


# An artificial intelligence grading system of apical periodontitis in cone-beam computed tomography data

Tianyin Zhao, MD<sup>1,2,3,4,†</sup>, Huili Wu, MD<sup>1,2,3,4,†</sup>, Diya Leng, MD<sup>1,2,3,4</sup>, Enhui Yao, MD<sup>1,2,3,4</sup>, Shuyun Gu, MD<sup>1,2,3,4</sup>, Minhui Yao, MD<sup>1,2,3,4</sup>, Qinyu Zhang, MD<sup>1,2,3,4</sup>, Tong Wang, MD<sup>1,2,3,4</sup>, Daming Wu , PhD<sup>1,2,3,4,\*</sup>, Lizhe Xie, PhD<sup>2,3,4,\*</sup>

<sup>1</sup>Department of Endodontics, The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing, 210029, China

<sup>2</sup>Department of Oral & Maxillofacial Imaging, The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing, 210029, China

<sup>3</sup>Jiangsu Province Key Laboratory of Oral Diseases, Nanjing, 210029, China

<sup>4</sup>Jiangsu Province Engineering Research Center of Stomatological Translational Medicine, Nanjing, 210029, China

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: Daming Wu, PhD, Department of Endodontics, The Affiliated Stomatological Hospital of Nanjing Medical University, 1 Shanghai Road, Nanjing, 210029, People's Republic of China (wdming@njmu.edu.cn) and Lizhe Xie, PhD, Department of Oral & Maxillofacial Imaging, The Affiliated Stomatological Hospital of Nanjing Medical University, 1 Shanghai Road, Nanjing, 210029, People's Republic of China (xielizhe@njmu.edu.cn)

## Abstract

**Objectives:** In order to assist junior doctors in better diagnosing apical periodontitis (AP), an artificial intelligence AP grading system was developed based on deep learning (DL) and its reliability and accuracy were evaluated.

**Methods:** One hundred and twenty cone-beam computed tomography (CBCT) images were selected to construct a classification dataset with four categories, which were divided by CBCT periapical index (CBCTPAI), including normal periapical tissue, CBCTPAI 1-2, CBCTPAI 3-5, and young permanent teeth. Three classic algorithms (ResNet50/101/152) as well as one self-invented algorithm (PAINet) were compared with each other. PAINet were also compared with two recent Transformer-based models and three attention models. Their performance was evaluated by accuracy, precision, recall, balanced F score (F1-score), and the area under the macro-average receiver operating curve (AUC). Reliability was evaluated by Cohen's kappa to compare the consistency of model predicted labels with expert opinions.

**Results:** PAINet performed best among the four algorithms. The accuracy, precision, recall, F1-score, and AUC on the test set were 0.9333, 0.9415, 0.9333, 0.9336, and 0.9972, respectively. Cohen's kappa was 0.911, which represented almost perfect consistency.

**Conclusions:** PAINet can accurately distinguish between normal periapical tissues, CBCTPAI 1-2, CBCTPAI 3-5, and young permanent teeth. Its results were highly consistent with expert opinions. It can help junior doctors diagnose and score AP, reducing the burden. It can also be promoted in areas where experts are lacking to provide professional diagnostic opinions.

**Keywords:** apical periodontitis; classification diagnosis; cone-beam CT; deep learning; periapical tissue.

## Introduction

The most common cause of apical periodontitis (AP) is the uncontrolled infection in the root canal system,<sup>1,2</sup> which appears on radiographs as a radiolucency around the root tip.<sup>3,4</sup> It has been reported that AP is prevalent in adults worldwide, with 52% having at least one tooth with a periapical lesion,<sup>5</sup> and highest prevalence in the middle-aged ones.<sup>6</sup> Radiology is an important adjunctive diagnostic tool in clinic. However, traditional 2D images have many disadvantages compared with cone-beam computed tomography (CBCT), such as superimposition of anatomical structures, geometrical distortion, anatomical noise, and indeed of their 2D nature.<sup>3,7,8</sup> Junior doctors often make incorrect diagnoses when reading periapical radiography films. According to experimental reports, when diagnosing AP, the accuracy rate of junior doctors when interpreting periapical radiography films for the first time was only 66.75–71.75%.<sup>9</sup> When interpreting panoramic radiographs, the accuracy rate was 68.7–71.3%.<sup>10</sup> CBCT is the only method that provides early and accurate detection of all periapical lesions in the jaws, with the lowest risk of false positives.<sup>3</sup>

In the studies of AP, the periapical index (PAI) devised by Brynolf in 1967 is commonly used as a measure to assess AP severity.<sup>11–13</sup> It helps researchers observe the root apex through periapical radiography films from a 2D perspective. However, the process of PAI scoring is time-consuming and requires trained manpower. Further, the calculated average kappa values were only in the fair to moderate range of agreement.<sup>14,15</sup> In 2008, Estrela proposed cone-beam computed tomography PAI (CBCTPAI), which can increase the reliability of epidemiologic studies, especially AP prevalence and severity, compared with PAI conducted in 1967.<sup>4</sup> CBCTPAI provides a 6-point (0–5) scoring system ranging from health to severe AP based on the interpretation of 3D radiographic images.<sup>16</sup>

The integration of artificial intelligence (AI) and the medical field has deepened in recent years.<sup>17–19</sup> Deep learning (DL), a sub-discipline of machine learning, has quickly become a widely used tool for AI medical image analysis due to its regular self-learning approach, which is easy to automate. DL has been used to diagnose AP on periapical radiography films, panoramic radiographs, and CBCT data. Li et al

developed a DL model for automatic diagnosis of caries and periapical disease on periapical radiography films.<sup>9</sup> Song et al developed a segmentation model that can detect periapical lesions on panoramic radiographs.<sup>20</sup> Kirnbauer et al developed a DL model that first segmented individual teeth from the dental arch and then performed object detection of periapical lesions in CBCT data.<sup>21</sup> Although these studies diagnosed AP in different modalities, they lacked a score for the severity of AP. Some researchers have also linked periapical radiography films with PAI by DL models to achieve automatic PAI scoring.<sup>14</sup> However, due to the inherent flaws of PAI's 2D data, AI models developed based on manually annotated data will also suffer from the problem of high false negatives. Few researchers have performed AI automated scoring of AP based on CBCTPAI.

Furthermore, it should be noted that young permanent teeth with underdeveloped roots and wide pulp chambers exhibit localized low-density imaging around the root tips. The restricted hypointense translucency is a normal characteristic of the papilla during the development of the roots of young permanent teeth, but it is similar to the radiolucency of AP in CBCT images. It is possible for junior clinicians without professional training to misdiagnose this condition.

Therefore, the aim of this study was to develop an AP grading system by CBCTPAI based on DL and evaluate its reliability and accuracy. The hypothesis is that the AP grading system could assist junior doctors in better diagnosing AP and reduce the burden. It can also be promoted in areas where experts are lacking to provide professional diagnostic opinions. In particular, it would be a scoring system for AP rather than a single diagnosis.

## Methods

### Data collection

This study received ethical approval from the Ethics Review Committee of the Affiliated Stomatological Hospital of Nanjing Medical University (PJ2021-150-001). Patients have informed consent for their CBCT data to be used. CBCT data taken at the Department of Oral & Maxillofacial Imaging from March 2021 to March 2023 were collected. The CBCT data was acquired by NewTom VGi (QR SRL, Verona, Italy) with the following scanning parameters: 110 kVp, 8-15 mA, exposure time 3.6-4.8 s, voxel 0.2-0.3 mm, FOV of 16 × 18 cm or 15 × 12 cm. All CBCT examinations were taken by the same experienced radiologist and the data were saved in DICOM 3.0 format.

