

Received 13 June 2023, accepted 5 July 2023, date of publication 12 July 2023, date of current version 28 July 2023. *Digital Object Identifier* 10.1109/ACCESS.2023.3294617

# **RESEARCH ARTICLE**

# **Position Weighted Convolutional Neural Network** for Unbalanced Children Caries Diagnosis

# XIAOJIE ZHOU<sup>®</sup><sup>1</sup>, XUEOU FENG<sup>2</sup>, QINGMING LI<sup>2</sup>, QIYUE YIN<sup>®</sup><sup>2</sup>, JUN YANG<sup>®</sup><sup>3</sup>, GUOXIA YU<sup>®</sup><sup>1,4</sup>, (Member, IEEE), AND QING SHI<sup>5</sup>

<sup>1</sup>Department of Stomatology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China <sup>2</sup>Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

<sup>3</sup>Department of Automation, Tsinghua University, Beijing 100084, China

<sup>4</sup>Department of Stomatology, National Clinical Research Center for Respiratory Diseases, Beijing Children's Hospital, Capital Medical University,

National Center for Children's Health, Beijing 100045, China <sup>5</sup>Beijing Stomatological Hospital, Capital Medical University, Beijing 100050, China

beijing Stomatological Hospital, Capital Wedical Oniversity, Beijing 100050, C

Corresponding author: Guoxia Yu (yuguoxia@bch.com.cn)

This research was funded by Respiratory Research Project of National Clinical Research Center for Respiratory Diseases (No. HXZX-20210402) and National Natural Science Foundation of China (No. 81800925).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of Beijing Children's Hospital, Capital Medical University, National Center for Children's Health under IRB No.: [2022]-E-044-R.

**ABSTRACT** Panoramic radiograph is one of the most widely used inspection tools for dentists making caries diagnosis, especially for teeth that are hard to be diagnosed through visual inspection. Recently, several deep learning methods, e.g., based on convolutional neural network (CNN) or transformer network, have been proposed for automatic caries diagnosis on dental panoramic radiographs, and promising results have been achieved. However, current approaches use all the teeth equally when training their models, which results in performance degeneration because of unbalanced classification difficulties for different tooth positions. The objective of this study is to introduce a position weighted CNN to alleviate the above problem for more accurate caries diagnosis. The position weighted module evaluates and revises the output of a specially designed CNN to incorporate position information. In addition, a novel data augmentation method is used to balance data with uneven decayed and normal teeth, which is one of the reasons leading to unbalanced classification difficulty. To verify the proposed method, a children panoramic radiograph database is collected and labeled with more than 6,000 teeth. The proposed approach outperforms the stateof-the-art caries diagnosis methods with the accuracy, precision, recall, F1 and area-under-the-curve being 0.8859, 0.8875, 0.8932, 0.8903 and 0.9315, respectively. Specially, the proposed model displays higher diagnosis performance compared with two attending doctors with more than five-year clinical experience but with different diagnosis patterns, showing a potential tool for assisting dentists.

**INDEX TERMS** Caries diagnosis, CNN, transformer, position embedding, panoramic radiograph.

# I. INTRODUCTION

Dental caries is one of the most popular dental diseases. Usually, it is caused by a long time of interaction between acid-producing bacteria and the residual fermentable carbohydrates on teeth [1]. Dental caries can happen to anyone regardless of his/her age [2]. Even a baby with a few primary teeth is likely to have caries. What's worse is

The associate editor coordinating the review of this manuscript and approving it for publication was Sathish Kumar<sup>10</sup>.

the negative impact for the growth of permanent tooth if a primary tooth is decayed [3]. Accordingly, children caries diagnosis is paid more and more attention with various media methods such as dental maxillofacial images and 3D oral images. However, children usually refuse to cooperate when being diagnosed, making panoramic radiograph a widely used inspection tool for dentists diagnosing caries [4]. Currently, a quantity of researchers are studying automatic caries diagnosis on panoramic radiographs.



FIGURE 1. An example to show unbalanced classification difficulty. (a) is caries ratio for each tooth of the collected panoramic radiograph dataset; (b), (c), and (d) are diagnosis accuracy of each tooth for the Resnet, S-Transformer, and T2S-Transformer models, respectively.

Usually, performing caries diagnosis on panoramic radiograph consists of two steps [5], [6]. Firstly, each tooth is extracted from the panoramic radiograph, which consists of all the teeth along with useless information such as bone and jaw structure around the teeth. After extracting each tooth, a pattern recognition task is conducted to identify whether a tooth is a caries. For tooth extraction, many methods have been proposed like the most popular region-CNN methods [7], [8], [9], [10] and conventional approaches such as image post-processing [11] and genetic algorithms [12]. Since a simply trained data annotation worker can accurately extract each tooth from a panoramic radiograph, we move our study to the second step, which requires professional dentists, hoping to assist them for more accurate caries diagnosis.

