained classification ability, the bilinear convolutional neural network (BCNN) model combines the features extracted from the 2 feature extractors through the outer product to produce a high-dimensional covariance

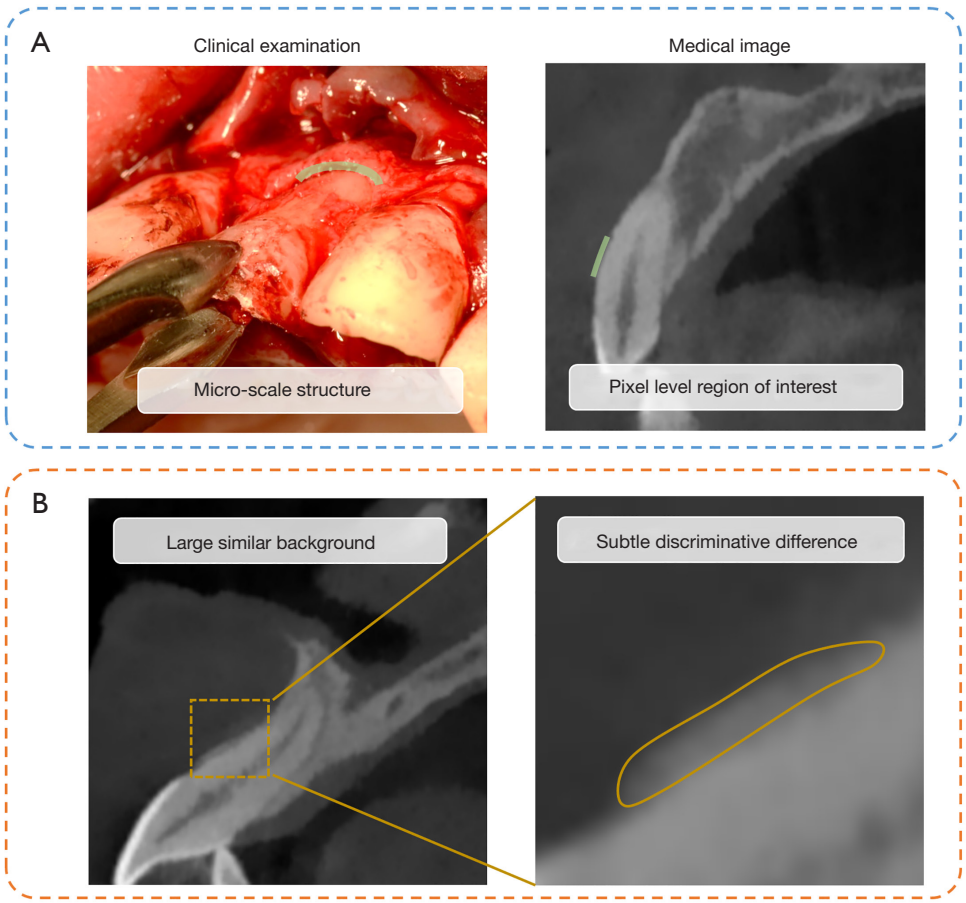


Figure 1 Schematics of the challenges in applying deep learning algorithm in quantitative micro-scale structure analysis in low-resolution images. The first challenge is that the micro-scale structure (i.e., the buccal bone of the anterior maxilla) with a few millimeters in the clinical practice represents as a few pixels in medical images (i.e., the few pixels in the CBCT images). The pixel-level error (i.e., missing a few pixels) may be magnified as a large quantitative variance (A). The second challenge is that the subtle but clinically significant inter-class difference (i.e., the thickness of the buccal bone wall) is easy to be overwhelmed by large but similar intra-class difference (i.e., the angulation, the shape of teeth and bone and the adjacent anatomic structures) (B). CBCT, cone-beam computed tomography.

matrix, and therefore has a great ability to magnify the most discriminative region and narrow the weight on the large similar factors (12). Furthermore, the architecture of BCNN realizes end-to-end gradients calculation without manual annotation (11). As a consequence, the BCNN is competent for the quantitative analysis of micro buccal bone.

In this study, we aimed to precisely analyze the binary thickness of the buccal bone of the anterior maxilla via the BCNN model and to compare the accuracy and efficiency between the BCNN and the dentists. This work provides a paradigm for the deep learning-assisted quantitative analysis of micro-scale structures in low-resolution medical images.

Methods

Ethical approval and data collection

This study received ethical approval (No. KQEC-2020-29-04) from the Medical Ethics Committee of the Hospital of Stomatology of Sun Yat-sen University. The requirement of informed consent from participants was exempted due to the retrospective nature of the study. This study collected the CBCT (NewTom VG; QR s.r.l., Verona, Italy) which were performed from 1 October 2019 to 10 July 2020 in the Hospital of Stomatology of Sun Yat-sen University. The inclusion criteria were as follows: (I) patients more than 18 years old; (II) the left and right central and lateral

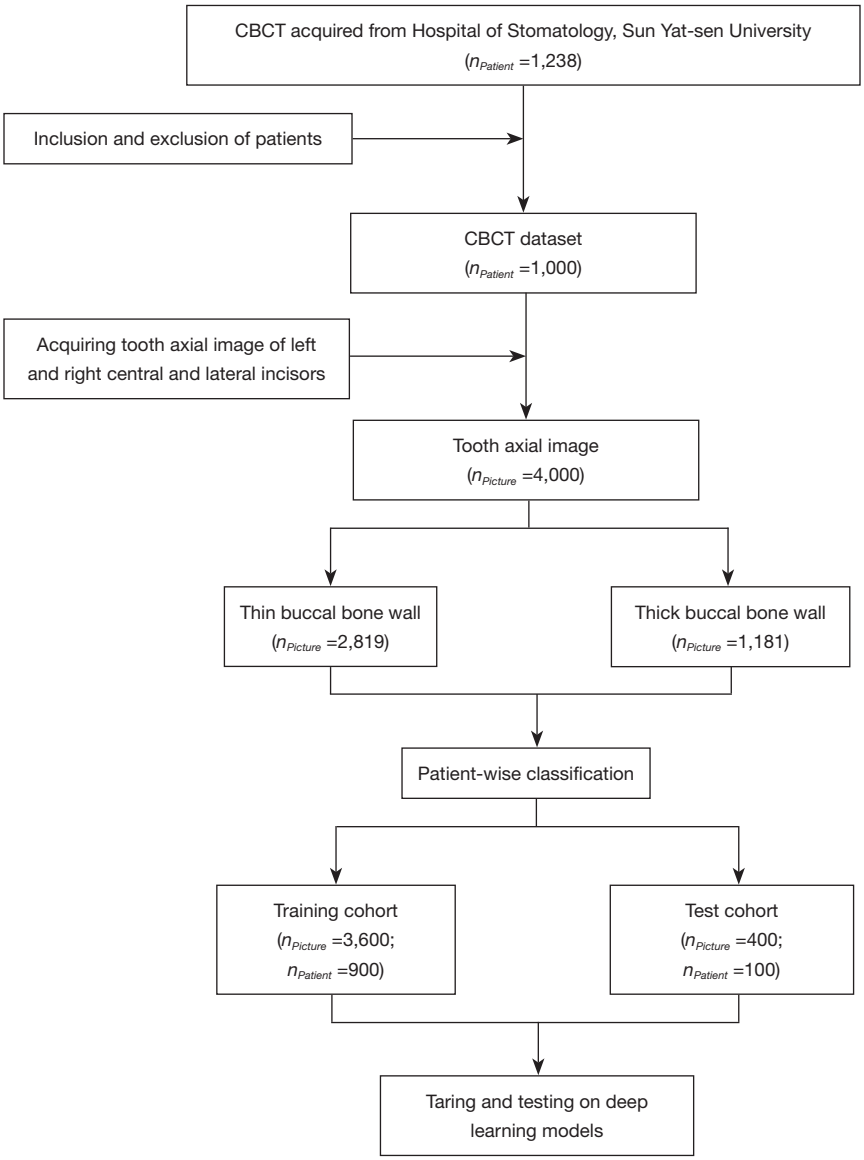


Figure 2 Flowchart of the data collection. The CBCT was acquired from Hospital of Stomatology of Sun Yat-sen University. After the inclusion and exclusion, this study included CBCT of 1,000 patients. Each piece of CBCT was pre-processed to output 4 tooth axial images of maxillary incisors, and totally 4,000 images were obtained. Then the images were divided into training cohort and test cohort (9:1) in a patient-wise manner. CBCT, cone-beam computed tomography.

incisors (#12–#22 teeth) were natural. The exclusion criteria were as follows: (I) tooth and supporting structure deformities, including severe alveolar defect, periodontitis, crowded dentition, root resorption, fracture, and periapical periodontitis; (II) severe motion artifact and metal artifact caused by orthodontic appliance or restoration that the buccal bone cannot be manually identified. The included CBCT were imported into the coDiagnostiX™ (DentalWings, Montreal, Canada) and the anterior tooth axial images of #12–#22 were obtained (*Figure 2*).