According to Estrela et al, CBCTPAI  $\geq 1$  was assessed as a periapical lesion. The larger the diameter of the lesion, the higher the score.<sup>4</sup> CBCTPAI 1-2 has diameter of periapical radiolucency  $< 2$  mm, which is too small to make no mistake. Therefore, CBCTPAI 1-2 was divided into a separate category from CBCTPAI 3-5. The diagnoses of young permanent teeth were based on the Nolla stage of tooth development,<sup>22</sup> and young permanent teeth with Nolla stage 8-9 were selected, which were simultaneously shown on CBCT images as apical around a restricted hypodense translucency (an image of the tooth papilla during normal root formation) with a continuous surrounding bone sclerite. Young permanent teeth with periapical lesions were not included in the study. In addition, unclear images, motion artefacts, severe

periodontitis such as expansion or destruction of periapical cortical bone, which is described as CBCTPAI n + E or CBCTPAI n + D ( $n = 2-5$ ) based on the manuscript published by Estrela in 2008,<sup>4</sup> root fractures, jawbone pathology, or previous surgical intervention were excluded. The criteria for inclusion and exclusion were described as:

- 1) The teeth divided into normal periapical (CBCTPAI 0) should be permanent teeth having intact periapical bone structures.
- 2) The teeth divided into CBCTPAI 1-2 should be permanent teeth having diameter of periapical radiolucency smaller than 2 mm.
- 3) The teeth divided into CBCTPAI 3-5 should be permanent teeth having diameter of periapical radiolucency larger than 2 mm.
- 4) The teeth divided into young permanent teeth should be young permanent teeth with Nolla stage 8-9, without periapical lesions.
- 5) Periapical lesions with expansion or destruction of periapical cortical bone, root fractures, jawbone pathology, or previous surgical intervention should be excluded.
- 6) Unclear images, motion artefacts, heavy metal artefacts should be excluded.
- 7) The root fillings and crowns were allowed to be exist in any categories.

A total of 120 CBCT data were collected, including 40 subjects for normal periapical, 30 subjects for CBCTPAI 1-2, 30 subjects for CBCTPAI 3-5, and 20 subjects for young permanent teeth.

### Construction and efficacy evaluation of diagnostic classification models for periapical imaging Image annotation

All CBCT data were desensitized (ie, all information about the patient was withheld). Image reconstruction was performed using ITK-Snap ([www.itksnap.org](http://www.itksnap.org)), adjusting for appropriate contrast, greyscale, and cut layer thickness. The sagittal slices reconstructed from raw data per category were used as the study object for constructing the dataset. Sagittal slices of the subjects' teeth were selected. The selected images need to include clear and complete images of the root and periapical area of the tooth, and the whole area of periapical lesion or papilla. No more than five images were used for a tooth. Regions of interest (ROI) of the selected images were annotated with the requirement that the limited field of view contained a complete image of the root and periapical tissues of a tooth. Then the images were saved in JPEG format. The acquisition and classification of the above image data were jointly performed by two professional radiologists after unified training within 2 weeks. In case of disagreement, an expert radiologist with extensive experience in the field made the decision. The final selection was 600 images from 120 subjects, encompassing 284 teeth.

All data were randomly divided into training set, validation set, and test set by 8:1:1 according to holdout cross validation. In dividing the dataset, it was ensured that several images of the same tooth existed only in one of the training set, validation set, and test set. This process can prevent false excellence in model performance caused by data leaks.<sup>23</sup> The

training set had 120 images in each category, the validation set had 15 images in each category, and the test set had 15 images in each category. The training set would be involved in the training process of the DL model, the validation set would not be involved in the training, with the purpose of monitoring the change of the model accuracy during the training process, and the test set would be used to evaluate the performance of the model after the training is completed.

### Data enhancement

Data enhancements were performed on the dataset using cv2 (opencv-python, <https://opencv.org>, version 4.6.0.66). The data enhancements for the training set included rotating 5° counterclockwise, brightening by 1.3 times, zooming by 1.5 times, and adding salt noise. The images generated from each type of data enhancement would participate in the next data enhancement process. The data enhancement methods taken for the validation set and test set included adding Gaussian noise, darkening by 0.9 times, and zooming by 2.0 times, and the images generated from each type of data enhancement would participate in the process of the next type of data enhancement. It is not a common practice to perform data enhancements on validation set and test set. If the amount of original data on the validation set and test set was relatively small, the observed accuracy change step would be relatively large. Data enhancements different from the training set were performed on the validation and test set in order to solve this problem. On the one hand, the change step of the observed accuracy became smaller. On the other hand, the validation set and test set became more different from the training set, which can test the generalization ability of the model. 7680 images of the training set, 480 images of the validation set, and 480 images of the test set were obtained after the data enhancements.

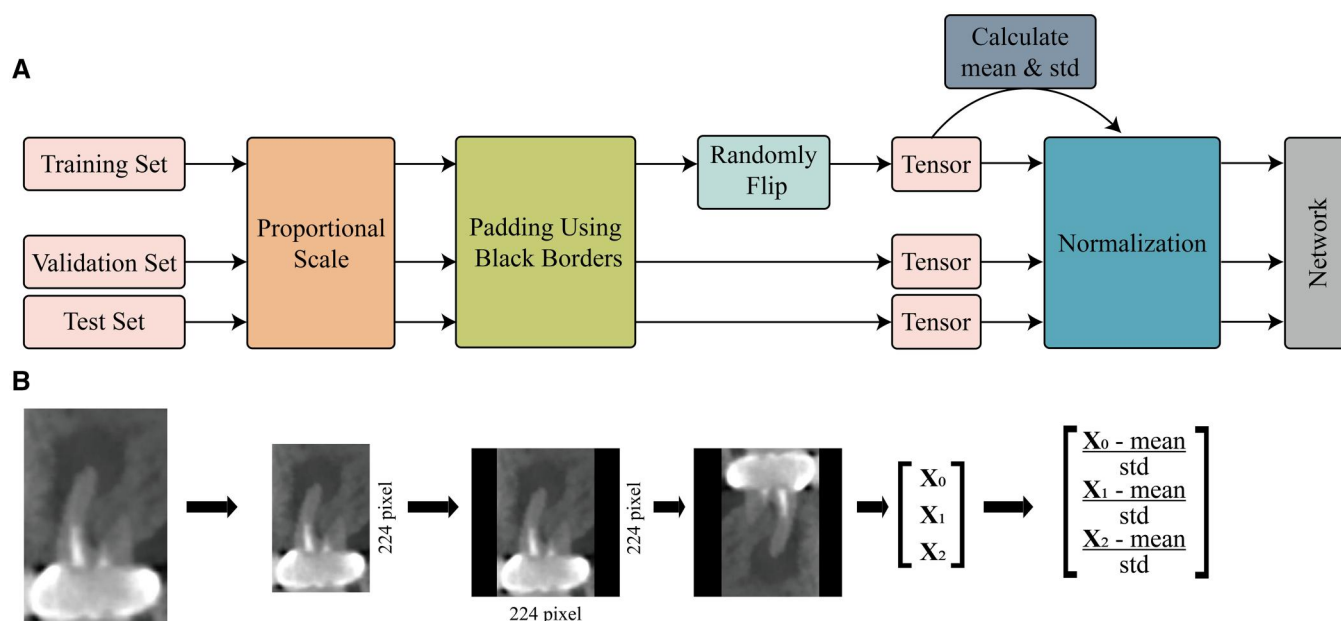
### Image pre-processing

All images were pre-processed before being passed into the network using the torchvision library (<https://pytorch.org>, version 0.14.1+cu116) with the PyTorch library (<https://pytorch.org>, version 1.13.1+cu116) (Figure 1A). All images would be scaled proportionally to ensure that the longer side became 224 pixels. Each image would then be placed in the centre of a 224 × 224 pixels black background. Images in the training set were randomly flipped horizontally and vertically. (Figure 1B) The images in the validation set and test set did not involve this process. All images were converted into tensors. The data in the training set were used to calculate the mean and standard deviation of each of the three RGB channels. Finally, all image data were normalized using the calculated mean and standard deviation before entering the network.

