Precise Dental Staging Method through Panoramic Radiographs Based on Deep Learning

Hongcheng Han[1,2], Shaoyi Du[1]*, Dong Zhang[1,3], Hong Long[1,2], Yucheng Guo[1, 4, 5]

[1]Institute of Artificial Intelligence and Robotics, College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710049, P.R. China

[2]School of Software Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710049, P.R. China

[3]School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710049, P.R. China

[4]Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710004, P.R. China

[5]Department of Orthodontics, Stomatological Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710004, P.R. China

*Corresponding author. E-mail: dushaoyi@gmail.com

2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CAC51589.2020.9327719

***Abstract*—In this paper, we use a deep learning method to realize automatic dental staging, which is an image classification task. Sample imbalance is found to influence the staging accuracy of deep learning method. Aiming at the problem, an optimized method for dental staging is proposed. Firstly, based on the thought of coarse-to-fine, the dental staging task is converted into a binary-classification and two multi-classification tasks to reduce the difficulty of each classification task. Secondly, data augmentation is used in multi-classification to balance the samples in different categories. Besides, the Easy-Ensemble method is used to reduce the influence of sample imbalance in the binary-classification task. According to the experiment results, the dental staging effect of the proposed method is improved in accuracy, recall, and accuracy compared with the unoptimized method, which verifies the availability of the proposed method.**

Keywords—Dental Staging; Image Classification; Deep Learning; Sample Imbalance; Coarse to Fine; Easy-ensemble

I. Introduction

Dental staging is to determine the stage of growth and development of teeth according to the shape of every single tooth in the panoramic radiograph[1]-[2]. It is of great significance in the Demirjian method of age estimation and the diagnosis of orthodontic effects in forensic medicine[3]-[5]. In medicine, From germination to maturity, the development of each tooth is divided into 8 stages, as shown in Fig. 1, each tooth in each stage has corresponding morphological characteristics[6]-[7]. Manual dental staging relies on the experience of the technicians and the work efficiency is limited[8]-[9]. The automatic realization of dental staging by the computer will greatly improve the work efficiency, and also provide a reference for the technicians. In recent years, computer vision technology has been widely used in medical image analysis. Due to the outstanding performance of deep convolutional neural networks in image recognition tasks, medical image recognition based on deep learning has become a research hotspot[10]. This paper applies the deep learning method to realize dental staging, which is a tooth image classification task.

There are some differences between medical image recognition and natural image recognition, which will affect the application of deep learning. In terms of evaluation indicators, natural image recognition application scenarios are relatively diverse, and recognition speed is pursued in many tasks. Medical image recognition is mostly used to assist diagnosis and forensic identification, which are related to the life and health of patients and the investigation and trial of criminal cases[11]-[12]. Therefore, the recognition accuracy is the most important indicator of medical image recognition. In terms of training data, natural images have a wide range of sources and are easy to collect, which provides a very large dataset for training and testing. Most of the medical images are collected from patients by hospitals, and the amount is usually limited. Especially for some rare pathological features, even fewer images can be collected. This brings some special challenges to the application of deep learning in medical image recognition. In the dental staging task of this paper, since the teeth in the early stages of development only appear in children, the number that can be collected is much smaller than that of adults, and most of their teeth have been developed. Therefore, the collected data has the problem of sample imbalance, which will influence the dental staging effect.