To perform caries diagnosis on each tooth coming from panoramic radiographs, a classification task should be performed. Various classification methods have been proposed for caries diagnosis using conventional pattern recognition methods [13], [14] and the most popular deep learning methods such as using CNN and transformer network [15], [16], [17], and promising performance has been achieved. However, we find that current methods cannot reach the professional dentists' diagnosis level [15], [16]. One possible reason is that current approaches use all the teeth equally when training their models, which results in performance degeneration because of the unbalanced classification difficulties of different tooth positions.

To show the unbalanced classification difficulty, we collect a children panoramic radiograph dataset consisting of more than 6,000 teeth with balanced carious and normal teeth. Fig. 1 (a) shows the caries ratio of each tooth. Because different teeth have different probabilities to be caries [18], the caries ratio cannot be guaranteed to be balanced even the overall ratio is balanced. Fig. 1 (b), (c), and (d) display the classification accuracy of several state-of-the-art caries diagnosis methods on each tooth, i.e., Resnet [19], S-Transformer [16], and T2S-transformer [16]. Form the figure, it can be seen that there is a big performance gap among different tooth positions. The reasons may lie in the uneven number of carious and normal teeth (e.g., position 53), and the tooth itself (e.g., position 52).

To cope with the unbalanced classification difficulty problem, this study presents a position weighted CNN method for unbalanced children caries diagnosis. More specifically, a position weighted module is proposed to evaluate and revise the outputs of a specially designed CNN to bring in position information, which provides a compensation for the unbalanced problem. In addition, a novel data augmentation method is used to balance data with uneven carious and normal teeth, which further improves the final performance. A children panoramic radiograph database is collected and labeled with more than 6,000 teeth, based on which, accuracy, precision, recall, F1 and area-under-the-curve metrics are calculated to show the effectiveness of the proposed model. The major contributions of our method are:

- The problem of unbalanced classification difficulty for each tooth position is firstly noticed and alleviated in caries diagnosis.
- A position weighted module is proposed to embed position information in a specially designed CNN framework, which improves the overall performance.
- A novel data augmentation method is developed to balance tooth positions with uneven decayed and normal teeth, which may provide a novel view for tooth data augmentation.
- A children panoramic radiograph database with more than 6,000 teeth is used to verify the proposed method, which performs better than the state-of-the-art caries diagnosis methods, and shows higher diagnosis performance compared with two attending doctors with more than five-year clinical experience.

#### **II. LITERATURE REVIEW**

Caries diagnosis on panoramic radiograph can be naturally regarded as a pattern recognition problem. Conventional methods use typical feature descriptors to encode key features, and then train classifiers for final classification. For example, Saravanan et al. [13] found that the pixel intensities (from a histogram) are concentrated in different ranges for carious and normal teeth, and the spectrum of caries has high frequency components compared to the spectrum of the normal tooth, which are key features to diagnosis caries in its early stage. Virupaiah and Sathyanarayana [14] utilized Gaussian low pass filter to preprocess the dental X-ray images, and used support vector machine as a classifier for normal and decayed teeth classification, which shows a promising result for the automatic diagnosis of dental caries.

Recently, deep learning, especially the CNN technique, shows big performances improvement on image analysis compared with conventional pattern recognition methods such as support vector machine, and are now widely used for caries diagnosis [20], [21]. Vinayahalingam et al. [22] used



FIGURE 2. Example of data annotation. For a child panoramic radiograph with primary dentition, each tooth is extracted with a rectangular box, and marked with its position based on FDI annotation.

CNN MobileNet V2 for classification of caries in third molars on panoramic radiographs, and a high accuracy of 0.87 is achieved on a test set consisting of 100 cropped panoramic radiographs. Bui et al. [5] extracted features by combing deep activated features using a deep pre-trained model and geometric features using mathematical formulas, and the fused features are used to train a support vector machine classifier for caries diagnosis. Sinmilar with [5], Haghanifar et al. [23] used various pre-trained deep learning models through transfer learning to extract relevant features, and then trained a capsule network for classification. Imak et al. [24] used a multi-input deep convolutional neural network to extract features from the original image and its intensity colormap, and a score-based fusion framework is used for final caries classification. Zhou et al. [15] revised the Resnet model by using the features of adjacent teeth, and fused all the information through attention networks for classification.

Apart from the CNN technique, transformer, as a key component for various foundation models, are now widely used for computer vision tasks such as image classification [25], [26], [27]. Based solely on feed-forward neural network and multi-head self-attention, vision transformer overcomes the drawbacks of CNN, e.g, lacking long-range dependencies and high-order spatial interactions. Currently, a few researchers have started to rely on transformer for caries detection. For example, jiang et al. [28] proposed a fast caries detection method, called RDFNet, for caries detection, by embedding the transformer mechanism into the detection network for better feature learning. Zhou et al. [16] proposed a tooth type enhanced transformer for children caries diagnosis, where a swin transformer is used as the backbone classification network, and tooth types are used to revise the backbone using its specific and shared parts, and promising diagnosis performance has been achieved compared to convolutional neural networks.