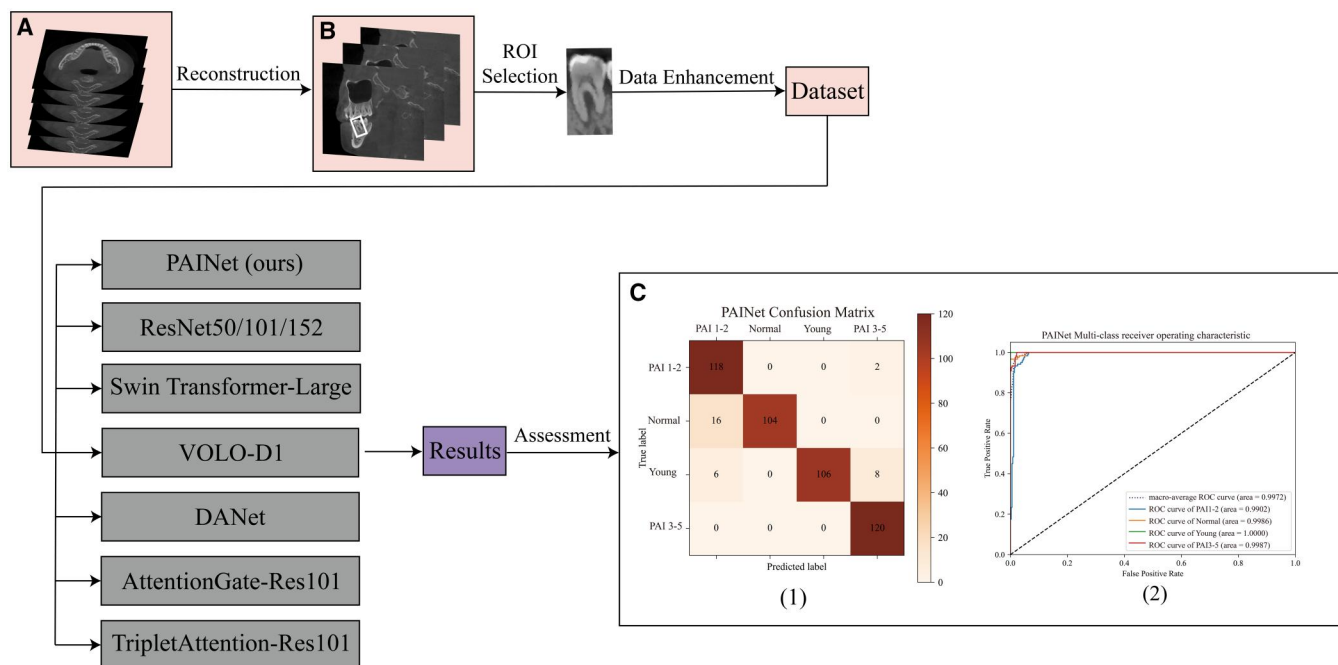
### Model construction and optimization

Convolutional neural networks (CNNs) of three classic architectures (ResNet50, ResNet101, and ResNet152<sup>24</sup>), two transformer-based architectures (Swin Transformer-Large<sup>25</sup> and VOLO-D1<sup>26</sup>), and a self-invented network (PAINet) were selected for complete training and model evaluation on the dataset. In addition, three other attention-based methods (DANet<sup>27</sup>, Attention Gate, and Triplet Attention<sup>28</sup>) are also used for comparison (Figure 2).

ResNet is a major invention in the field of computer vision. Residual blocks were used to enable the training of large models to overcome problems such as vanishing/exploding gradients. ResNet50/101/152 consists of four conv layers, each of which consists of several different numbers of bottleneck residual blocks (Figure 3). The self-invented network contained four layers, each of which contained several basic blocks (Figure 4C). Layers 1-4 included 3, 4, 23, and 3 basic blocks, which was consistent with ResNet101. PAINet and



**Figure 1.** The method of image pre-processing. (A) Schematic diagram of image pre-processing in the dataset. std, standard deviation. (B) An example of image pre-processing of an image from the training set.



**Figure 2.** The classification task. (A) Origin images; (B) reconstructed sagittal images. Tooth marked with a white box was of interest; (C) metrics used for evaluation, confusion matrix (1) and receiver operating curve (ROC) (2).

ResNet101 had the similar backbone framework. The bottleneck residual blocks in ResNet were improved by inserting the Convolutional Block Attention Module (CBAM<sup>29</sup>) at the end of each basic block. CBAM is added before the tensor addition calculation, forming a basic block using CBAM (Figure 4B). CBAM can increase the weight of important features on channels and space. The training process for each architecture was independent of another.

In recent years, transformer-based models have also performed well in the field of computer vision. A pure transformer (Swin Transformer-Large) and a CNN-Transformer (VOLO-D1) were chosen to have a comparison with PAINet. In addition to CBAM, many other attention methods have been reported, such as DANet,<sup>27</sup> Attention UNet,<sup>30</sup> Reverse Attention,<sup>31</sup> Triplet Attention,<sup>28</sup> and Multi-scale guided attention.<sup>32</sup> DANet, Attention UNet, and Triplet Attention were chosen to be compared with CBAM. DANet is originally a segmentation task network. ResNet101 was used as the backbone network, increased the attention structure of DANet, removed the upsampling operation, and changed the output layer to a fully connected layer to output the classification structure. The attention method in Attention UNet is Attention Gate. AttentionGate-Res101 was designed by adding an attention gate to each residual structure. Triplet Attention is a three-branch attention strategy. ResNet101 was used as the backbone network and added a Triplet Attention structure behind the four layers to calculate attention on width, height, and channel respectively, and performed fusion calculations of the three kinds of attention.

CNNs training was performed on a server at Google Colaboratory (<https://colab.research.google.com>) using a NVIDIA A100 TENSOR CORE GPU 40G for accelerating training. In the training process, regularization was used to

prevent overfitting. Batch normalization (BN) was used to accelerate model convergence and prevent overfitting. A mini-batch of eight was set. A low learning rate was chosen (0.00001). Meanwhile, cross-entropy was chosen as the loss function, the back-propagation algorithm was used to propagate the loss gradient, and the Adam optimizer was used to update the model parameters. The models of the above nine architectures were trained until the models converged on the training set. The accuracy on the validation set was the reference for selecting the best model parameters.

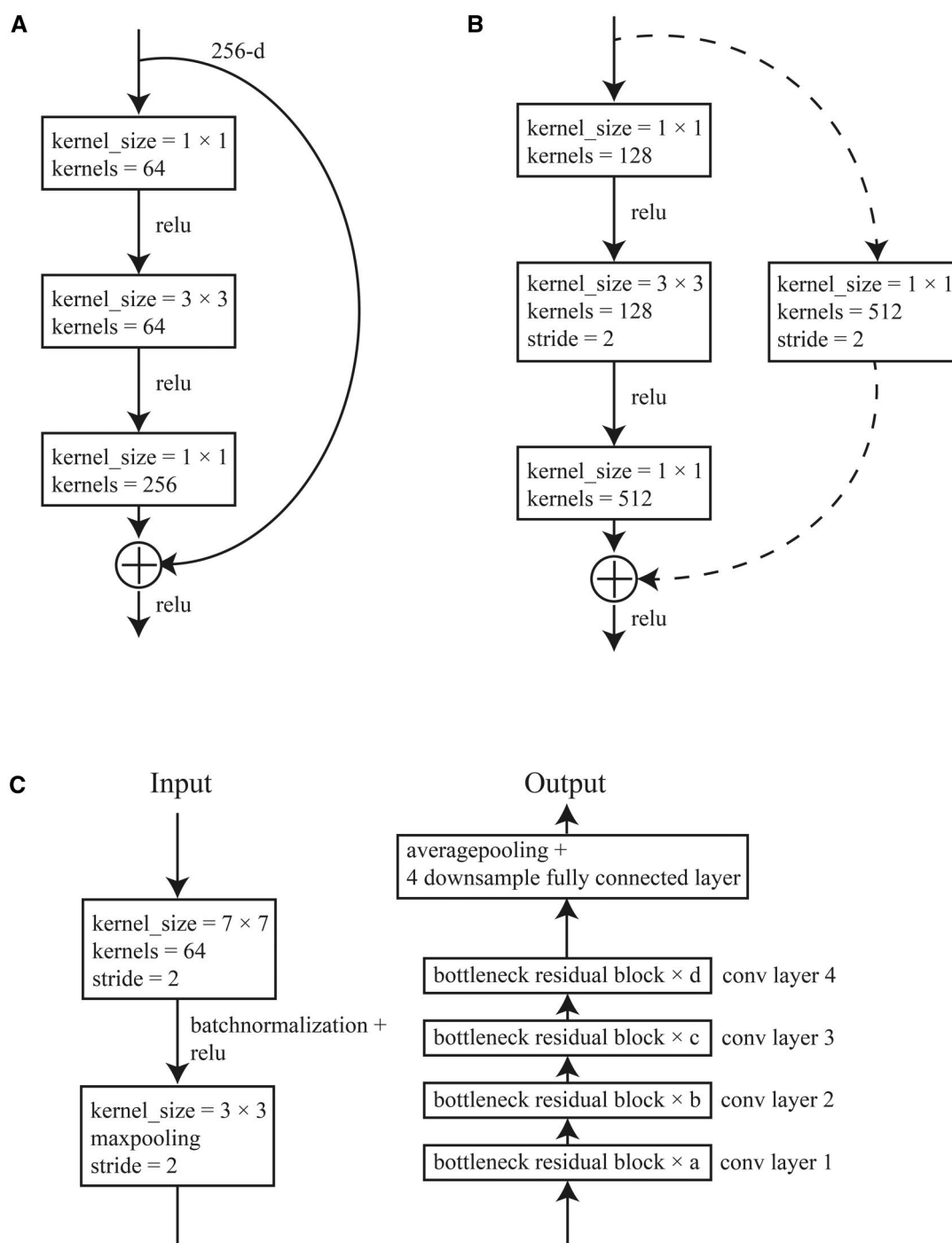
### Statistical analyses

The data were analysed using Python (<http://www.python.org>, version 3.7). The results of model evaluation in the test set were calculated to determine overall diagnostic accuracy, precision, recall, and balanced F score (F1-score). The confusion matrix visualizations were plotted and the area under the macro-average receiver operating curve (AUC) was calculated.