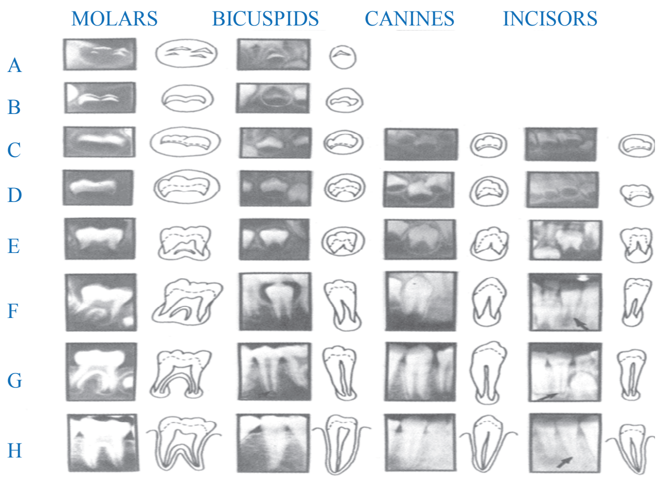

Fig. 1. Morphological characteristics of teeth in different stages

If the amount of data in each category is large, the problem of sample imbalance can be solved by under-sampling the categories with large amounts of data. For tasks with small amounts of data in categories, simply using the under-sampling method makes the amounts of data involved in the training small, and the data utilization rate is low, and it is prone to overfitting, so that accurate classification cannot be achieved. To solve this problem, this paper proposes a dental staging algorithm based on coarse-to-fine classification, data augmentation, and Easy-ensemble[13]. First of all, according to the idea of coarse-to-fine, the categories with small numbers of samples and the categories with large numbers of samples are divided into 2 groups. Performing the fine classification separately in each group, and performing the coarse classification with 2 groups together, the overall classification model is composed of a coarse classifier and 2 fine classifiers. Second, in the fine classification, the data augmentation is used to reduce the impacts of sample imbalance, and in the coarse classification, the Easy-ensemble method is used to further reduce the impacts of sample imbalance. Finally, a precise dental staging model is obtained.

The rest of this paper is arranged as follows. Section 2 details the proposed method for precise dental staging. Section 3 introduces the experiment and analysis of its results, as well as the conclusions drawn.

II. Proposed Method

A. The framework of Our Method

The advancement of the deep convolutional neural network (CNN) structure has made a great contribution to the improvement of the accuracy of image classification, and it is also important to choose the appropriate training method for tasks with different characteristics. In the dental staging task in this paper, due to the limitation of data sources, there are small amounts of samples and 8 categories, and the numbers of samples in 8 categories are imbalanced. In the training, to pursue a smaller loss, the update of network parameters is biased towards categories with a large number of samples, and relatively ignore categories with a small number of samples, which influence the classification effect.

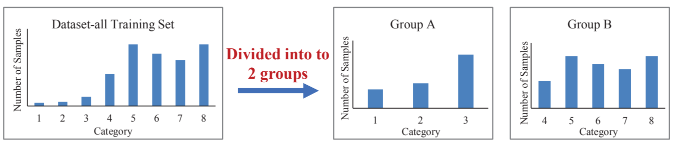

Fig. 2. Dataset is divided into 2 groups

The original dataset is divided into two groups, the categories with small sample numbers are classified as a group, and the categories with larger sample numbers are classified as another group, as shown in Fig. 2, in each group, the sample numbers of categories in each group become more balanced. Performing classification in each group separately, the impacts of sample imbalance is reduced. The classification of 2 groups is coarse classification, and the classification of categories in each group is fine classification, this is a coarse-to-fine method. It is worth noting that, the teeth categories with small numbers of samples are generally in the early development stage of the teeth, and the categories with large numbers of samples are generally in the mature stage. The process of teeth growth and development is continuous, and the characteristics of teeth images of adjacent categories are relatively close. Therefore, the teeth image features difference in the same group are relatively small, and the teeth image features differences across groups are large, which is beneficial for the classification of Group A and Group B.