The thickness of the buccal bone wall was measured on the level of 2 mm apical to the cementoenamel junction (CEJ) due to its clinical significance using Adobe Illustrator (Adobe Systems Inc., San Jose, CA, USA). Each image was measured by 2 independent researchers who were well-trained in dentistry (had completed dentistry courses and

around 3-year of experience of dental imaging analysis) and were familiar with Adobe Illustrator (well trained by experienced dentists) with an intraclass correlation efficient (ICC) of 0.899 (Table S1). An agreement on measurement results was reached by 2 researchers. According to the genuine thickness, the buccal bone walls were classified into thick type (≥1 mm) and thin type (<1 mm) (Figure S1).

Construction and training of the deep learning models

The images from the anterior tooth axial image dataset were first cropped into 187×187 pixels to contain tooth and alveolar bone, and were resized to 224×224 as inputs of the following networks. The cropped images were divided into a training cohort (900 patients, 3,600 cropped images) and a test cohort (100 patients, 400 cropped images) in a patient-wise fashion (16). The *K*-fold cross-validation method was adopted during the training process. The training cohort was spilt into 5-, 8-, and 10-fold patient-wise. In every iteration, *K*-1 folds were used to train the model and the remaining 1-fold was used for validation. The iteration repeated *K* times and finally, *K* models were saved. The performance metrics of every model on their respective validation sets were averaged.

The BCNN architecture consists of a quadruple

$$\mathrm{B} = \left(f_A, f_B, P, C\right) \quad [1]$$

including 2 parallel feature extractors as *feature functions* (f_A,f_B), a *bilinear pooling function* (P), and a *classification function* (C). In the bilinear pooling module f, the outer product of the feature extracted from 2 backbones f_A,f_B at each location were calculated to produce a covariance matrix. Then, the bilinear feature was pooled and turned into a bilinear vector, followed by the signed square root step and L_2 normalization (12). The classifier C predicted the probability of each classification according to the bilinear feature. This study adopted pretrained visual geometry group (VGG)16, ResNet 18, ResNet 34, ResNet 50, ResNet 101, and ResNeXt 50 as backbones of BCNN (*Figure 3*). The model complexity was reduced by reconstructing the VGG16 backbones to only 1 classification head with 1 linear layer. Further, mincing the second-order pooling method, the 2 feature extractors were identical (i.e., fully-shared parameters). The 2 identical feature maps were calculated as the outer product to enrich the local feature representation and improve the computational efficacy (17). At the training stage, the conventional convolutional neural networks (CNNs) that had been pretrained on the ImageNet 1K were also applied to our dataset as the baseline, including VGG16, ResNet 18, ResNet 34, ResNet 50, ResNet 101, and ResNeXt 50.

Evaluation of the performance of the deep learning models

The model ensemble was used at the test stage. All the K models saved during the K-fold cross-validation output their prediction on the test cohort, and their predictions were averaged to produce a final outcome. Furthermore, the loss reweight method was used on the best model to try to solve the problem of sample imbalance. The models were optimized by the stochastic gradient descent (SGD) method. The initial learning rate was 0.05 with a weight decay of 1e–4, and the batch size was 32. All the experiments were performed in PyTorch on a workstation with Nvidia A30 GPU (Nvidia, Santa Clara, CA, USA). The code was released at https://github.com/xiaohhuiiiii/Project_code.

The performance metrics of the buccal bone type prediction included accuracy, precision, sensitivity, specificity, F1 score, area under the receiver operator characteristic (ROC) curve (AUC), and area under the precision-recall curve (AUPRC). The detailed formulas are listed below. The TP, TN, FP, and FN referred to true positive, true negative, false positive, and false negative, respectively.

$$\text{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \quad [2]$$

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \quad [3]$$

$$\text{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \quad [4]$$

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \quad [5]$$

$$\text{Fl score} = 2\times\frac{\text{Presicion}\times\text{Sensitivity}}{\text{Presicion}+\text{Sensitivity}} \quad [6]$$

The bootstrap method was used for calculating 95% confidence interval (95% CI). The samplings with replacement were repeated for 1,000 iterations with a sampling size of 400 images.

Visualization of the model outputs

This study applied post hoc visualization methods including

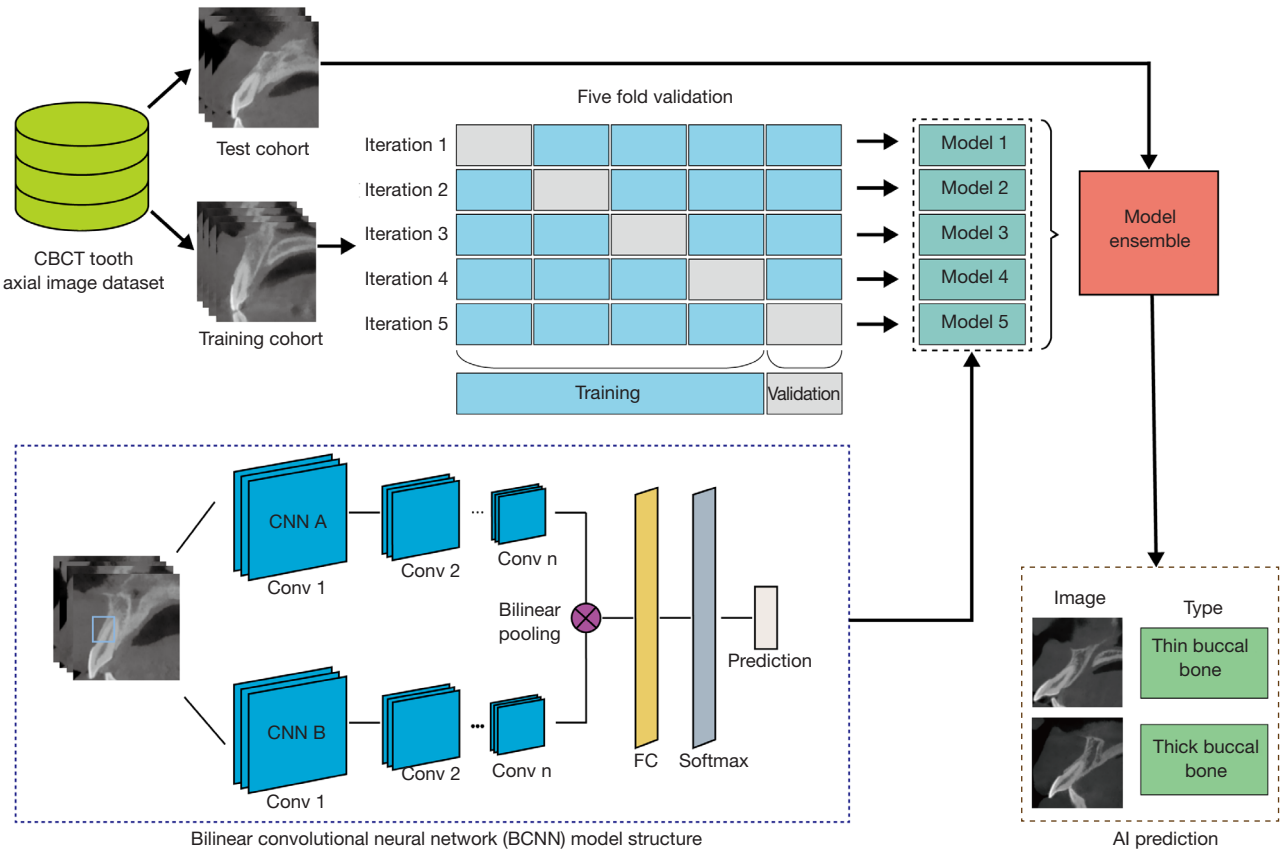


Figure 3 Workflow of the training and test process of the BCNN model for quantitative analysis of buccal bone wall of anterior maxilla. The BCNN consists of 2 parallel CNN backbones, a bilinear pooling module and a classifier. The training cohort is split into 5 folds. In each iteration, 4 folds are used for training and the remaining one for validation. It repeated 5 times and 5 models are saved with different validation folds. The test cohort is input into the 5 trained models. The model ensemble process outputs the final AI prediction by averaging the results of the 5 models on the test cohort. CBCT, cone-beam computed tomography; CNN, convolutional neural network; FC, fully connected layer; BCNN, bilinear convolutional neural network; AI, artificial intelligence.

