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.
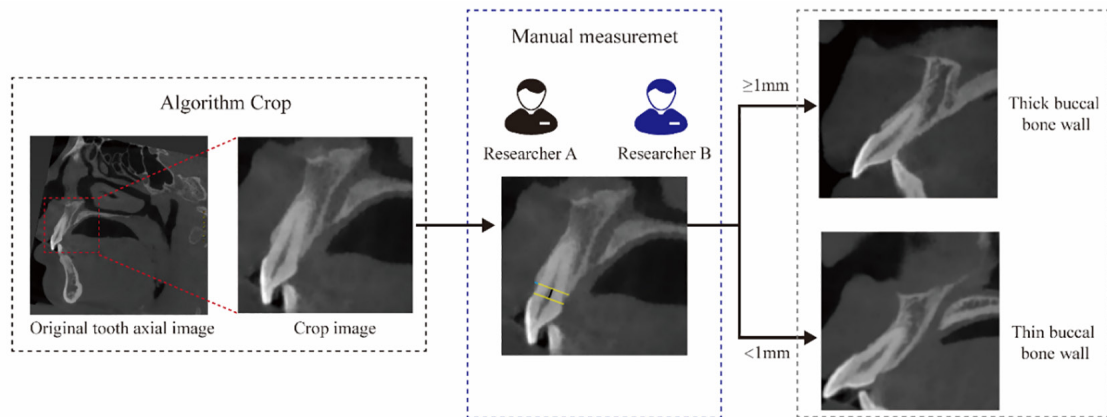

Figure S1 The manual measurement and classification process of buccal bone wall.

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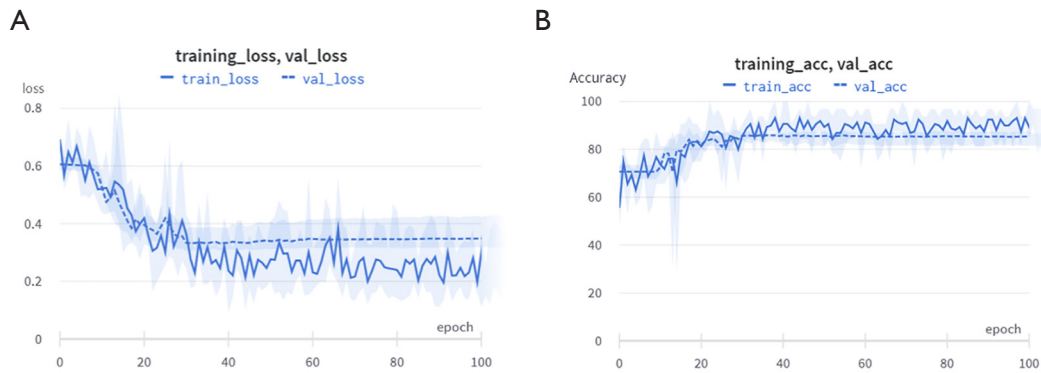
Figure S2 The convergence plot of loss (A) and accuracy (B) of the BCNN with VGG16 backbone in the training cohort. BCNN, bilinear convolutional neural network; VGG, visual geometry group.

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.