Overall, we find that above methods cannot reach the professional dentists' diagnosis level, and most models only

perform well on several tooth positions as shown in Fig. 1, which leads to sub-optimal solutions. Thus, in this paper, we aim to alleviate this problem for more accurate caries diagnosis.

#### **III. METHOD**

This study presents a novel position weighted CNN for unbalanced children caries diagnosis. The proposed method consists of three components, i.e., a data augmentation module to balance tooth positions with uneven normal and decayed teeth, a shared CNN module to encode features of the current tooth and its adjacent teeth, a position weighted module to revise the outputs of CNN for final classification. In the following, we will first introduce the dataset, and then illustrate all the components, followed by the overall algorithm.

### A. DATASET

A children panoramic radiograph dataset is used, which was collected in Beijing Children's Hospital, Capital Medical University, National Center for Children's Health from year 2015 to 2021. The dataset consists of 304 panoramic radiographs with about 6,000 teeth, where there are 3039 and 2989 teeth being decayed and normal, respectively. The teeth are manually extracted from the panoramic radiograph based on annotation tool via [29], which is one of the most popular computer vision annotation tools. The labeling example is shown in Fig. 2, where teeth with unfixed rectangular boxes and their positions are extracted. Using unfixed rectangular box can better crop the tooth due to the different sizes between teeth, and the position is encoded with the FDI notation, which is a universal method [30]. For the labels of each extracted tooth, we use the discharge diagnosis report to provide the true labels. Noted tooth missing is common for children, so we just ignore the missing positions.

The dataset is used with the approval of the Institutional Review Board (IRB) of Beijing Children's Hospital, Capital



FIGURE 3. Overall framework of the proposed method. There are three key components, i.e., a data augmentation module, a shared CNN module, and a position weighted module. In the framework, we take tooth position 54 and its diagnosis as an example.

#### TABLE 1. Data characteristics.

Data	Num. caries	Num. non-caries	Total num.
Training	2432	2401	4833
Validation	320	279	599
Testing	287	309	596
Total	3039	2989	6028

Medical University (IRB No.: [2022]-E-044-R). With the collected dataset, we split all the teeth into training, validation and testing sets with 8:1:1 like most previous study, and the final data characteristics are summarized in Table 1.

# **B. SHARED CNN BACKBONE MODULE**

The backbone of the proposed method is a CNN, which encodes the high dimensional tooth image into a vector representation. Various CNN architectures can be used like previous methods, such as Alexnet, Googlenet and Resnet, and in this paper we use Resnet due to its promising performance in a variety of image analysis tasks. Usually, a typical Resnet consists of 18 layers with a convolutional layer, a pooling layer, several residual connected convolutional layers and a fully connected layer, as shown in the top of Fig. 3. Specifically, there are two kinds of residual connected convolutional layers as shown in Fig. 4. Apart from the regular residual block to alleviate the degradation problem, the residual layers with stride=2 will reduce the feature map size to limit the calculation complexity.

In [15], researchers found that using information of adjacent teeth can promote overall classification performance.



FIGURE 4. Two kinds of residual connected convolutional layers of the Resnet. The batch normalization layer and the ReLU layer are omitted due to space limitation.

The intuition behind the model is that caries of a tooth often affects its adjacent teeth due to the same growing environment. Accordingly, a tooth will more likely to be a caries if there are caries around it. In the original paper, an attention network is used to fuse information from adjacent teeth. In this paper, we just concatenate all the features to avoid additional network parameters. Specifically, a Resnet with shared parameters f is used to encode an image  $x_i$  to  $f(x_i)$ , and its final representation  $r_i$  is calculated as:

$$\boldsymbol{r}_i = [f(\boldsymbol{x}_i), f(\boldsymbol{x}_{i1}), \dots, f(\boldsymbol{x}_{ik})]$$
(1)

where  $f(\mathbf{x}_{ij}), j \in 1, ..., k$  is the feature of k adjacent teeth for  $\mathbf{x}_i$ . In our experiments, we set k to 3. Taking tooth 54 as an example, its three adjacent teeth are positions 55, 53 and 84.

With the final representation, a fully connected layer with a softmax activation function is used to obtain a classification distribution  $c_i^s$ .

# C. POSITION WEIGHTED MODULE

In previous sections, we illustrate the unbalanced classification difficulties of different tooth positions, and in this subsection, we introduce position weighted module to solve this problem. Specifically, the module is used to measure the classification difficulty of the current tooth position, and then revise the classification distribution of the shared CNN backbone. The calculation is shown in the right bottom of Fig. 3, where there are several normalization and fully connected layers followed by a softmax activation function.