Gradient Weighted Class Activation Mapping (Grad-CAM), Guided Grad-CAM, and layer-wise relevance propagation (LRP) to exhibit the region on which the models focused (18). The Grad-CAM produces the feature map by the weight of the gradient and highlights the area in which the model is interested. The Guided Grad-CAM is the combination of the Grad-CAM and guide-backpropagation, displaying the more specific pixels which impact the model. Instead of using the gradients, the LRP relies on the relevance scores between the neurons (18). This study presented the 3 visualization results and their corresponding original images. To further analyze the model's activation level toward different categories, the activation analysis was conducted by normalizing the activation score of a couples of truly or falsely predicted images (e.g., TP and TN pair, FP and FN pair) based on the larger score of the 2 images, rather than their respectively largest score. Then, the Grad-CAM heatmap was drawn based on the adjusted normalized activation scores of different categories.

The comparison between dental practitioners, BCNN, and human-machine fusion

This study invited 6 dental practitioners, including 2 senior dentists (more than 5-year of clinical experience), 2 junior dentists (less than 5-year of clinical experience), and 2 dental students (without clinical experience) to diagnose the buccal bone wall type from the test cohort through visual estimation. The speciality and familiarity in dental imaging increases as the level improves. To further validate the effect of the

deep learning models in the clinical application, human-machine fusion experiments were carried out using the 'Or' strategy (19), that is, the buccal bones were diagnosed as thick bone type when either the practitioner or the deep learning model interpreted that they were. The thin bone type was diagnosed only when the practitioner and the deep learning model both considered the buccal bones were (Table S2). The performance of the dentists and human-machine fusion were compared to that of the deep learning models.

Results

The demographic characteristics of the dataset

The dataset contained CBCT from 1,000 patients, among which 388 were males and 612 were females. The average age of the included patients was 36.15±13.09 years. There were 4,000 teeth images in total including left and right central and lateral incisors (1,000 images for each type of tooth). The thick bone wall type included 1,181 teeth (29.5%) and the thin bone wall type included 2,819 teeth (70.5%) (*Figure 2*).

The construction of BCNN and its performance on the training cohort

We successfully constructed BCNN with 2 identical parallel CNN backbones, followed by a bilinear pooling module and classification module. BCNN with VGG16 backbone (BCNN-VGG16) reached the best general performance in the training cohort, with an AUC of 0.916 (95% CI: 0.903–0.929), AUPRC of 0.833 (95% CI: 0.806–0.858), and accuracy of 0.869 (95% CI: 0.855–0.883). The BCNN-VGG16 outperformed all the other backbones except for the sensitivity, which was lower than that of the BCNN-ResNet18 (Table S3). The loss and accuracy of BCNN-VGG16 reached convergence at the epochs of approximately 40 during the training process (Figure S2).

Compared to the conventional CNNs, the BCNN-VGG16 had comparable training cost yet better performance. The floating-point operations (FLOPs) of BCNN-VGG16 and VGG16 were 15.38 and 15.53 GFLOPs, respectively (Table S4). The AUC of the conventional CNN models were around 0.90, with the highest AUC of 0.908 (95% CI: 0.894–0.921) and AUPRC of 0.821 (95% CI: 0.793–0.847) acquired by ResNet 18 at the training stage, which were lower than the BCNN-VGG16 (Table S5).

The performance of BCNN on the test cohort

The BCNN-VGG16 achieved the top general performance on the test cohort on almost all metrics. The accuracy, sensitivity, specificity, F1-score, AUC, and AUPRC of BCNN-VGG16 were 0.870 (95% CI: 0.838–0.902), 0.843 (95% CI: 0.776–0.906), 0.701 (95% CI: 0.617–0.783), 0.943 (95% CI: 0.914–0.968), 0.765 (95% CI: 0.700–0.825), 0.924 (95% CI: 0.896–0.948), and 0.859 (95% CI: 0.803–0.903), respectively. However, the sensitivity of the BCNN-VGG16 (0.701) was slightly lower than that of BCNN-ResNet50 (0.706) (*Table 1*). The results of different *K*-fold cross-validation and loss reweighting methods implied that the setting of 5-fold cross-validation and 1:1 loss weight ratio achieved the optimized AUC (Tables S6,S7). Meanwhile, approximately 45% of the misclassified cases (FN and FP) were related with the poor image quality, including blurriness and motion artifact, whereas the remaining cases were mainly due to the abnormality in anatomical structures and the actual thickness of buccal bone around the cut point of the binary classifications (Table S8, Figure S3).

Visualization of the outcome of the BCNN

The visualization of the conventional CNNs and BCNN was performed on the finetuned VGG16 and BCNN-VGG16. Both the Grad-CAM, Guided Grad-CAM, and LRP of VGG16 depicted large scale of the focused region, including the tooth, alveolar bone, soft tissue, and background. On the contrary, the visualization results of BCNN-VGG16 presented small and local regions, but precisely focused on the buccal bone regardless of the predicted outcome (*Figure 4*).

Besides, the activation analysis of feature maps showed that the difference in activation score was somewhat large between TP and TN, yet small between FP and FN. The heatmap of the normalized activation score showed that the attention strength of TP was greater than TN, whereas the attention strength of FP and FN was similar (Figure S4).

The comparison between dental practitioners, BCNN, and human-machine fusion

The BCNN surpassed the 3 levels of dental practitioners on almost all performance metrics. Specifically, the sensitivity of BCNN was lower than that of the junior dentists (*Table 2*). The difference existed between dental practitioners, as the senior dentists performed better than

Table 1 The performance of BCNN with different backbones in the test cohort

Model	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 score (95% CI)	AUC (95% CI)	AUPRC (95% CI)
BCNN-VGG16	0.870 (0.838, 0.902)	0.843 (0.776, 0.906)	0.701 (0.617, 0.783)	0.943 (0.914, 0.968)	0.765 (0.700, 0.825)	0.924 (0.896, 0.948)	0.859 (0.803, 0.903)
BCNN-Resnet18	0.857* (0.825, 0.890)	0.803* (0.734, 0.868)	0.702 (0.617, 0.783)	0.926* (0.893, 0.954)	0.748* (0.681, 0.811)	0.906* (0.871, 0.937)	0.829* (0.769, 0.885)
BCNN-Resnet34	0.858* (0.825, 0.887)	0.808* (0.737, 0.878)	0.691* (0.608, 0.767)	0.929* (0.896, 0.957)	0.744* (0.684, 0.804)	0.911* (0.877, 0.943)	0.843* (0.788, 0.892)
BCNN-Resnet50	0.854* (0.818, 0.885)	0.782* (0.712, 0.850)	0.706* (0.625, 0.792)	0.915* (0.882, 0.946)	0.741* (0.678, 0.802)	0.906* (0.872, 0.937)	0.837* (0.785, 0.884)
BCNN-Resnet101	0.855* (0.820, 0.885)	0.810* (0.741, 0.875)	0.674* (0.592, 0.750)	0.932* (0.904, 0.957)	0.735* (0.670, 0.796)	0.914* (0.881, 0.944)	0.851* (0.791, 0.903)
BCNN-ResNeXt50	0.850* (0.820, 0.880)	0.813* (0.745, 0.883)	0.648* (0.567, 0.733)	0.936* (0.907, 0.964)	0.721* (0.654, 0.785)	0.900* (0.866, 0.934)	0.830* (0.776, 0.880)

*, the result of BCNN-VGG16 is statistically significantly different from the result of all contrast models with *t*-test P<0.05. BCNN, bilinear convolutional neural network; CI, confidence interval; AUC, area under curve; AUPRC, area under the precision-recall curve; BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group; BCNN-Resnet18, bilinear convolutional neural network with Resnet18 as its backbone; BCNN-Resnet34, bilinear convolutional neural network with Resnet34 as its backbone; BCNN-Resnet50, bilinear convolutional neural network with Resnet50 as its backbone; BCNN-Resnet101, bilinear convolutional neural network with Resnet101 as its backbone; BCNN-ResNeXt50, bilinear convolutional neural network with ResNeXt50 as its backbone.

the other practitioners except for sensitivity. The time cost of BCNN was profoundly less than that of the dental practitioners (*Table 3*). Further, in different intervals of confusing realm of the buccal bone thickness, where all practitioners were poorly predicted, the accuracy of BCNN surpassed all levels of practitioners, either in the narrow or large intervals (*Table 4*).

After human-machine fusion, all the practitioners acquired remarkable advancement in sensitivity, AUC, and AUPRC, whereas their specificity decreased mildly (*Table 2*, *Figure 5*). The ROC and precision-recall (P-R) curves revealed that BCNN outperformed all the practitioners alone and human-machine fusion, indicating that the model performed at the expert level (*Figure 5*).