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.
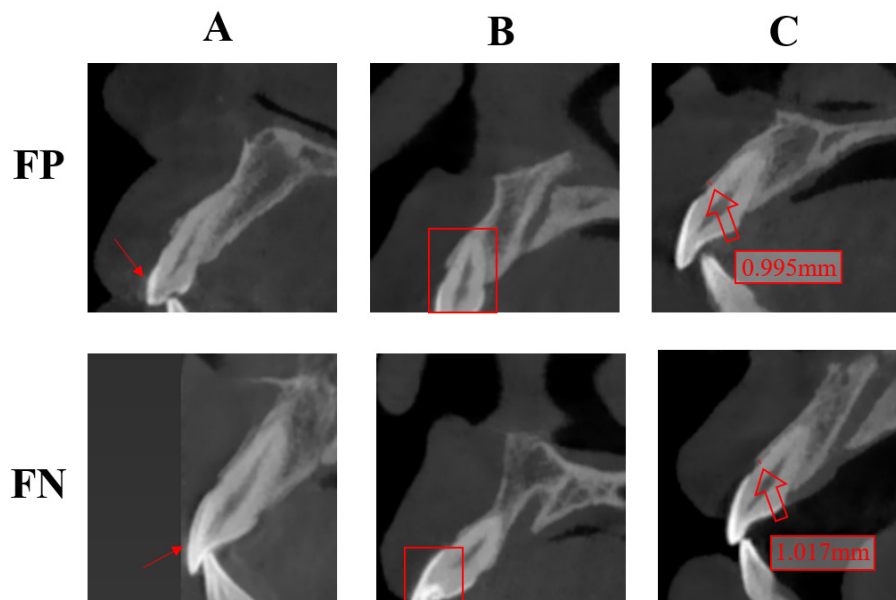
Figure S3 The typical cases of misclassified samples by BCNN-VGG16 in the test cohort, including the FP and FN samples. The column A represent the poor image quality (i.e., the fuzziness and artifacts depicted by the arrow), the column B represent the anatomic abnormity (i.e., the incomplete crown and 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.

A

| | True Positive | True Negative | | False Positive | False Negative |
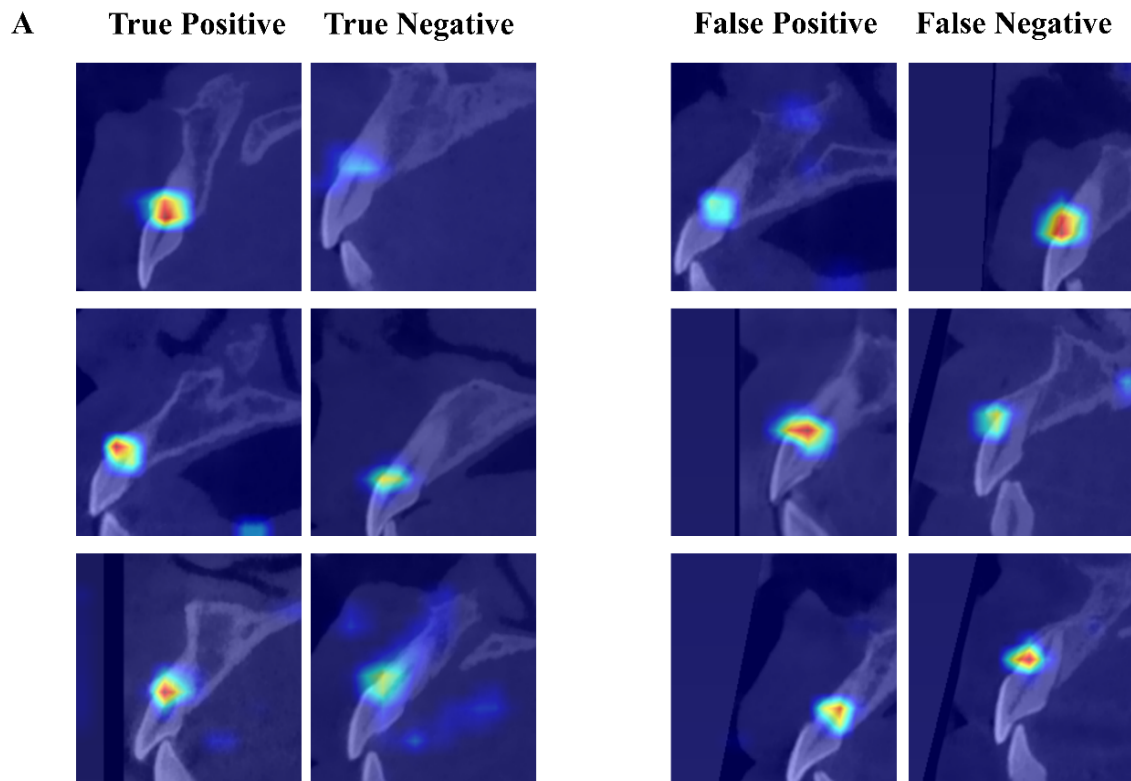
B

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.