Firstly, the position code (e.g, 54 based on FDI annotation) of a tooth is normalized to the range of (0,1]. Secondly, the normalized code is mapped with a fully connected layer with ReLU activation function. Thirdly, another fully connected layer without using activation function is used to map its input to a two-dimensional vector, which is activated by a softmax function to obtain a distribution  $c_i^w$ . Finally, the classification distribution of tooth *i* is obtained with:

$$\boldsymbol{c}_i = \boldsymbol{c}_i^s \odot \boldsymbol{c}_i^w \tag{2}$$

where  $\odot$  is the Hadamard product. With the modified classification distribution, a cross entropy loss can be used as the final objective.

# D. DATA AUGMENTATION MODULE

Caries ratio for each tooth position is shown in Fig. 1 (a). It can be seen that there are some positions with uneven decayed and normal teeth (see positions 71-71 and 81-83), even when the caries ration for the whole tooth data is balanced. This observation is reasonable because different teeth have distinct probabilities to be caries. The uneven class distribution for a position makes the classification difficult, and data augmentation [31] is always performed to alleviate the defect. Conventional data augmentation methods usually operate the image itself, e.g. rotating the image to some degrees, which we believe cannot capture the characteristics of the tooth image.

Generally, teeth in a panoramic radiograph are usually symmetric, so a possible data augmentation method is to use its symmetric tooth as the current tooth. For example, as shown in the left bottom of Fig. 3, for a tooth position 54, if the decayed teeth number is much less than the normal teeth number, we can put some caries of position 64 as the caries image for position 54. Then the number of caries image for position 54 will increase. With this operation, the uneven tooth positions will be even without operating the tooth images with the current positions, e.g., 54 in the above example. We believe this operation can largely increase the data diversity, and we will see its effect in the Experiment part.

With all the modules, the pseudocode of the proposed position weighted CNN is summarized in Algorithm 1.

77038

Algorithm 1 Position Weighted CNN for Unbalanced Children Caries Diagnosis

Input: The model (parameters) shown in Section III-B-C, and the dataset shown in Section III-A.

Output: Children caries diagnosis result.

# Train the proposed model:

Perform data augmentation as shown in Section III-D; while not converged **do** 

Train the model parameters including the shared CNN backbone and the position weighted modules as shown in Section III-B and C using the dataset training part;

Save the model (parameters) based on the best result (accuracy) on the dataset validation part;

Test the proposed model:

Calculate all the metrics of the saved model on the dataset testing part.

# **IV. RESULTS**

In this section, we try to figure out three important questions: (1) is the proposed method performing better than the state-of-the-art caries diagnosis methods? (2) How does each component contribute to the proposed method, such as the data augmentation module and the position weighted module? (3) how does the proposed method perform compared with professional dentists? In the following, we will first introduce the experimental settings, and then answer the questions above one by one.

#### A. EXPERIMENTAL SETTINGS

The experiments were conducted on the dataset as shown in Section III-A. All the methods including the proposed method and the compared methods were trained on the training set, where the model is saved based on the best performances on the validation set. Then the results of the model is reported by using the testing set. To provide a comprehensive testing, five classical classification metrics are utilized like in previous works, i.e., accuracy, precision, recall, F1-score and area-under-the-curve (AUC).

The five metrics are calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(3)

$$Precision = \frac{TT}{TP + FP}$$
(4)

$$\operatorname{Recall} = \frac{TP}{TP + FN}$$
(5)

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$
(6)

where TP, TN, FP, and FN are the true positive, true negative, false positive, and false negative, respectively. Suppose the labels for the normal and decayed teeth are 0 and 1, then: the true positive is the number of teeth that are labeled as 1 and predicted as 1; the true negative is the number of teeth that

TABLE 2. Results of the proposed method and the compared methods.

Methods	Accuracy	Precision	Recall	F1	AUC
AlexNet	0.6040	0.6181	0.6181	0.6181	0.6547
GoogleNet	0.6376	0.6317	0.7217	0.6737	0.6633
SeNet	0.7836	0.8000	0.7767	0.7882	0.8520
ResNet	0.7768	0.8056	0.8049	0.8052	0.8490
CA-CNN-2	0.8020	0.8051	0.8155	0.8155	0.8537
CA-CNN-3	0.8272	0.8538	0.8770	0.8652	0.9005
CA-CNN-5	0.8104	0.8452	0.8738	0.8593	0.8725
S-Transformer	0.8272	0.8576	0.7994	0.8275	0.8991
T2S-Transformer	0.8557	0.8832	0.8317	0.8567	0.9223
PW-CNN	0.8859	0.8875	0.8932	0.8903	0.9315

are labeled as 0 and predicted as 0; the false positive is the number of teeth that are labeled as 0 and predicted as 1; the false negative is the number of teeth that are labeled as 1 and predicted as 0. For the metric AUC, the receiver operating characteristic (ROC) curve should be plotted, and the AUC is calculated as the area under the ROC. Generally, ROC is plotted based on the true positive rate (TPR) and false positive rate (FPR), where TPR equals to recall, and FPR is calculated as:

$$FPR = \frac{FP}{FP + TN}$$
(7)

By setting different thresholds, different TPR and FPR values are calculated, which serve as the vertical and horizontal coordinates of the ROC curve.