Discussion

In this study, we successfully introduced the BCNN to the micro-scale structures analysis in medical imaging and validated its success in the semi-quantitative analysis of the buccal bone from low-resolution CBCT. The best performance was achieved by BCNN-VGG16 with the AUC of 0.924 (95% CI: 0.896–0.948). The visualization showed that the BCNN precisely distinguished the buccal bone wall region from the tooth or alveolar bone region. The BCNN performed at expert-level and the dentists' overall diagnosis performance was improved when combined with BCNN.

The application of fine-grained image recognition network to quantitative micro-scale structures analysis

The primary contribution of this study is to innovatively apply the fine-grained image recognition network to the analysis of binary buccal bone thickness in the low-resolution CBCT. The fine-grained image recognition networks had previously been successfully validated in a series of competitions of natural images identification (11), whereas no attempt had been made in the quantitative analysis. In this study, the BCNN performed end-to-end encoding with minimal annotations and training cost, successfully located micro-scale region even in low-resolution medical images, and yielded satisfying performance.

The bilinear-based high-order feature interactions enable the end-to-end encoding for quantitative feature analysis with few annotations and training prerequisites. The classical texture representation methods [e.g., scale invariant feature transform (SIFT) and CNN] perform just passable results in fine-grained features analysis for their failure to learn the underlying features in

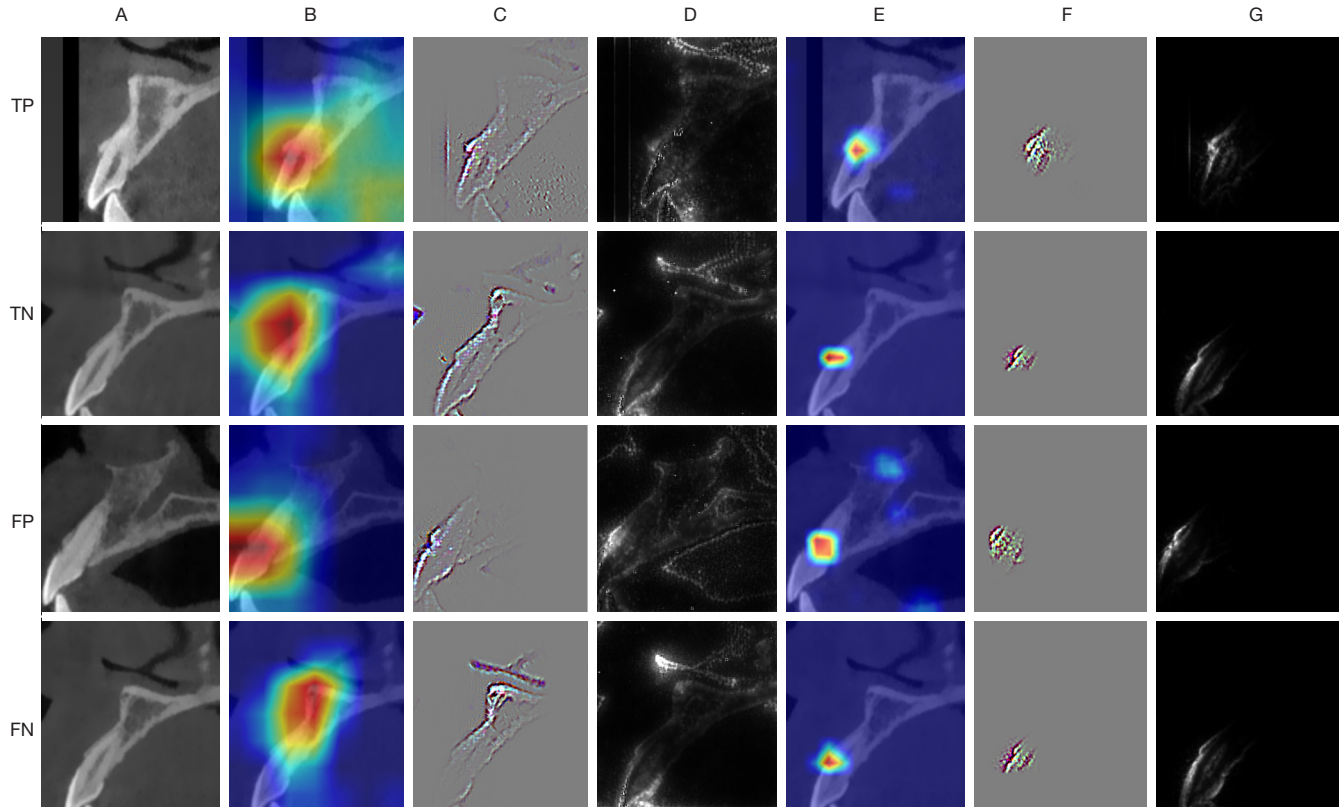


Figure 4 The visualization of the conventional CNN and BCNN. The raw image is listed in (A). As for the conventional CNN models (i.e., VGG16), the Grad-CAM (B), Guide Grad-CAM (C), and LRP (D) results show the large focus region including the whole teeth and alveolar bone. The Grad-CAM (E), Guide Grad-CAM (F), and LRP (G) of BCNN with VGG16 backbones show smaller but more precisely located regions on the buccal bone wall. TP, true positive; TN, true negative; FP, false positive; FN, false negative; CNN, convolutional neural network; VGG, visual geometry group; BCNN, bilinear convolutional neural network; Grad-CAM, Gradient Weighted Class Activation Mapping; LRP, layer-wise relevance propagation.

Table 2 The comparison of performance metrics pre- and post-human-machine fusion

Predictor	Fusion	Accuracy	Precision	Sensitivity	Specificity	F1 score
Dental student	Pre	0.71	0.51	0.73	0.70	0.60
	Post	0.74	0.54	0.87	0.69	0.67
Junior dentist	Pre	0.73	0.54	0.84	0.68	0.65
	Post	0.74	0.55	0.89	0.68	0.67
Senior dentist	Pre	0.80	0.69	0.62	0.87	0.64
	Post	0.83	0.69	0.80	0.85	0.74
BCNN-VGG16	–	0.87	0.84	0.70	0.94	0.77

BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group.

Table 3 The time cost of dental practitioners and BCNN

Predictor	Time cost per image (s)
Dental student	109
Junior dentist	113
Senior dentist	163
BCNN-VGG16	0.06

BCNN, bilinear convolutional neural network; BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group.

an end-to-end manner. On the contrary, the bilinear architecture is a directed acyclic graph, which enables better back propagation (12). In this way, the BCNN facilitates end-to-end learning of the discriminative local textures (20). To alleviate the potential exponential explosion of the computational cost incurred by high-order encoding, the parameters of the 2 identical backbones are fully-shared to improve the efficacy by reducing the parameters number while acquiring matchable result as 2 different backbones without parameters sharing (12,21). For

Table 4 The performance of dental practitioners and BCNN in different intervals of confusing realm

Intervals of confusing realm	Included patients, n (%)	Predictor	Accuracy	Precision	Sensitivity	Specificity	F1 score
0.9–1.1 mm	40 (10.00%)	Dental student	0.54	0.71	0.52	0.58	0.60
		Junior dentist	0.58	0.72	0.65	0.42	0.67
		Senior dentist	0.54	0.81	0.44	0.73	0.54
		BCNN-VGG16	0.67	0.82	0.52	0.86	0.64
0.8–1.2 mm	88 (22.00%)	Dental student	0.64	0.70	0.57	0.72	0.63
		Junior dentist	0.64	0.67	0.70	0.56	0.67
		Senior dentist	0.64	0.79	0.46	0.85	0.56
		BCNN-VGG16	0.72	0.87	0.55	0.90	0.68
0.7–1.3 mm	128 (32.00%)	Dental student	0.66	0.68	0.59	0.72	0.63
		Junior dentist	0.66	0.67	0.73	0.59	0.68
		Senior dentist	0.67	0.78	0.50	0.85	0.59
		BCNN-VGG16	0.75	0.84	0.63	0.87	0.72
0.6–1.4 mm	157 (39.25%)	Dental student	0.67	0.66	0.62	0.71	0.64
		Junior dentist	0.72	0.66	0.76	0.59	0.69
		Senior dentist	0.71	0.80	0.54	0.86	0.63
		BCNN-VGG16	0.78	0.86	0.66	0.90	0.75
0.5–1.5 mm	191 (47.75%)	Dental student	0.68	0.66	0.64	0.72	0.65
		Junior dentist	0.70	0.67	0.78	0.63	0.71
		Senior dentist	0.72	0.80	0.55	0.87	0.64
		BCNN-VGG16	0.81	0.88	0.68	0.92	0.77

BCNN, bilinear convolutional neural network; BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group.