### Results

The visualizations of the confusion matrix and Cohen's kappa values on the test set are shown in Figure 5A-I. The macro-average receiver operating curves (ROCs) of the nine models on the test set are shown in Figure 6. The AUC values of ResNet50, ResNet101, ResNet152, and PAINet were 0.9593, 0.9734, 0.9531, and 0.9972, respectively (Figure 6A-D). The AUC values of Swin Transformer-Large and VOLO-D1 were 0.9900 and 0.9935, respectively (Figure 6E-F). The AUC values of DANet, AttentionGate-Res101, and TripletAttention-Res101 were 0.9736, 0.9650, and 0.9815, respectively (Figure 6G-I). The normal periapical and young



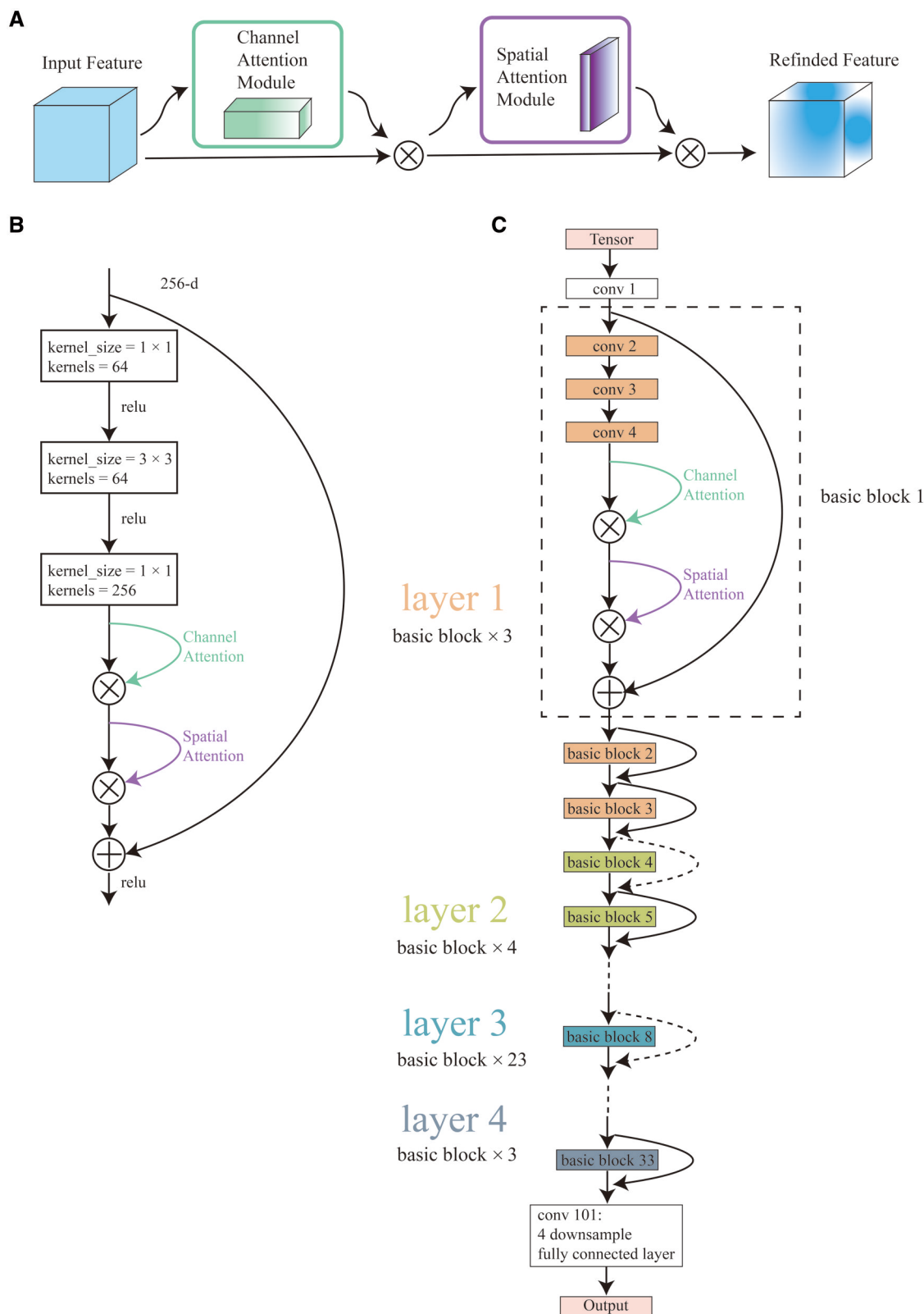


**Figure 3.** The structure of ResNet50/101/152. (A) A bottleneck residual block; (B) another kind of residual block with a downsampling convolutional layer on the shortcut. The extra convolutional layer here to ensure the same shape between the outputs from master branch and shortcut. (C) The structure of ResNet50/101/152, [a, b, c, d] equals to [3, 4, 6, 3], [3, 4, 23, 3], [3, 8, 36, 3].

permanent teeth showed better classification results in the nine models. The classification of PAI1-2 was relatively difficult in all architectures except ResNet50. According to Cohen's kappa, ResNet50, ResNet101, and ResNet152 were substantial consistency with the values of 0.731, 0.778, and 0.706. PAI1Net shown almost perfect consistency with expert opinions with the value of 0.911. Two transformer-based models and two attention models (DANet and

TripletAttention-Res101) also performed excellently with almost perfect consistency. AttentionGate-Res101 was not improved compared with ResNet101.

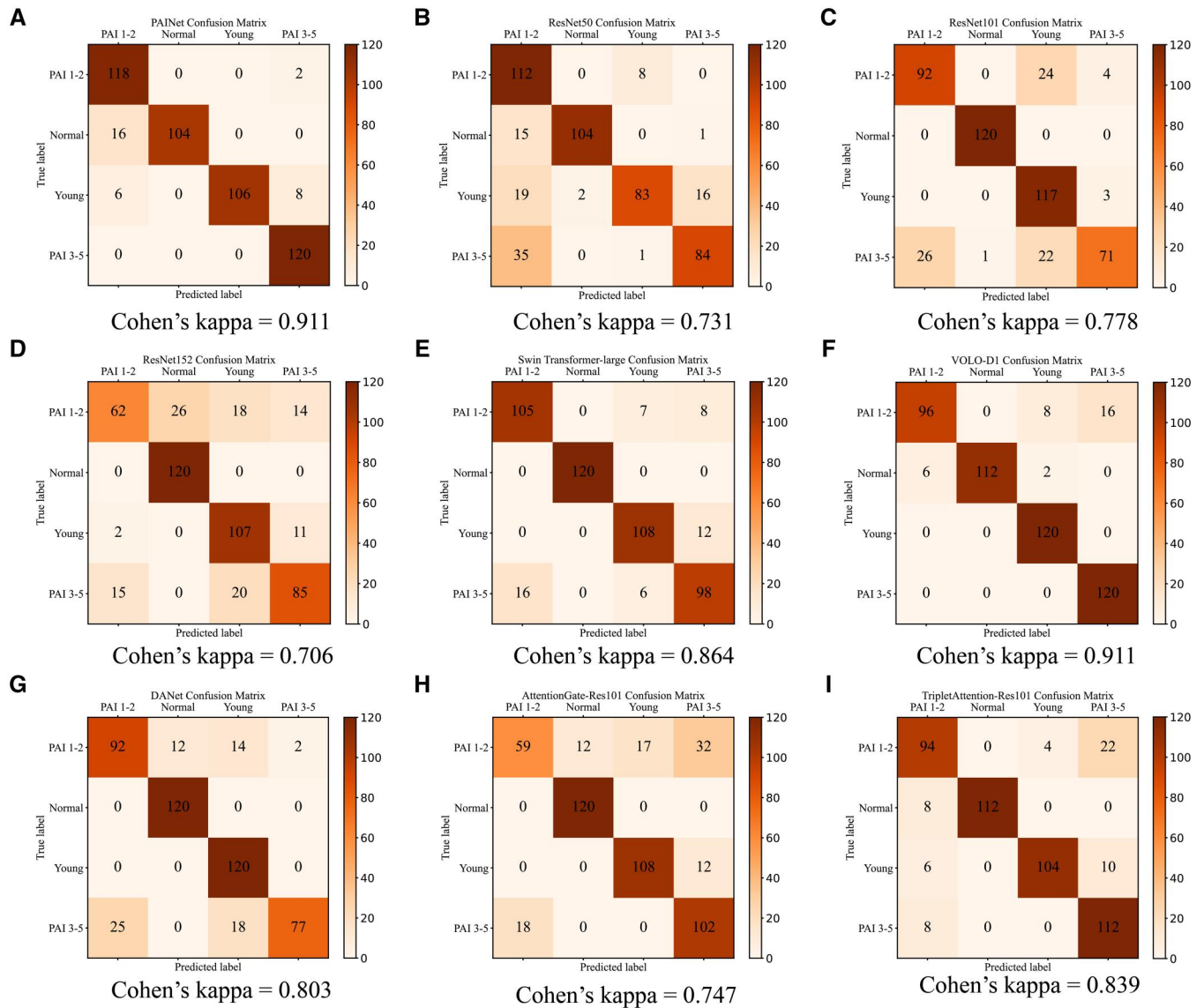
The classification performance of the nine models on the test set is shown in Table 1. In the three classic architectures, different evaluation indicators show the same distribution trend. ResNet101 performed the best, while ResNet152 performed the worst. The classification performance of