Apart from above metrics, we introduce another metric, called kappa coefficient [32], to measure the agreement between two classification patterns, e.g. the classification results between the proposed model and professional dentists. The kappa coefficient is defined as:

$$\kappa = \frac{P_a - P_e}{1 - P_e} \tag{8}$$

where  $P_a$  is the percent agreement, and  $P_e$  is the expected agreement. Usually, kappa value lies in [-1.0,1.0], and the higher the perfect agreement. If a value of [0.8,1.0] is obtained, an almost perfect agreement is reached.

We compare our method with several typical deep learning methods, including CNN based methods such as AlexNet [33], GoogleNet [34], SeNet [35] and ReseNet [19], and transformer based methods such as swin transformer [26]. Besides, several state-of-the-art caries diagnosis methods are used as baselines. Specifically, CA-CNN [15] is a context aware convolutional neural network for caries diagnosis, which improves ResNet by considering adjacent teeth using attention technique for information fusing. CA-CNN-X means X adjacent teeth being considered. T2S-Transformer [16] is a tooth type enhanced swin transformer for caries diagnosis, which improves swin transformer

TABLE 3. Hyper-parameters used for our model training.

Hyper-parameters	Value
Learning rate	0.001
Mini-batch size	32
Maximum epoch number	10000
fc layer size in our model	128
Optimization approach	Stochastic gradient decent

by taking tooth types information into account using shared and specific parameters of transformer.

For all the compared methods, we follow the settings of their original paper to train the models from scratch without using the pre-trained models. For the proposed method, a similar training method is used, and the hyper-parameters used for all the experiments are summarized in Table 3. Besides, we resize the original tooth images to  $224 \times 224$ , which is necessary because we use different sized rectangular boxes for tooth extraction. For the training platform, a workstation was used, which consists of  $2 \times \text{Intel}(R)$  Xeon(R) Gold 6240R CPU and  $4 \times \text{NVIDIA}$  RTX 2080 Ti GPU. For the training code, Python + Pytorch was used, and can be requested with academic research use.

#### **B. CARIES DIAGNOSIS PERFORMANCE**

In this subsection, we try to answer question (1) by comparing our method (PW-CNN) with the state-of-the-art caries diagnosis methods. The results in terms of accuracy, precision, recall, F1 and AUC are shown in Table 2. Overall, it can be seen that the proposed method outperforms all the compared methods.

Compared with CNN based methods, i.e., AlexNet, GoogleNet, SeNet, ResNet, CA-CNN-2, CA-CNN-3 and CA-CNN-5, a big improvement of our method can be achieved, i.e., 7.1%, 3.9%, 1.8%, 2.9% and 3.4% percentages improvements of accuracy, precision, recall, F1 and AUC compared with the second best method CA-CNN-3. It can



FIGURE 5. Classification of the proposed method for each tooth compared with CA-CNN-3 and T2S-Transformer, which are best caries diagnosis methods based on CNN and Transformer network respectively.



**FIGURE 6.** ROC curves of the proposed method compared with CA-CNN-3 and T2S-Transformer.

be proved that using position weighted module and data augmentation module can boost the CNN performance.

Compared with transformer network based methods, i.e, S-Transformer and T2S-Transformer, some improvement can be obtained, i.e, 3.5%, 0.5%, 7.4%, 3.9% and 1.0% percentages improvements of accuracy, precision, recall, F1 and AUC compared with the second best method T2S-transformer. Even though the transformer network based

77040

methods are a bit better than the CNN based methods, PW-CNN still shows performance improvement, which again validates the advantages of using position weighted module and data augmentation module.

To further show the superiority of the proposed method, we plot the ROC curves of the proposed method compared with CA-CNN-3 and T2S-transformer, as shown in Fig. 6. CA-CNN-3 and T2S-Transformer are best caries diagnosis methods based on CNN and Transformer network. From the figure, it can be seen the advantage of the proposed method in a wide range of (1-secificity, secificity) area.

Finally, the accuracy of each tooth position for the proposed method and CA-CNN-3 and T2S-Transformer methods is shown in Fig. 5. The proposed method outperforms both the two best CNN and transformer network based methods for 9 tooth positions. To show the proposed method can alleviate the unbalanced classification problem, we calculate the classification standard deviation of typical methods, as shown in Table 4. It can be seen that the proposed method has the best performance along with a minimum standard deviation. This implicitly reflects that the proposed method can alleviate the unbalanced classification problem to some extent.