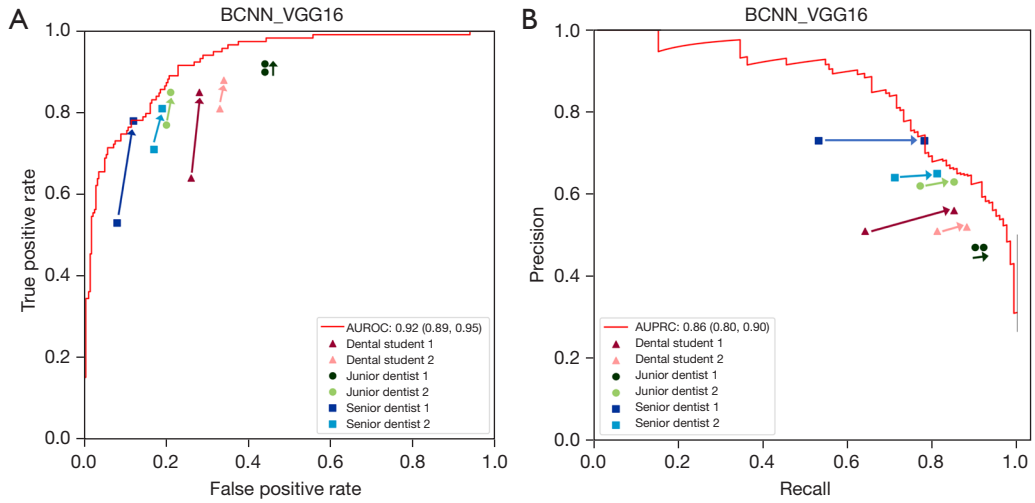


Figure 5 ROC curve (A) and P-R curves (B) of the BCNN and dental practitioners. The shape of triangle, circle and square refers to the dental student, junior dentist, and senior dentist, respectively. Each participant of the same roles is in different colors. The arrows show the advancement of pre-human-machine fusion to post fusion. BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group; AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; ROC, receiver operating characteristic; P-R, precision-recall; BCNN, bilinear convolutional neural network.

the simplification of the backbones, for example, the prune of classification head is another contributor to reducing training cost. In this way, BCNN realizes the high efficacy in an end-to-end manner.

The BCNN exhibits precise micro-scale features learning even in low-resolution medical images. Even though the CBCT is widely used in dentistry with irreplaceable advantages including lower radiation dose, lower cost, and shorter scanning time (22), its main drawback is the lower spatial resolution compared with the planning computed tomography (CT) (23,24). In this study, the resolution of CBCT was as low as 196.5 μm per pixel and the region of buccal bone accounted for only 20×5 pixels. The covariance matrix-based representation of BCNN generates numerous high-order features in a non-linear manner. The combination of 2 feature maps realizes the derivation of specific features regardless of the pixel size and generates more features that improve the classification performance. As the visualization results depicted, the BCNN concentrated on the buccal bone wall, rather than concerning the whole tooth as conventional CNNs. The activation analysis also confirmed that BCNN paid more attention to the region of interest under the truly predicted conditions and less attention under the falsely predicted condition. This means that the model makes the prediction based on the identification and extraction of the buccal bone wall region rather than irrelevant features.

The BCNN in this study achieved comparable performance to that of the related work. Mastouri *et al.* used BCNN with 2 streams (VGG16 and VGG19) followed by a support vector machine (SVM) to classify the lung nodules on CT (512×512 pixels) into non-nodules, micro-nodules (<3 mm), and masses (≥3 mm). The accuracy of their study reached 0.9199 and the AUC reached 0.959 (25). Zhao *et al.* applied BCNN and fast BCNN to assess the wound depth and granulation tissue amount from pictures of diabetes patients and achieved the best accuracy of 0.846 (26). Huang *et al.* developed the bilinear MobileNet-V3 model to diagnose breast cancer from histopathological slide images and achieved an accuracy of 0.88 (27). In this study, the top-performing model (BCNN-VGG16) achieved an accuracy of 0.870 (95% CI: 0.838–0.902), an AUC of 0.924 (95% CI: 0.896–0.948), and an AUPRC of 0.859 (95% CI: 0.803–0.903). The potential reasons for the advanced performance may be that the backbones of BCNN in this study were fully finetuned, which possessed better generalization ability and outperformed the transfer learning methods (28). Besides, the VGG backbones may have outperformed the advanced CNN backbones due to its simple and straightforward architecture that facilitates its ability in fine-grained and

local feature representation.

The significant potential for the clinical application of BCNN

The general performance of BCNN is equivalent to expert-level with higher efficiency. Recently, a meta-analysis depicted that the deep learning models possess higher sensitivity and specificity than doctors in medical images-derived diagnosis (29). In this study, the BCNN-VGG16 achieved significantly higher accuracy (0.87 *vs.* 0.80), precision (0.84 *vs.* 0.69), specificity (0.94 *vs.* 0.87), and F1 score (0.77 *vs.* 0.65) compared to all levels of dental practitioners, whereas its time-cost was thousands of times less than that of humans (*Tables 2,3*). In the dental scenario, 38.9% and 47.7% of patients in the 55–64 and 65–74 years groups in China, respectively, have unrestored missing teeth (30), and 96% of the elderly (≥60 years) have missing teeth in southern Vietnam (31). The BCNN may accelerate the clinical workflow and reduce numerous time-associated costs in the mechanically repeated medical image analysis on so many patients.

Another interesting finding is that human-machine fusion is promising in improving the dentist's diagnostic ability, especially towards the thick buccal bone type. The fusion profoundly increased the sensitivity from 0.62 to 0.80 maximumly, indicating that the machine rescued many missed thick samples by the human. The slight decrease in specificity (from 0.87 to 0.85) hinted that the fusion did not hazard prediction towards thin samples. What should be noticed is that the human-machine fusion narrowed the difference caused by individual and empirical factors and precisely produced consistent results regardless of the operators. This would be helpful for numerous junior and basic dentists to reduce the disparities and limit the risk of misdiagnosis.

Interestingly, the BCNN-VGG16 obtained high specificity (0.94) yet somehow inferior sensitivity (0.70), which was slightly lower than the mean sensitivity of dental practitioners (0.73). The imbalance between sensitivity and specificity resulted from the trade-off between sensitivity and specificity and the imbalance of thin and thick types (around 2.4:1), which can be adjusted through the loss reweight method. The increased ratio of loss weight from 1:1 to 1:3 led to the increase of sensitivity (from 0.719 to 0.749) and decrease of specificity (from 0.944 to 0.922) with the minor cost of AUC (from 0.925 to 0.920) (Table S7). From the clinical aspect, the higher specificity is helpful in the exclusion of the thin buccal bone type which is not suitable for immediate implant placement. The dentist can take advantage of the loss reweight method to adjust the preference towards the thin or thick type according to the clinical demand.

Implications of BCNN-based micro-scale structure analysis and limitations

The success of BCNN-based buccal bone wall classification hints that the fine-grained image recognition network has the potential in quantitative analysis of other important millimeter-to-micron structures, including small blood arteries abnormities (internal diameter smaller than 100 μm) (32), small bronchiole (internal diameter is less than 2 mm) (33), and so forth. Other kinds of medical images such as X-ray, ultrasound, and magnetic resonance imaging (MRI), are able to be quantitatively analyzed regardless of whether it is of high- or low-resolution. However, this study attempted to turn a regression task into a classification task. In the future, more deep learning methods can be explored in the regression-based micro-scale structure analysis to achieve better performance.

Conclusions

We succeeded in applying BCNN to the quantitative analysis of the binary thickness of the buccal bone wall and yielded clinically acceptable performance with high specificity. The BCNN precisely located the buccal bone wall region and identified the subtle difference in its thickness. The model outperformed the human experts generally whereas the diagnostic ability of all levels of dentists improved with the assistance of BCNN. This study took the automatic classification of buccal bone as a paradigm and laid a stepping-stone to the quantitative micro-scale structure analysis of low-resolution medical images.

Acknowledgments

The authors wish to thank Shangyou Wen, Jiayu Li, Jiaxin Xie, Wei Wei, Chenghao Zhang, Xinyi Yang, and Diallo Mariama from Guanghua School of Stomatology, Sun Yat-sen University for their contribution in data collection.

Funding: This project was supported by the Guangzhou Science and Technology Project (No. 2023B03J1232), Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation (Nos. pdjh2021b0013 and pdjh2023b0013), and the National Undergraduate Training Program for Innovation and

Entrepreneurship (Nos. 20212001 and 202210780).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-744/coif). The authors have no conflicts of interest.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and received ethical approval (No. KQEC-2020-29-04) from the Medical Ethics Committee of the Hospital of Stomatology of Sun Yat-sen University. The requirement of informed consent from participants was exempted due to the retrospective nature of the study.