**Figure 4.** The structure of PAINet. (A) The structure of Convolutional Block Attention Module (CBAM); (B) a basic block using CBAM; (C) the overall structure of PAINet. The entire network included four layers. Layers 1-4 had 3, 4, 23, and 3 basic blocks using CBAM, respectively. The dashed line represented the use of basic blocks with downsampling. There was a total of 101 convolutional layers in the backbone part of the network. The last layer was a fully connected layer with four nodes.

PAINet greatly improved over the three classic architectures. The accuracy, precision, recall, and F1-score of PAINet on the test set were 0.9333, 0.9415, 0.9333, and 0.9336,

respectively. Both the transformer-based models performed very well. Among them, VOLO-D1 and PAINet achieved the same accuracy with the remaining evaluation indicators

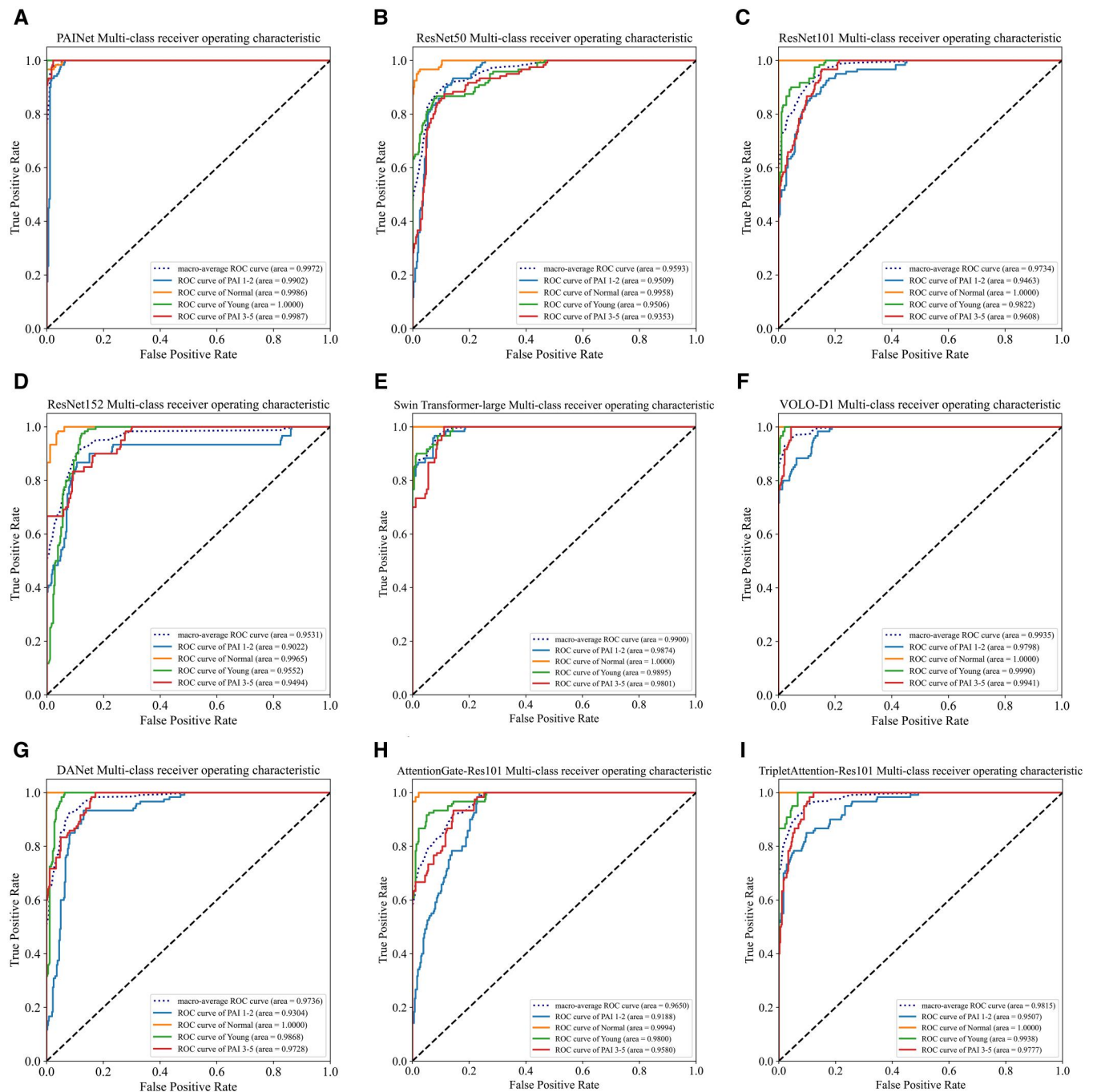


**Figure 5.** Visualizations of confusion matrix and Cohen's kappa for CNNs on the test set. (A) PAINet; (B) ResNet50; (C) ResNet101; (D) ResNet152; (E) Swin Transformer-Large; (F) VOLO-D1; (G) DANet; (H) AttentionGate-Res101; (I) TripletAttention-Res101.

very similar. VOLO is advanced architecture based on CNN+Transformer in recent years. The performance of PAINet was very close to VOLO, and slightly higher than VOLO in various evaluation indicators. The evaluation indicators of the three improved attention models based on ResNet101 have also been improved compared to ResNet101, except for AttentionGate-Res101. As a modular attention strategy, CBAM improved the performance of the model greatly compared to DANet, Attention Gate, and Triplet Attention.

The Grad-CAM<sup>33</sup> class activation heat map (Figure 7A-D) shows the key regions that PAINet focused on when making diagnosis. For categories diagnosed as having no periapical radiolucency, the model focused more on low-density areas such as the areas around the crown, the trabeculae around root, and the pulp chamber. Categories CBCTPAI 1-2 focused more on the high-density parts of the crown or canal. The CBCTPAI 3-5 category focused more on the low-density areas around the root. For young permanent teeth, more

attention was paid to the low density of the pulp chamber in the anterior teeth, while more attention was paid to the uniform density of the dental hard tissue in the posterior teeth. For images that were misclassified by PAINet, Grad-CAM can also more intuitively display the areas of concern to the model. In Figure 8A, a PAI 1-2 tooth was misclassified as PAI 3-5. The size of the AP of this tooth was at the critical level. The mistake made by PAINet was similar to that of human. At the same time, the loose trabeculae around the root were paid special attention by PAINet, leading to misclassification. In Figure 8B-D, the attention of the apical area of the tooth was extremely low compared to the correctly classified normal category. PAINet misjudged the maxillary sinus, relatively loose trabeculae and dental follicles. In Figure 8E, the low-density image of the root tip was correctly focused. However, it was misjudged as AP. This might be related to the lack of obvious contrast between the loose trabeculae and the low-density image area of the root canal. This is consistent with the mistakes human make.



**Figure 6.** ROCs of CNNs on the test set. (A) PAINet; (B) ResNet50; (C) ResNet101; (D) ResNet152; (E) Swin Transformer-Large; (F) VOLO-D1; (G) DANet; (H) AttentionGate-Res101; (I) TripletAttention-Res101.