#### C. EFFECT OF THE MODULES

In this subsection, we try to answer question (2) by performing ablation experiment. In our proposed method, apart from the shared CNN module used as a backbone for caries classification, there are two important modules, i.e.,

TABLE 4. Classification accuracy (mean value  $\pm$  standard deviation) for all tooth positions.

Methods	Acuracy (mean $\pm$ standard deviation)
Resnet	$0.7768 \pm 0.0559$
CA-CNN-3	$0.8272 \pm 0.0692$
S-Transformer	$0.8272 \pm 0.0799$
T2S-Transformer	$0.8557 \pm 0.0943$
PW-CNN	$\textbf{0.8859} \pm \textbf{0.0465}$



**FIGURE 7.** Effect of the position embedding and data augmentation modules for caries diagnosis.

data augmentation module and position weighted module. To show how these two modules effect the caries diagnosis performance, we remove either the data augmentation module (w/o augmentation) or the position embedding module (w/o position), and report their performances. In addition, the proposed method without the two modules serves as the baseline. The results are shown in Fig. 7.

Without considering data augmentation and position embedding modules, the proposed method has the worst performance. However, by adding a module, either the data augmentation or the position embedding module, the proposed method obtains a better performance. From the histogram result, the position embedding module seems more important than the data augmentation module. Considering both modules, the proposed method reaches the best performance. However, we can see the precision and recall metrics seem to drop compared with the one only considering position



**FIGURE 8.** Effect of the data augmentation module for caries diagnosis for tooth positions with uneven decayed and normal teeth.

embedding. On possible reason is that the position embedding module may dominate the most performance improvement.

Finally, we want to see if the data augmentation module can improve the performance for tooth positions with uneven caries ratios. From Fig. 1 (a), the collected dataset has uneven decayed and normal teeth for positions 71, 72, 73, 81, 82 and 83, so we calculate the classification accuracy for those positions using our method with only the shared CNN module (w/o augmentation) and the CNN plus the data augmentation module (w augmentation). The result is shown in Fig. 8. From the comparison, it further verifies the usefulness of the data augmentation module.

# D. COMPARISON WITH DENTISTS

In this subsection, we try to answer question (3) by comparing our method with professional dentists. We invite two attending doctors from Department of Stomatology, Beijing Children's Hospital, Capital Medical University to perform a comparison. All the test panoramic radiographs and their cropped tooth images are provided to the dentists for caries diagnosis. After recording their judgement on each tooth, we calculate the classification metrics, which are shown in Table 5. It can be seen that the proposed method reaches the professional dentists' level, which is promising because it provides a chance for clinical application. In addition, we can see the diagnosis time used for our method is much shorter compared with professional dentists (1.4464 second vs more than 1 minute on a panoramic radiograph). This further validates the advantage of the proposed method.

After reaching the professional dentists' level, we want to see if the proposed method behaves the same pattern as the professional dentists. First, we give the classification accuracy for each tooth position, which is shown in Fig. 9. From the figure, we can see that the proposed method performs better or equal for tooth positions 55, 54, 65, 64, 75, 74, 85, 61, 63, 73, 83. A different classification pattern is observed, and the proposed method is more advantageous for molars and canines. Second, we calculate the kappa value of the proposed method with the dentists, to further show the differences among the model and human. The result is shown

#### TABLE 5. Results of the proposed method and the invited attending doctors.



FIGURE 9. Classification of the proposed method for each tooth position compared with professional dentists (average performance is reported).



FIGURE 10. Classification differences for all the teeth, and the Kappa coefficients of the two professional dentists and the proposed method with dentists.

in Fig. 10. From the result, the diagnosis patterns for the two dentists are similar while the diagnosis patterns of proposed model and professional dentists are different. Considering the promising results of the proposed method, especially for the molars and canines, the proposed model may serve as a good tool for dentists.

# **V. DISCUSSION**

CNN and transformer network are popular deep learning techniques used for computer vision tasks, which are now brought into the medical image analysis field. By applying the state-of-the-art networks or modifying the networks based on domain specific knowledge, promising caries diagnosis performance has been achieved. However, none of the previous works consider the unbalanced classification difficulties of different tooth positions, leading to under-performing performance compared with professional dentists. This paper proposes probably the first position weighted CNN method for unbalanced children caries diagnosis.

The detailed comparison is displayed in Table 2. The proposed method outperforms the state-of-the-art caries diagnosis methods in terms of accuracy, precision, recall, F1 and AUC. The comparison with professional dentists is shown in Table 5, Fig. 9 and Fig. 10. From the results, the proposed

method reaches the professional dentists' level. However, the classification results show different patterns, and the model performs better for molars and canines, which may provide a supplement or tool to assist dentists in clinical application.