References

1. Tsigarida A, Toscano J, de Brito Bezerra B, Geminiani A, Barmak AB, Caton J, Papaspyridakos P, Chochlidakis K. Buccal bone thickness of maxillary anterior teeth: A systematic review and meta-analysis. J Clin Periodontol 2020;47:1326-43.
2. Ferrus J, Cecchinato D, Pjetursson EB, Lang NP, Sanz M, Lindhe J. Factors influencing ridge alterations following immediate implant placement into extraction sockets. Clin Oral Implants Res 2010;21:22-9.
3. Chappuis V, Araújo MG, Buser D. Clinical relevance of dimensional bone and soft tissue alterations post-extraction in esthetic sites. Periodontol 2000 2017;73:73-83.
4. Morton D, Chen ST, Martin WC, Levine RA, Buser D. Consensus statements and recommended clinical procedures regarding optimizing esthetic outcomes in implant dentistry. Int J Oral Maxillofac Implants 2014;29 Suppl:216-20.
5. Rojo-Sanchis J, Soto-Peñaloza D, Peñarrocha-Oltra D, Peñarrocha-Diago M, Viña-Almunia J. Facial alveolar bone thickness and modifying factors of anterior maxillary teeth: a systematic review and meta-analysis of cone-beam computed tomography studies. BMC Oral Health 2021;21:143.
6. Westheimer G. Visual acuity: information theory, retinal image structure and resolution thresholds. Prog Retin Eye Res 2009;28:178-86.
7. Zhang LJ, Wang Y, Schoepf UJ, Meinel FG, Bayer RR 2nd, Qi L, Cao J, Zhou CS, Zhao YE, Li X, Gong JB, Jin Z, Lu GM. Image quality, radiation dose, and diagnostic accuracy of prospectively ECG-triggered high-pitch coronary CT angiography at 70 kVp in a clinical setting: comparison with invasive coronary angiography. Eur Radiol 2016;26:797-806.
8. Wang G, Liu X, Shen J, Wang C, Li Z, Ye L, et al. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images. Nat Biomed Eng 2021;5:509-21.
9. Yue W, Zhang H, Zhou J, Li G, Tang Z, Sun Z, Cai J, Tian N, Gao S, Dong J, Liu Y, Bai X, Sheng F. Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging. Front Oncol 2022;12:984626.
10. Huang Y, Bert C, Sommer P, Frey B, Gaipl U, Distel LV, Weissmann T, Uder M, Schmidt MA, Dörfler A, Maier A, Fietkau R, Putz F. Deep learning for brain metastasis detection and segmentation in longitudinal MRI data. Med Phys 2022;49:5773-86.
11. Wei XS, Song YZ, Aodha OM, Wu J, Peng Y, Tang J, Yang J, Belongie S. Fine-Grained Image Analysis With Deep Learning: A Survey. IEEE Trans Pattern Anal Mach Intell 2022;44:8927-48.
12. Lin TY, RoyChowdhury A, Maji S. Bilinear CNN Models for Fine-Grained Visual Recognition. 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE; 2015:1449-57.
13. Brauwers G, Frasincar F. A General Survey on Attention Mechanisms in Deep Learning. IEEE Trans Knowl Data Eng 2023;35:3279-98.
14. Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc.; 2017:3859-69.
15. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-

Excitation Networks. IEEE Trans Pattern Anal Mach Intell 2020;42:2011-23.
16. Askar H, Krois J, Rohrer C, Mertens S, Elhennawy K, Ottolenghi L, Mazur M, Paris S, Schwendicke F. Detecting white spot lesions on dental photography using deep learning: A pilot study. J Dent 2021;107:103615.
17. Carreira J, Caseiro R, Batista J, Sminchisescu C. Semantic Segmentation with Second-Order Pooling. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C. editors. Computer Vision – ECCV 2012. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2012.
18. de Vries BM, Zwezerijnen GJC, Burchell GL, van Velden FHP, Menke-van der Houven van Oordt CW, Boellaard R. Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review. Front Med (Lausanne) 2023;10:1180773.
19. Zheng X, Wang R, Zhang X, Sun Y, Zhang H, Zhao Z, et al. A deep learning model and human-machine fusion for prediction of EBV-associated gastric cancer from histopathology. Nat Commun 2022;13:2790.
20. Lin TY, RoyChowdhury A, Maji S. Bilinear Convolutional Neural Networks for Fine-grained Visual Recognition. IEEE Trans Pattern Anal Mach Intell 2018;40:1309-22.
21. Liu W, Juhas M, Zhang Y. Fine-Grained Breast Cancer Classification With Bilinear Convolutional Neural Networks (BCNNs). Front Genet 2020;11:547327.
22. Hou X, Xu X, Zhao M, Kong J, Wang M, Lee ES, Jia Q, Jiang HB. An overview of three-dimensional imaging devices in dentistry. J Esthet Restor Dent 2022;34:1179-96.
23. Oyama A, Kumagai S, Arai N, Takata T, Saikawa Y, Shiraishi K, Kobayashi T, Kotoku J. Image quality improvement in cone-beam CT using the super-resolution technique. J Radiat Res 2018;59:501-10.
24. Liu J, Yan H, Cheng H, Liu J, Sun P, Wang B, Mao R, Du C, Luo S. CBCT-based synthetic CT generation using generative adversarial networks with disentangled representation. Quant Imaging Med Surg 2021;11:4820-34.
25. Mastouri R, Khlifa N, Neji H, Hantous-Zannad S. A bilinear convolutional neural network for lung nodules classification on CT images. Int J Comput Assist Radiol Surg 2021;16:91-101.
26. Zhao X, Liu Z, Agu E, Wagh A, Jain S, Lindsay C, Tulu B, Strong D, Kan J. Fine-grained diabetic wound depth and granulation tissue amount assessment using bilinear convolutional neural network. IEEE Access 2019;7:179151-62.
27. Huang J, Mei L, Long M, Liu Y, Sun W, Li X, Shen H, Zhou F, Ruan X, Wang D, Wang S, Hu T, Lei C. BM-Net: CNN-Based MobileNet-V3 and Bilinear Structure for Breast Cancer Detection in Whole Slide Images. Bioengineering (Basel) 2022;9:261.
28. Mastouri R, Khlifa N, Neji H, Hantous-Zannad S. Transfer Learning Vs. Fine-Tuning in Bilinear CNN for Lung Nodules Classification on CT Scans. Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition 2020:99-103.
29. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019;1:e271-97.
30. Sun H, Du M, Tai B, Chang S, Wang Y, Jiang H. Prevalence and associated factors of periodontal conditions among 55- to 74-year-old adults in China: results from the 4th National Oral Health Survey. Clin Oral Investig 2020;24:4403-12.
31. Nguyen TC, Witter DJ, Bronkhorst EM, Truong NB, Creugers NH. Oral health status of adults in Southern Vietnam - a cross-sectional epidemiological study. BMC Oral Health 2010;10:2.
32. Bosetti F, Galis ZS, Bynoe MS, Charette M, Cipolla MJ, Del Zoppo GJ, Gould D, Hatsukami TS, Jones TL, Koenig JI, Lutty GA, Maric-Bilkan C, Stevens T, Tolunay HE, Koroshetz W; "Small Blood Vessels: Big Health Problems" Workshop Participants. "Small Blood Vessels: Big Health Problems?": Scientific Recommendations of the National Institutes of Health Workshop. J Am Heart Assoc 2016;5:e004389.
33. Ikezoe K, Hackett TL, Peterson S, Prins D, Hague CJ, Murphy D, LeDoux S, Chu F, Xu F, Cooper JD, Tanabe N, Ryerson CJ, Paré PD, Coxson HO, Colby TV, Hogg JC, Vasilescu DM. Small Airway Reduction and Fibrosis Is an Early Pathologic Feature of Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med 2021;204:1048-59.

Cite this article as: Gong Z, Li X, Shi M, Cai G, Chen S, Ye Z, Gan X, Yang R, Wang R, Chen Z. Measuring the binary thickness of buccal bone of anterior maxilla in low-resolution cone-beam computed tomography via a bilinear convolutional neural network. Quant Imaging Med Surg 2023;13(12):8053-8066. doi: 10.21037/qims-23-744

Supplementary

Table S1 ICC between the two researchers

Comparisons	ICC
Two measurements of researcher 1	0.943
Two measurements of researcher 2	0.954
Measurements between researcher 1 and researcher 2	0.899

ICC, intra-class correlation.

Table S2 The human-machine fusion result using the 'Or' strategy

Predictions	Dentist_Thick	Dentist_Thin
BCNN_Thick	Thick	Thick
BCNN_Thin	Thick	Thin

The fusion result is listed in the cells of this table. Dentist_Thick/Dentist_Thin: the sample is considered thick/thin by the dentist; BCNN_Thick/BCNN_Thin: the sample is considered thick/thin by the BCNN. BCNN, bilinear convolutional neural network.