## Discussion

In this study, CBCT data were used to reconstruct images, which completely displayed the root and periapical tissue, excluding the problems of image overlap and distortion that 2D radiographs. The sagittal plane of the tooth of interest was reconstructed. 3D data were compressed into several 2D planes. The computational complexity of the models was greatly reduced by this approach.<sup>34</sup> It could help the model to be promoted in areas with slightly backward computer equipment.

During the compression process, 3D data suffered from data loss. However, DL can not only use the image features of the apical area, but also use the features of pericoronal and root canal areas, such as artificial crowns, crown fillings, caries, or root canal fillings. The inherent hidden connections between various regions of the image were also features that DL can exploit. Therefore, although data were lost, the surrounding hidden information was supplemented to a certain extent. It can be seen from the performance of PAINet on



**Table 1.** Classification performance of CNNs on the test set.

| Overall apical detection on test set | True positives |       | False negatives |       | Diagnostic performances |        |         |       |          |
|--------------------------------------|----------------|-------|-----------------|-------|-------------------------|--------|---------|-------|----------|
|                                      | <i>n</i>       | %     | <i>n</i>        | %     | ACC (%)                 | SE (%) | PPV (%) | AUC   | F1-score |
| PAINet (ours)                        | 448            | 93.33 | 32              | 6.67  | 93.33                   | 93.33  | 94.15   | 0.997 | 0.9336   |
| ResNet50                             | 383            | 79.79 | 97              | 20.21 | 79.79                   | 79.79  | 83.34   | 0.959 | 0.8019   |
| ResNet101                            | 400            | 83.33 | 80              | 16.67 | 83.33                   | 83.33  | 84.99   | 0.973 | 0.8282   |
| ResNet152                            | 374            | 77.92 | 106             | 22.08 | 77.92                   | 77.92  | 77.93   | 0.953 | 0.7680   |
| Swin Transformer-Large               | 431            | 89.79 | 49              | 10.21 | 89.79                   | 89.79  | 89.77   | 0.990 | 0.8978   |
| VOLO-D1                              | 448            | 93.33 | 32              | 6.67  | 93.33                   | 93.33  | 93.67   | 0.994 | 0.9320   |
| DANet                                | 409            | 85.21 | 71              | 14.79 | 85.21                   | 85.21  | 86.49   | 0.974 | 0.8462   |
| AttentionGate-Res101                 | 389            | 81.04 | 91              | 18.96 | 81.04                   | 81.04  | 80.95   | 0.965 | 0.8000   |
| TripletAttention-Res101              | 422            | 87.92 | 58              | 12.08 | 87.92                   | 87.92  | 88.78   | 0.982 | 0.8807   |

Abbreviations: ACC = accuracy,  $ACC = (TP + TN)/(TP + TN + FP + FN)$ ; AUC = area under the receiver operating characteristic curve; FN = false negatives; FP = false positives; F1-score = balanced F score,  $1/F1\text{-score} = 1/2 \times (1/precision + 1/recall)$ ; PPV = positive predictive value, equals to precision,  $PPV = TP/(TP + FP)$  = precision; SE = sensitivity, equals to recall,  $SE = TP/(TP + FN)$  = recall; TN = true negatives; TP = true positives. For the overall performances,  $TN = FP = 0$ .

the test set that the partial loss of 3D data did not affect the model's judgement of the overall situation.

In addition, it may be related to the direction of expansion of AP lesions. Smaller AP has been reported to be more spherical and thus have a similar cross-sectional shape in all directions. Larger AP will appear oval or irregular in shape. According to the research, the boundary where sphericity changes is approximately 3 mm in diameter.<sup>35</sup> Therefore, it was reasonable to choose CBCTPAI 2 as the cut-off in this study. Reasonable level division also allowed the model to still perform well even when it could only get information from the sagittal position.

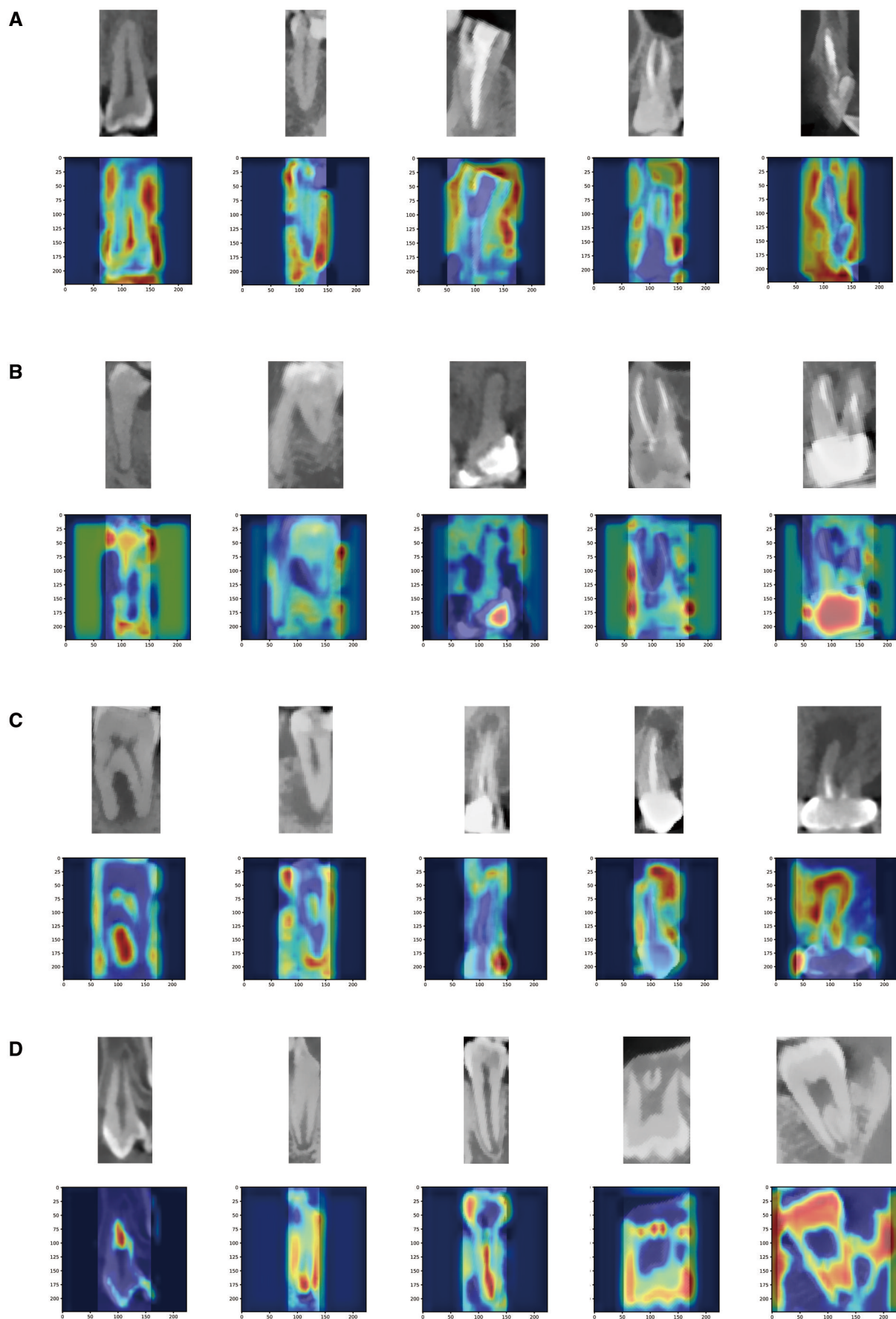
Scoring the severity of AP is helpful for treatment. In the past,  $PAI \geq 3$  was judged to be refractory AP, requiring periapical surgery or extraction of the affected tooth.<sup>13</sup> There are relatively few studies on the relationship between CBCTPAI and prognosis. It has been found that larger AP has poorer sphericity, which can help clinicians plan and perform apical surgery.<sup>35</sup> As mentioned previously, the cut-off for significant changes in sphericity is approximately 3 mm in diameter, which is consistent with the scoring division in this study. The AI scoring system in this study may be able to play a certain role in treating AP by clinicians.