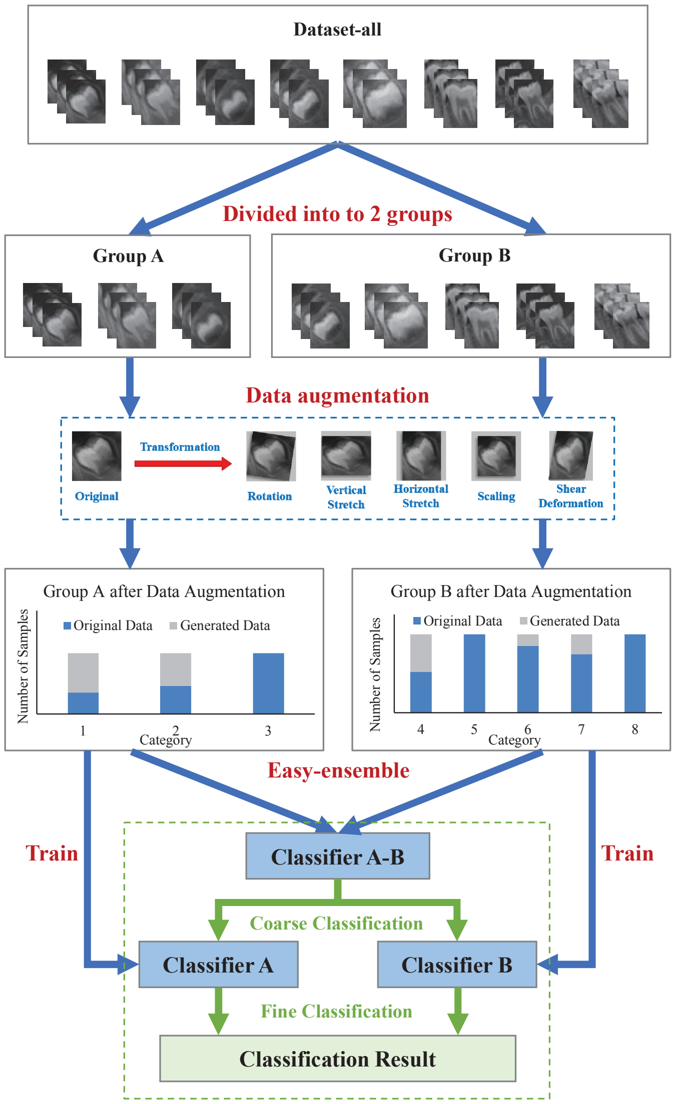

Fig. 3. The framework of the proposed algorithm

Based on the above considerations, a precise dental staging method is proposed. The framework of the algorithm is shown in Fig. 3. Firstly, according to the classification method from coarse to fine, the original 8 categories are divided into group A with a small number of samples in each category and group B with a large number of samples in each category. Secondly, data augmentation is performed in 2 groups. Training the classifier A and the classifier B separately in group A and group B to achieve the fine classification of each category. And using the group A and the group B together to train the classifier A-B to achieve the coarse classification of the group A and the group B. In the training of the classifier A-B, the

Easy-ensemble method is used to reduce the impacts of sample imbalance. classifier A-B, classifier A, and classifier B constitute our dental staging model.

B. Data augmentation

The coarse-to-fine classification method reduces the gap in the numbers of samples of each category in a group, making them within an order of magnitude, but the impacts of sample imbalance still exist. In order to further reduce its impacts, data augmentation is used. Data augmentation is to transform the existing images to a certain extent to generate new images and participate in training when the existing data is limited, which requires that the generated images should be distinguished from the original images to achieve the diversity of the data, but also to be representative of this type of data to achieve the effectiveness of the data augmentation.

Before data expansion, the image is preprocessed. Images are converted into a uniform $r \times r$ size. And To facilitate the solution using the gradient descent method, all pixel values are normalized:

$$\mathbf{X}_1 = \frac{1}{x_{\max} - x_{\min}} (\mathbf{X} - x_{\min} \mathbf{A}) \tag{1}$$

Where $\mathbf{X}$ is the $r \times r$ image matrix, $\mathbf{X}_1$ is the image matrix after pixel value normalization, $x_{\max}$ is the maximum pixel value in $\mathbf{X}$, and $x_{\min}$ is the minimum pixel value in $\mathbf{X}$, and $\mathbf{A}$ is the $r \times r$ matrix in which all elements are 1.