#### **VI. CONCLUSION**

In this paper, we have proposed a position weighed CNN for unbalanced children caries diagnosis, and a big caries diagnosis dataset has been used, with which, the proposed method outperforms the state-of-the-art deep learning methods such as various CNN and transformer networks. More importantly, the proposed method reaches the professional dentists' diagnosis level (based on the classification metrics) but with different classification patterns. The model performs better for molars and canines, while dentists diagnose well for incisor. This provides an opportunity for our model to assist dentists for more accurate caries diagnosis.

In the future, apart from improving the classification performance for more accurate caries diagnosis, we aim to provide a depth caries diagnosis such as different level of caries.

#### REFERENCES

- [1] F. Oztekin, O. Katar, F. Sadak, M. Yildirim, H. Cakar, M. Aydogan, Z. Ozpolat, T. T. Yildirim, O. Yildirim, O. Faust, and U. R. Acharya, "An explainable deep learning model to prediction dental caries using panoramic radiograph images," *Diagnostics*, vol. 13, no. 2, p. 226, Jan. 2023.
- [2] H. Colak, C. T. Dulgergil, M. Dalli, and M. M. Hamidi, "Early childhood caries update: A review of causes, diagnoses, and treatments," *J. Natural Sci., Biol., Med.*, vol. 4, no. 1, pp. 29–38, 2013.
  [3] M. S. Skeie, M. Raadal, G. V. Strand, and I. Espelid, "The relationship
- [3] M. S. Skeie, M. Raadal, G. V. Strand, and I. Espelid, "The relationship between caries in the primary dentition at 5 years of age and permanent dentition at 10 years of age—A longitudinal study," *Int. J. Paediatric Dentistry*, vol. 16, no. 3, pp. 152–160, 2006.
- [4] M. Nandeesh and B. Naveen, "A literature review on carries detection and classification in dental radiographs," *Ind. Eng. J.*, vol. 13, no. 6, Jul. 2020. [Online]. Available: https://www.researchgate. net/publication/357015814\_A\_literature\_Review\_on\_Carries\_Detection\_ and\_Classification\_in\_Dental\_Radiographs
- [5] T. H. Bui, K. Hamamoto, and M. P. Paing, "Deep fusion feature extraction for caries detection on dental panoramic radiographs," *Appl. Sci.*, vol. 11, no. 5, p. 2005, Feb. 2021.
- [6] M. P. Muresan, A. R. Barbura, and S. Nedevschi, "Teeth detection and dental problem classification in panoramic X-ray images using deep learning and image processing techniques," in *Proc. IEEE 16th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Cluj-Napoca, Romania, Sep. 2020, pp. 457–463.
- [7] H. Zhu, Z. Cao, L. Lian, G. Ye, H. Gao, and J. Wu, "CariesNet: A deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image," *Neural Comput. Appl.*, vol. 35, pp. 16051–16059, 2023.
- [8] B. Thanathornwong and S. Suebnukarn, "Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks," *Imag. Sci. Dentistry*, vol. 50, no. 2, p. 169, 2020.
- [9] J.-H. Lee, S.-S. Han, Y. H. Kim, C. Lee, and I. Kim, "Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs," *Oral Surg., Oral Med., Oral Pathol. Oral Radiol.*, vol. 129, no. 6, pp. 635–642, Jun. 2020.
  [10] M. Chung, J. Lee, S. Park, M. Lee, C. E. Lee, J. Lee, and Y.-G. Shin,
- [10] M. Chung, J. Lee, S. Park, M. Lee, C. E. Lee, J. Lee, and Y.-G. Shin, "Individual tooth detection and identification from dental panoramic X-ray images via point-wise localization and distance regularization," *Artif. Intell. Med.*, vol. 111, Jan. 2021, Art. no. 101996.
- [11] R. Kaur, R. S. Sandhu, A. Gera, and T. Kaur, "Edge detection in digital panoramic dental radiograph using improved morphological gradient and MATLAB," in *Proc. Int. Conf. Smart Technol. Smart Nation (SmartTech-Con)*, Bengaluru, India, Aug. 2017, pp. 793–797.