Convolutional Block Attention Module is a combination of two dimensions, channel attention module (CAM) and spatial attention module (SAM).<sup>29</sup> Based on these two modules, adaptive feature extraction and enhancement of the input dataset in the channel direction and spatial direction is achieved, reinforcing the primary features and weakening the secondary features. The addition of CBAM enabled the model to focus on the content and location information of the target, thus improving all the performances of the model. With the help of CBAM, PAINet in this study focused more on learning important features. The results shown that the accuracy, precision, recall, and F1-score of PAINet on the test set were improved greatly compared with ResNet50/101/152. The AUC reached 0.9972 on average, which was close to completely correct. Transformer performs very well in the field of natural language processing. In recent years, scholars have tried to use transformer directly for computer vision and developed some excellent models, such as Vision Transformer and Swin Transformer.<sup>36</sup> VOLO is a transformer-based model proposed by Yuan et al in 2021.<sup>26</sup> This model uses CNN as the feature extractor and then

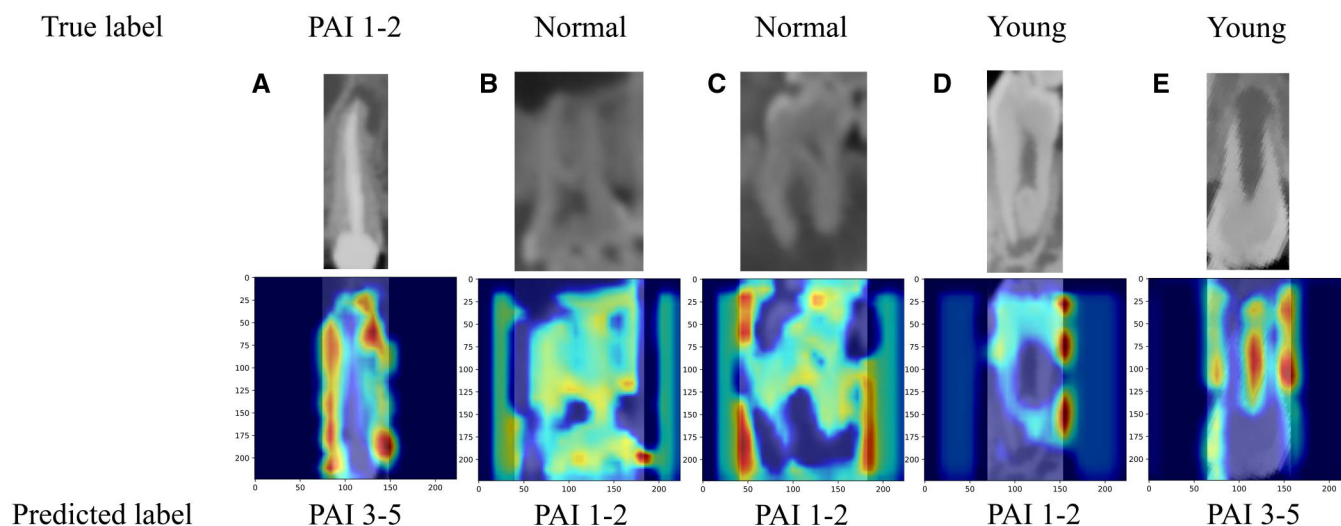
connects to the transformer. It is the most outstanding model architecture currently. PAINet was slightly better than VOLO in various evaluation indicators. PAINet was compared with three attention models. The performance improvement of PAINet was more significant than other models. The attention structure of AttentionGate-Res101 was the attention gate, which originally existed in Attention Unet. The performance of AttentionGate-Res101 was declined compared to ResNet101, which may be related to the fact that the attention gate structure is not suitable for the residual structure.

Current mainstream AI algorithms lack interpretability.<sup>37</sup> They are black-box models. This has led to difficulties in generalizing AI technology in some serious fields, such as medicine. A model that lacks intrinsic explanation is difficult to be accepted by patients.<sup>38</sup> We tried to explore the interpretability of the constructed classification algorithms using Grad-CAM. The class activation heatmap shown the areas where the model focuses its attention. No conflicts with existing knowledge were found. In a past study using AI to diagnose caries and AP, researchers used a similar approach.<sup>9</sup>

For the four categories of diagnosis, the three classic architectures and PAINet all shown similar trends (Figure 6). The diagnostic accuracy for normal periapical tissue and young permanent teeth was high, followed by CBCTPAI 3-5, and the worst was CBCTPAI 1-2. We guessed that the information contained in the images of the dataset was partially lost in the process of downsampling to  $224 \times 224$  size. For CBCTPAI 1-2, the lesion areas were small. The lost information accounted for a greater proportion of the original information, which effected the model's understanding of the features. We conjectured whether using a multiplexed convolutional network, sampling the crown 1/3, middle 1/3, and root 1/3 of the affected teeth separately. A high sampling rate would be used in the root 1/3, in order to overcome the problem of minor lesion information loss and the computational complexity of the model would not increase excessively at the same time. Some researchers have used similar structures to handle two classification problems simultaneously.<sup>9</sup> In contrast, we will use a three-passway neural network to handle one complex classification problem. In addition, the dataset constructed in this study was a single-centre small sample, which limited the generalization ability of the model to a



**Figure 7.** Four classification class activation heat maps of PAI-Net on test set. (A) Normal apices; (B) CBCTPAI 1-2; (C) CBCTPAI 3-5; (D) Young permanent teeth.



**Figure 8.** The class activation heat maps of misclassified samples. (A) A PAI 1-2 tooth misclassified as PAI 3-5; (B) a normal tooth misclassified as PAI 1-2; (C) a normal tooth misclassified as PAI 1-2; (D) a young permanent tooth misclassified as PAI 1-2; (E) a young permanent tooth misclassified as PAI 3-5.

certain extent. It will be one of our subsequent research directions to collect multicentre data. The AP grading system constructed in this study was based on human experts' opinions. The system itself does not have the ability to outperform human experts. This system is also unable to give correct results for some images that are too small and too blurry for human experts to judge. We look forward to the further development of AI technology.

## Conclusions

An AP grading system by CBCTPAI based on DL was developed and its reliability and accuracy were evaluated. This scoring system can use CBCT data to divide AP into two parts and distinguish it from radiolucent periapical tissue and young permanent teeth. It can assist junior doctors in better diagnosing and scoring AP and reduce the burden, playing a certain role in treatment. It can also be promoted in areas where experts are lacking to provide professional diagnostic opinions.

## Author contributions

Tianyin Zhao and Huili Wu contributed equally to this work. Tianyin Zhao, contribute to design, acquisition, analysis, and interpretation, drafted manuscript; Huili Wu, contribute to design, acquisition, and interpretation, drafted manuscript; Diya Leng, contribute to conception, acquisition, critically revised manuscript; Enhui Yao, contribute to acquisition, critically revised manuscript; Shuyun Gu, contribute to acquisition, critically revised manuscript; Minhui Yao, contribute to acquisition, critically revised manuscript; Qinyu Zhang, contribute to acquisition, critically revised manuscript; Tong Wang, contribute to acquisition, critically revised manuscript; Daming Wu, contribute to conception, design, and interpretation, critically revised manuscript; Lizhe Xie, contribute to design, and interpretation, critically revised manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

## Supplementary material

Supplementary material is available at *Dentomaxillofacial Radiology* online.

## Funding

The work was supported by the Scientific Research Project of Health Care for Cadres of Jiangsu Province (BJ21034), the Innovation and Entrepreneurship Training Program for College Students in Jiangsu Province (202110312098E), Jiangsu Province Capability Improvement Project through Science, Technology and Education-Jiangsu Provincial Research Hospital Cultivation Unit (YJXYJSDW4), and Jiangsu Provincial Medical Innovation Center (CXZX202227).