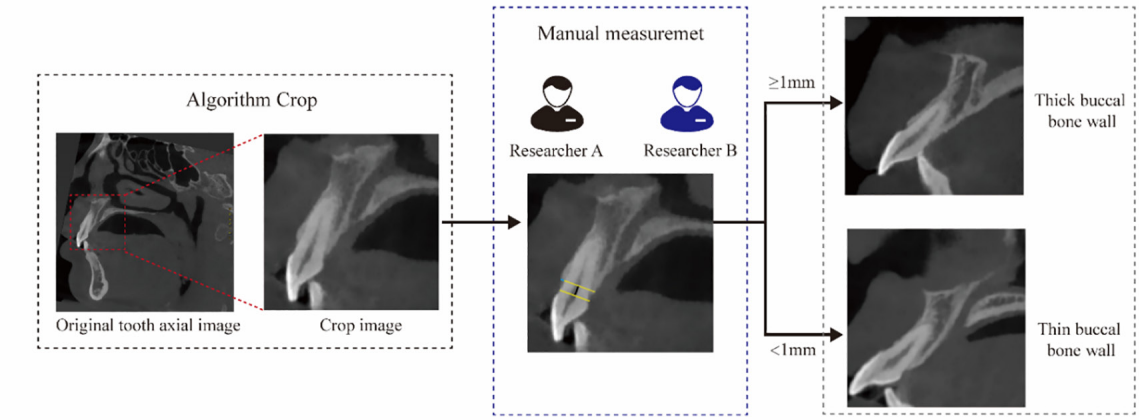


Figure S1 The manual measurement and classification process of buccal bone wall.

Table S3 The performance of BCNN with different backbones in the training cohort

Model	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 score (95% CI)	AUC (95% CI)	AUPRC (95% CI)
BCNN-VGG16	0.869 (0.855, 0.883)	0.817 (0.787, 0.846)	0.713 (0.677, 0.749)	0.933 (0.920, 0.946)	0.760 (0.734, 0.787)	0.916 (0.903, 0.929)	0.833 (0.806, 0.858)
BCNN-Resnet18	0.859* (0.844, 0.872)	0.780* (0.749, 0.810)	0.723* (0.687, 0.758)	0.915* (0.901, 0.929)	0.749* (0.719, 0.773)	0.906* (0.891, 0.920)	0.812* (0.782, 0.839)
BCNN-Resnet34	0.856* (0.840, 0.871)	0.800* (0.765, 0.832)	0.683* (0.642, 0.721)	0.927* (0.913, 0.941)	0.735* (0.703, 0.764)	0.896* (0.880, 0.912)	0.803* (0.774, 0.830)
BCNN-Resnet50	0.858* (0.844, 0.873)	0.778* (0.748, 0.810)	0.721* (0.685, 0.756)	0.914* (0.899, 0.929)	0.747* (0.721, 0.774)	0.902* (0.887, 0.918)	0.803* (0.771, 0.836)
BCNN-Resnet101	0.857* (0.844, 0.872)	0.799* (0.769, 0.830)	0.692* (0.654, 0.728)	0.925* (0.912, 0.937)	0.736* (0.708, 0.763)	0.906* (0.891, 0.920)	0.815* (0.786, 0.844)
BCNN-ResNeXt50	0.851* (0.837, 0.866)	0.795* (0.763, 0.829)	0.667* (0.628, 0.704)	0.929* (0.915, 0.942)	0.722* (0.693, 0.752)	0.902* (0.887, 0.916)	0.803* (0.770, 0.834)

*, the result of BCNN-VGG16 is statistically significantly different to the result of all contrast models with *t*-test P<0.05. BCNN, bilinear convolutional neural network; CI, confidence interval; AUC, area under the receiver operator characteristic curve; AUPRC, area under the precision-recall curve; BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group; BCNN-Resnet18, bilinear convolutional neural network with Resnet18 as its backbone; BCNN-Resnet34, bilinear convolutional neural network with Resnet34 as its backbone; BCNN-Resnet50, bilinear convolutional neural network with Resnet50 as its backbone; BCNN-Resnet101, bilinear convolutional neural network with Resnet101 as its backbone; BCNN-ResNeXt50, bilinear convolutional neural network with ResNeXt50 as its backbone.

https://dx.doi.org/10.21037/qims-23-744

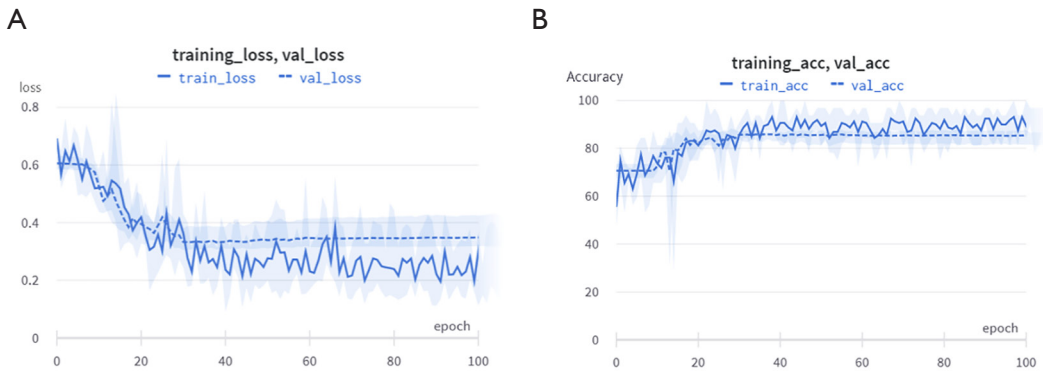


Figure S2 The convergence plot of loss (A) and accuracy (B) of the BCNN with VGG16 backbone in the training cohort. BCNN, bilinear convolutional neural network; VGG, visual geometry group.

Table S4 The parameters and the FLOPs of the top-performing model and traditional CNN model

Models	Parameters (MB)	FLOPs (GFLOPs)
BCNN-VGG16	15.2	15.38
VGG16	13.5	15.53

FLOPs, floating-point operations; CNN, convolutional neural network; BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group.

Table S5 The performance of BCNN-VGG16 and CNNs in the training cohort

Model	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 score (95% CI)	AUC (95% CI)	AUPRC (95% CI)
BCNN-VGG16	0.869 (0.855, 0.883)	0.817 (0.787, 0.846)	0.713 (0.677, 0.749)	0.933 (0.920, 0.946)	0.760 (0.734, 0.787)	0.916 (0.903, 0.929)	0.833 (0.806, 0.858)
VGG16	0.852* (0.837, 0.868)	0.776* (0.745, 0.810)	0.706* (0.669, 0.741)	0.912* (0.897, 0.928)	0.737* (0.707, 0.764)	0.903* (0.888, 0.917)	0.819* (0.789, 0.844)
Resnet18	0.852* (0.838, 0.867)	0.787* (0.757, 0.817)	0.694* (0.656, 0.731)	0.919* (0.905, 0.932)	0.729* (0.701, 0.757)	0.908* (0.894, 0.921)	0.821* (0.793, 0.847)
Resnet34	0.856* (0.842, 0.871)	0.784* (0.752, 0.814)	0.705* (0.669, 0.740)	0.918* (0.904, 0.932)	0.741* (0.714, 0.769)	0.895* (0.879, 0.910)	0.808* (0.781, 0.835)
Resnet50	0.852* (0.837, 0.867)	0.782* (0.749, 0.815)	0.689* (0.651, 0.724)	0.920* (0.904, 0.933)	0.731* (0.704, 0.760)	0.904* (0.890, 0.918)	0.816* (0.788, 0.841)
Resnet101	0.845* (0.829, 0.860)	0.772* (0.739, 0.804)	0.671* (0.632, 0.712)	0.917* (0.902, 0.931)	0.716* (0.684, 0.745)	0.894* (0.879, 0.910)	0.791* (0.759, 0.820)
ResNeXt50	0.853* (0.838, 0.867)	0.76* (0.731, 0.792)	0.730* (0.694, 0.764)	0.904* (0.888, 0.920)	0.744* (0.716, 0.770)	0.903* (0.889, 0.917)	0.805* (0.776, 0.833)

*, the result of BCNN-VGG16 is statistically significantly different to the result of all contrast models with *t*-test P<0.05. BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group; CNN, convolutional neural network; CI, confidence interval; AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve.

https://dx.doi.org/10.21037/qims-23-744

Table S6 The model performance with different folds of cross validation in the test cohort

Number of folds	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 score (95% CI)	AUC (95% CI)	AUPRC (95% CI)
5-fold	0.870 (0.838, 0.902)	0.843 (0.776, 0.906)	0.701 (0.617, 0.783)	0.943 (0.914, 0.968)	0.765 (0.700, 0.825)	0.924 (0.896, 0.948)	0.859 (0.803, 0.903)
8-fold	0.863 (0.832, 0.892)	0.819 (0.752, 0.887)	0.701 (0.617, 0.783)	0.933 (0.900, 0.961)	0.755 (0.690, 0.814)	0.922 (0.894, 0.948)	0.858 (0.801, 0.905)
10-fold	0.860 (0.828, 0.890)	0.794 (0.726, 0.863)	0.726 (0.650, 0.808)	0.919 (0.886, 0.950)	0.758 (0.695, 0.814)	0.922 (0.893, 0.947)	0.855 (0.797, 0.902)

CI, confidence interval; AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve.