- [12] A. Haghanifar, M. M. Majdabadi, and S. Ko, "Automated teeth extraction from dental panoramic X-ray images using genetic algorithm," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Seville, Spain, Oct. 2020, pp. 1–5.
  - [13] T. Saravanan, M. S. Raj, and K. Gopalakrishnan, "Identification of early caries in human tooth using histogram and power spectral analysis," *Middle-East J. Sci. Res.*, vol. 20, no. 7, pp. 871–875, 2014.
  - [14] G. Virupaiah and A. K. Sathyanarayana, "Analysis of image enhancement techniques for dental caries detection using texture analysis and support vector machine," *Int. J. Appl. Sci. Eng.*, vol. 17, no. 1, pp. 75–86, 2020.
  - [15] X. J. Zhou, G. X. Yu, Q. Y. Yin, Y. Liu, Z. L. Zhang, and J. Sun, "Context aware convolutional neural network for children caries diagnosis on dental panoramic radiographs," *Comput. Math. Methods Med.*, vol. 2022, Sep. 2022, Art. no. 6029245.
  - [16] X. Zhou, G. Yu, Q. Yin, J. Yang, J. Sun, S. Lv, and Q. Shi, "Tooth type enhanced transformer for children caries diagnosis on dental panoramic radiographs," *Diagnostics*, vol. 13, no. 4, p. 689, Feb. 2023.
  - [17] L. Lian, T. Zhu, F. Zhu, and H. Zhu, "Deep learning for caries detection and classification," *Diagnostics*, vol. 11, no. 9, p. 1672, Sep. 2021.
  - [18] S. Saravanan, I. Madivanan, B. Subashini, and J. W. Felix, "Prevalence pattern of dental caries in the primary dentition among school children," *Indian J. Dental Res.*, vol. 16, no. 4, pp. 140–146, 2005.
  - [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
  - [20] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.
  - [21] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022.
  - [22] S. Vinayahalingam, S. Kempers, L. Limon, D. Deibel, T. Maal, M. Hanisch, S. Berg, and T. Xi, "Classification of caries in third molars on panoramic radiographs using deep learning," *Evol. Intell.*, vol. 11, p. 12609, 2021.
  - [23] A. Haghanifar, M. M. Majdabadi, and S.-B. Ko, "PaXNet: Dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier," 2020, arXiv:2012.13666.
  - [24] A. Imak, A. Celebi, K. Siddique, M. Turkoglu, A. Sengur, and I. Salam, "Dental caries detection using score-based multi-input deep convolutional neural network," *IEEE Access*, vol. 10, pp. 18320–18329, 2022.
  - [25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, arXiv:2010.11929.
  - [26] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 9992–10002.
  - [27] K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, Y. Gao, Q. Wang, J. F. Zhang, and D. G. Shen, "Transformers in medical image analysis," *Intell. Med.*, vol. 3, no. 1, pp. 59–78, Feb. 2023.
  - [28] H. Jiang, P. L. Zhang, C. Che, and B. Jin, "RDFNet: A fast caries detection method incorporating transformer mechanism," *Comput. Math. Methods Med.*, vol. 2021, Nov. 2021, Art. no. 9773917.
  - [29] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proc. 27th ACM Int. Conf. Multimedia*, Nice, France, Oct. 2019, pp. 2276–2279.
  - [30] B. Y. Tekin, C. Ozcan, A. Pekince, and Y. Yasa, "An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105547.
- [31] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.
- [32] T. Gjørup, "The Kappa coefficient and the prevalence of a diagnosis," *Methods Inf. Med.*, vol. 27, no. 4, pp. 184–186, 1988.
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014, arXiv:1409.4842.

[35] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.



**XIAOJIE ZHOU** received the B.S. and M.S. degrees from Capital Medical University, China, in 2009 and 2011, respectively. She is currently an attending Doctor with the Department of Stomatology, Beijing Children Hospital, Capital Medical University, National Center for Children's Health. Her research interests include pediatric stomatology and machine learning on medical image analysis.



**QIYUE YIN** received the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2017. He is currently an Associate Professor with the Institute of Automation, Chinese Academy of Sciences. His research interests include machine learning, pattern recognition, and artificial intelligence on games and medical image.



**JUN YANG** received the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2011. He is currently an Associate Professor with the Department of Automation, Tsinghua University. His research interests include multi-agent reinforcement learning and game theory.



**XUEOU FENG** received the B.S. degree from the Nanjing University of Science and Technology, China, in 2021. She is currently an Algorithm Engineer with the Institute of Automation, Chinese Academy of Sciences. Her research interests include reinforcement learning and artificial intelligence on games and medical image.



**GUOXIA YU** (Member, IEEE) received the Ph.D. degree from the Peking University Health Science Center, China, in 2005. She is currently a Chief Physician with the Department of Stomatology, Beijing Children Hospital, Capital Medical University, National Center for Children's Health. She is in charge of the Department of Stomatology. Her research interests include oral and maxillofacial surgery and respiratory diseases.



**QINGMING L1** received the B.S. degree from the Nanjing University of Science and Technology ZiJin College, China, in 2019. He is currently an Algorithm Engineer with the Institute of Automation, Chinese Academy of Sciences. His research interests include reinforcement learning, continual learning, and artificial intelligence on games and medical image.



**QING SHI** received the Ph.D. degree from the Department of Stomatology, Capital Medical University, China, in 1983. She is currently a Chief Physician with Beijing Stomatological Hospital, Capital Medical University. Her research interests include caries diagnosis and prevention and treatment of dental pulp disease and periapical disease.