## Conflicts of interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

1. Arias Z, Nizami MZI, Chen X, et al. Recent advances in apical periodontitis treatment: a narrative review. *Bioengineering (Basel)*. 2023;10(4):488. <https://doi.org/10.3390/bioengineering10040488>
2. Siqueira JF Jr, Rôças IN. Present status and future directions: microbiology of endodontic infections. *Int Endod J*. 2022;55(Suppl 3):512-530. <https://doi.org/10.1111/iej.13677>
3. Cotti E, Schirru E. Present status and future directions: imaging techniques for the detection of periapical lesions. *Int Endod J*. 2022;55(Suppl 4):1085-1099. <https://doi.org/10.1111/iej.13828>
4. Estrela C, Bueno MR, Azevedo BC, Azevedo JR, Pecora JD. A new periapical index based on cone beam computed tomography. *J Endod*. 2008;34(11):1325-1331. <https://doi.org/10.1016/j.joen.2008.08.013>
5. Tibúrcio-Machado CS, Michelon C, Zanatta FB, Gomes MS, Marin JA, Bier CA. The global prevalence of apical periodontitis: a systematic review and meta-analysis. *Int Endod J*. 2021;54(5):712-735. <https://doi.org/10.1111/iej.13467>



6. Huuomonen S, Suominen AL, Vehkalahti MM. Prevalence of apical periodontitis in root filled teeth: findings from a nationwide survey in Finland. *Int Endod J*. 2017;50(3):229-236. <https://doi.org/10.1111/iej.12625>
7. Estrela C, Bueno MR, Leles CR, Azevedo B, Azevedo JR. Accuracy of cone beam computed tomography and panoramic and periapical radiography for detection of apical periodontitis. *J Endod*. 2008;34(3):273-279. <https://doi.org/10.1016/j.joen.2007.11.023>
8. Chogle S, Zuaitar M, Sarkis R, Saadoun M, Mecham A, Zhao Y. The recommendation of cone-beam computed tomography and its effect on endodontic diagnosis and treatment planning. *J Endod*. 2020;46(2):162-168. <https://doi.org/10.1016/j.joen.2019.10.034>
9. Li S, Liu J, Zhou Z, et al. Artificial intelligence for caries and periapical periodontitis detection. *J Dent*. 2022;122:104107. <https://doi.org/10.1016/j.jdent.2022.104107>
10. Nardi C, Calistri L, Pietragalla M, et al. Electronic processing of digital panoramic radiography for the detection of apical periodontitis. *Radiol Med*. 2020;125(2):145-154. <https://doi.org/10.1007/s11547-019-01102-z>
11. Orstavik D, Kerekes K, Eriksen HM. The periapical index: a scoring system for radiographic assessment of apical periodontitis. *Endod Dent Traumatol*. 1986;2(1):20-34. <https://doi.org/10.1111/j.1600-9657.1986.tb00119.x>
12. Sapmaz Ucan C, Dumani A, Unal I, Yilmaz S, Yoldas O. Effect of QMix as final irrigation protocol on periapical healing after single-visit root canal treatment: a randomised controlled clinical trial. *Aust Endod J*. 2023;49(Suppl 1):113-121. <https://doi.org/10.1111/aej.12701>
13. Wang L, Dong M, Shi D, et al. Role of PI3K in the bone resorption of apical periodontitis. *BMC Oral Health*. 2022;22(1):345. <https://doi.org/10.1186/s12903-022-02364-2>
14. Moidu NP, Sharma S, Chawla A, Kumar V, Logani A. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clin Oral Investig*. 2022;26(1):651-658. <https://doi.org/10.1007/s00784-021-04043-y>
15. Tarcin B, Gumru B, Iriboz E, Turkaydin DE, Ovecoglu HS. Radiologic assessment of periapical health: comparison of 3 different index systems. *J Endod*. 2015;41(11):1834-1838. <https://doi.org/10.1016/j.joen.2015.08.010>
16. Alves Dos Santos GN, Faria ESAL, Ribeiro VL, et al. Is the quality of root canal filling obtained by cone-beam computed tomography associated with periapical lesions? A systematic review and meta-analysis. *Clin Oral Investig*. 2022;26(8):5105-5116. <https://doi.org/10.1007/s00784-022-04558-y>
17. Smit JM, Krijthe JH, van Bommel J, Causal Inference for ICUC. The future of artificial intelligence in intensive care: moving from predictive to actionable AI. *Intensive Care Med*. 2023;49(9):1114-1116. <https://doi.org/10.1007/s00134-023-07102-y>
18. Bidzińska J, Szurowska E. See lung cancer with an AI. *Cancers (Basel)*. 2023;15(4):1321. <https://doi.org/10.3390/cancers15041321>
19. Aromiawura AA, Settle T, Umer M, et al. Artificial intelligence in cardiac computed tomography. *Prog Cardiovasc Dis*. 2023;81:54-77. <https://doi.org/10.1016/j.pcad.2023.09.001>
20. Song IS, Shin HK, Kang JH, et al. Deep learning-based apical lesion segmentation from panoramic radiographs. *Imaging Sci Dent*. 2022;52(4):351-357. <https://doi.org/10.5624/isd.20220078>
21. Kirnbauer B, Hadzic A, Jakse N, Bischof H, Stern D. Automatic detection of periapical osteolytic lesions on cone-beam computed tomography using deep convolutional neuronal networks. *J Endod*. 2022;48(11):1434-1440. <https://doi.org/10.1016/j.joen.2022.07.013>
22. Moorrees CF, Fanning EA, Hunt EE Jr. Age variation of formation stages for ten permanent teeth. *J Dent Res*. 1963;42(6):1490-1502. <https://doi.org/10.1177/00220345630420062701>
23. Wen J, Thibeau-Sutre E, Diaz-Melo M, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med Image Anal*. 2020;63:101694. <https://doi.org/10.1016/j.media.2020.101694>
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; June 27-30, 2016; Las Vegas, NV. IEEE; 2016; Piscataway, NJ. p. 770-778. <https://doi.org/10.48550/arXiv.1512.03385>
25. Liu Z, Lin Y, Cao Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows. Paper presented at: *IEEE International Conference on Computer Vision (ICCV)*; October 11-17, 2021; Montreal, QC. IEEE; 2021; Piscataway, NJ. p. 9992-10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
26. Yuan L, Hou Q, Jiang Z, Feng J, Yan S. VOLO: vision outlooker for visual recognition. *IEEE Trans Pattern Anal Mach Intell*. 2021;45(5):6575-6586. <https://doi.org/10.1109/TPAMI.2022.3206108>
27. Fu J, Liu J, Tian H, Fang Z, Lu H. Dual attention network for scene segmentation. Paper presented at: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; June 16-20, 2019; Long Beach, CA. IEEE; 2019; Piscataway, NJ. p. 3141-3149. <https://doi.org/10.1109/CVPR.2019.00326>
28. Misra D, Nalamada T, Arasanipalai AU, Hou Q. Rotate to attend: convolutional triplet attention module. Paper presented at: *IEEE Winter Conference on Applications of Computer Vision (WACV)*; January 5-9, 2021; Waikoloa, HI. IEEE; 2021; Piscataway, NJ. p. 3138-3147. <https://doi.org/10.1109/WACV48630.2021.00318>
29. Woo S, Park J, Lee J-Y, Kweon I-S. 2018. CBAM: convolutional block attention module. Paper presented at: *European Conference on Computer Vision (ECCV)*; September 8-14, 2018; Munich, Germany. Springer; 2018; Berlin, German. p. 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
30. Li C, Tan Y, Chen W, et al. Attention Unet++: a nested attention-aware U-Net for liver CT image segmentation. Paper presented at: *IEEE International Conference on Image Processing (ICIP)*; October 25-28, 2020; Abu Dhabi, United Arab Emirates. IEEE; 2020; Piscataway, NJ. p. 345-349. <https://doi.org/10.1109/ICIP40778.2020.9190761>
31. Fan D-P, Ji G-P, Zhou T, et al. 2020. PraNet: parallel reverse attention network for polyp segmentation. Paper presented at: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; October 4-8, 2020; Lima, Peru. Springer International Publishing; 2020; Berlin, German. p. 263-273. [https://doi.org/10.1007/978-3-030-59725-2\\_26](https://doi.org/10.1007/978-3-030-59725-2_26)
32. Sinha A, Dolz J. Multi-scale self-guided attention for medical image segmentation. *IEEE J Biomed Health Inform*. 2019;25(1):121-130. <https://doi.org/10.1109/JBHI.2020.2986926>
33. Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2016;128(2):336-359. <https://doi.org/10.1007/s11263-019-01228-7>
34. Wang C. A review on 3D convolutional neural network. Paper presented at: *IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*; January 6-8, 2023; Shenyang, China. IEEE; 2023; Piscataway, NJ. p. 1204-1208. <https://doi.org/10.1109/ICPECA56706.2023.10075760>
35. Boubaris M, Cameron A, Love R, George R. Sphericity of periapical lesion and its relation to the novel CBCT periapical volume index. *J Endod*. 2022;48(11):1395-1399. <https://doi.org/10.1016/j.joen.2022.08.009>
36. Liu Y, Zhang Y, Wang Y, et al. A survey of visual transformers. *IEEE Trans Neural Netw Learn Syst*. 2021;35(6):7478-7498. <https://doi.org/10.1109/tnnls.2022.3227717>
37. Liu Y. Explainable artificial intelligence and its practical applications. *ACE*. 2023;4(1):755-759. <https://doi.org/10.54254/2755-2721/4/2023419>
38. Vinuesa R, Sirmacek B. Interpretable deep-learning models to help achieve the sustainable development goals. *Nat Mach Intell*. 2021;3(11):926-926. <https://doi.org/10.1038/s42256-021-00414-y>