Table S7 The model performance with different loss reweight ratio in the test cohort

Loss reweight ratio	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 score (95% CI)	AUC (95% CI)	AUPRC (95% CI)
1:1	0.875 (0.843, 0.905)	0.846 (0.784, 0.909)	0.719 (0.633, 0.800)	0.944 (0.914, 0.968)	0.776 (0.715, 0.830)	0.925 (0.897, 0.950)	0.860 (0.805, 0.905)
1:1.5	0.868 (0.835, 0.900)	0.803 (0.737, 0.867)	0.742 (0.667, 0.817)	0.922 (0.889, 0.950)	0.771 (0.712, 0.829)	0.920 (0.887, 0.948)	0.855 (0.800, 0.905)
1:2	0.869 (0.835, 0.900)	0.799 (0.732, 0.865)	0.747 (0.675, 0.817)	0.919 (0.886, 0.950)	0.772 (0.715, 0.828)	0.922 (0.892, 0.949)	0.863 (0.814, 0.907)
1:2.5	0.865 (0.835, 0.895)	0.779 (0.714, 0.845)	0.765 (0.683, 0.842)	0.907 (0.871, 0.936)	0.772 (0.713, 0.827)	0.921 (0.892, 0.950)	0.858 (0.809, 0.902)
1:3	0.870 (0.838, 0.900)	0.805 (0.741, 0.871)	0.749 (0.667, 0.825)	0.922 (0.889, 0.954)	0.775 (0.716, 0.833)	0.920 (0.892, 0.947)	0.861 (0.809, 0.908)

CI, confidence interval; AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve.

Table S8 The misclassified samples and the potential reasons by BCNN-VGG16 in the test cohort

Misclassified categories	Potential reasons	Samples	Proportion
FP	Poor image quality (e.g., the artifacts, fuzziness)	9	45%
	Anatomic abnormity (e.g., the alveolar ridge is >2 mm away from the CEJ)	8	40%
	Genius thickness of buccal bone around the cut point of the binary classifications (i.e., the thickness of 1 mm)	3	15%
	Total	20	100%
FN	Poor image quality (e.g., the artifacts, fuzziness)	14	45.2%
	Anatomic abnormity (e.g., opacity lesions)	5	16.1%
	Genius thickness of buccal bone around the cut point of the binary classifications (i.e., the thickness of 1 mm)	12	38.7%
	Total	31	100.0%

BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group; FP, false positive; CEJ, cementoenamel junction; FN, false negative.

https://dx.doi.org/10.21037/qims-23-744

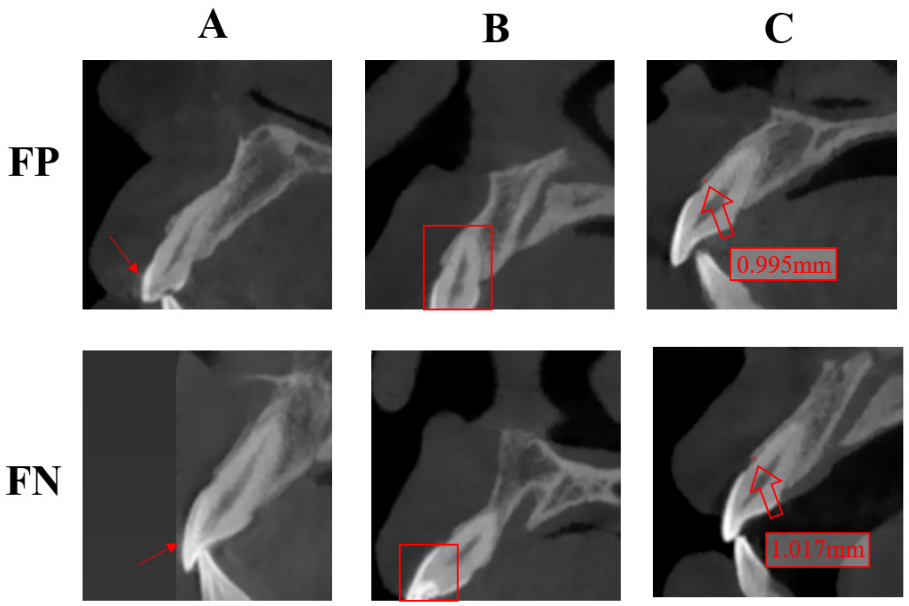
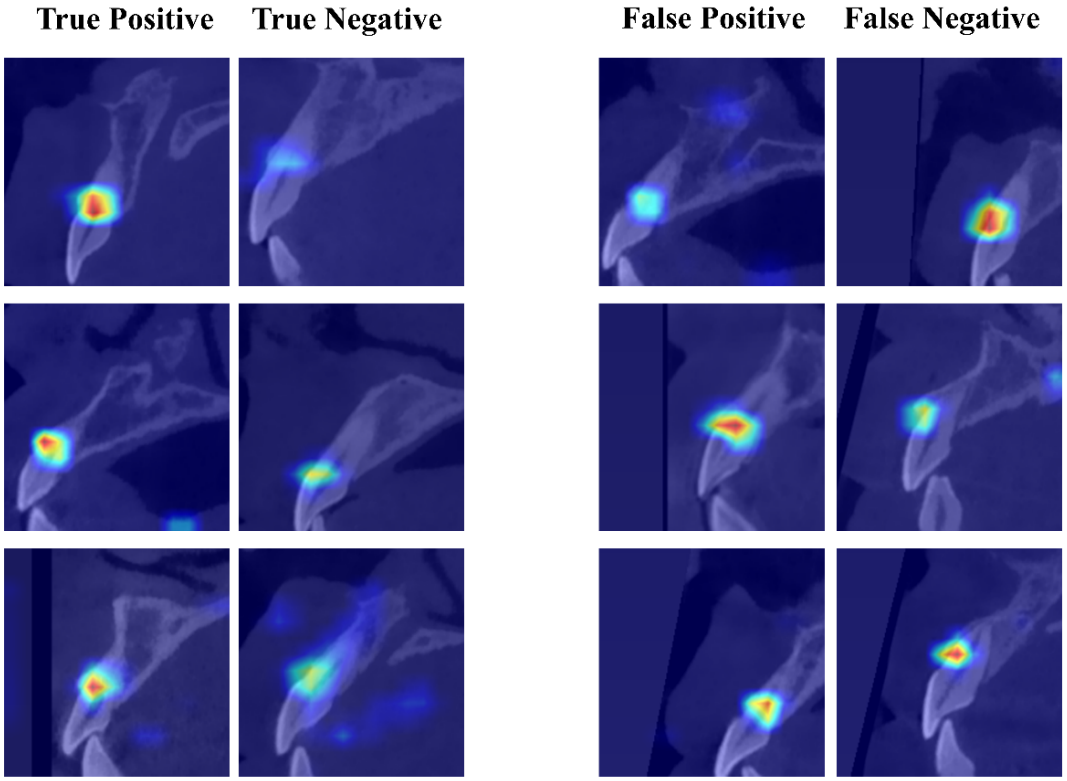


Figure S3 The typical cases of misclassified samples by BCNN-VGG16 in the test cohort, including the FP and FN samples. The column A represent the poor image quality (i.e., the fuzziness and artifacts depicted by the arrow), the column B represent the anatomic abnormity (i.e., the incomplete crown and opacity lesion, depicted by the red box), the column C represent the confusing samples which is near the cutting point (i.e., its thickness is close to 1 mm). FP, false positive; FN, false negative; BCNN-VGG16, bilinear convolutional neural network with VGG16 as its backbone; VGG, visual geometry group.

https://dx.doi.org/10.21037/qims-23-744

A



B

Category	True Positive	True Negative	False Positive	False Negative
Normalized activation score	0.012	0.007	0.013	0.011

Figure S4 The qualitative (A) and quantitative (B) activation analysis between different categories predicted by BCNN. BCNN, bilinear convolutional neural network.

https://dx.doi.org/10.21037/qims-23-744