A panoramic radiograph is an image that unfolds a curved surface into a plane. The image will be distorted to a certain extent during conversion. Therefore, the transformation of horizontal stretch, vertical stretch, and shear deformation is added. The size and growth angle of different persons' teeth are slightly different, so the transformation of image scaling and rotation is added. The above transformations are all performed on the original images with random parameters in a specific range:

$$\mathbf{X}_2 = \mathbf{T}_{\text{strech}} \cdot \mathbf{T}_{\text{shear}} \cdot \mathbf{T}_{\text{scaling}} \cdot \mathbf{T}_{\text{rotation}} \cdot \mathbf{X}_1 \tag{2}$$

Where $\mathbf{X}_1$ is the image matrix before the transformation, $\mathbf{X}_2$ is the image matrix after transformation, $\mathbf{T}_{\text{strech}}$ is the stretch matrix, and $\mathbf{T}_{\text{shear}}$ is the shear matrix, $\mathbf{T}_{\text{scaling}}$ is the scaling matrix, $\mathbf{T}_{\text{rotation}}$ is the rotation matrix.

C. Easy-ensemble in coarse classification

The coarse classification of Group A and Group B is a binary classification problem with sample imbalance. In order to reduce the negative impacts of sample imbalance, we use the Easy-ensemble method to train the classifier A-B. The Easy-ensemble method is proposed by Liu et al. which is a simple ensemble learning method for solving imbalanced classification problems, it is mainly used for binary classification problems with imbalanced samples. As shown in Fig. 4, this method is mainly based on the principle of under-sampling, but it is different from simply under-sampling. The sample number of Group A is much smaller than Group B. Group B is randomly divided into several groups whose sample number is close to Group A, which are expressed as Group B_i (i=1, 2, …, n). Training weak classifiers with Group A and each Group B_i (i=1, 2, …, n) separately, in this way, n weak classifiers are obtained, which are expressed as classifier i (i=1, 2, …, n).

Under the action of the Softmax activation function, the output of each weak classifier is the probabilities of the data belongs to class A and belongs to class B:

$$h(x) = \frac{\exp(f_{CNN,j}(x))}{\sum_{j=1}^{j=m} \exp(f_{CNN,j}(x))} \tag{3}$$

Where m is the number of groups, $f_{CNN,j}(x)$ Is the feature vector calculated by the CNN.

Cross-Entropy is used as the loss function. The prediction output by all weak classifiers are voted and integrated to provide the classification results of strong classifier:

$$H(x) = \frac{1}{\sum_{i=1}^{n} \frac{1}{L_i(x)}} \sum_{i=1}^{n} \left[\frac{1}{L_i(x)} h_i(x) \right] \tag{4}$$

Where $h_i(x)$ is the prediction output by the i th weak classifier, n is the number of the weak classifiers, $L_i(x)$ is the loss of the i th weak classifier in the training set.

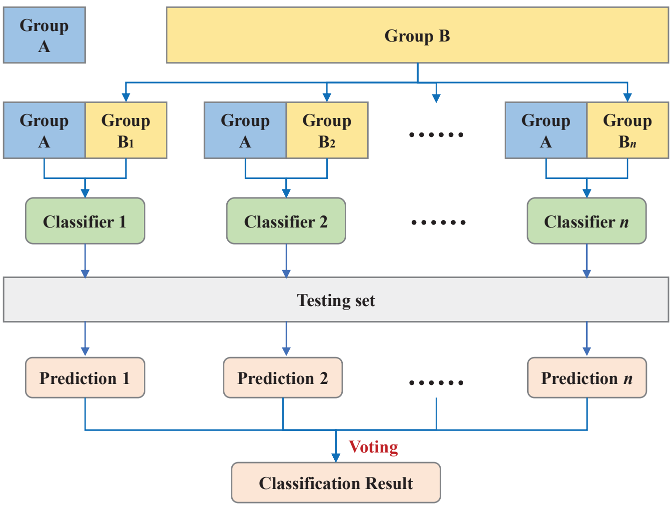

Fig. 4. The framework of the Easy-ensemble method

III. Experiments

A. Experimental Setup

Dataset: The tooth images in the experiments come from 9231 panoramic radiographs of people aged 0-25 years collected by the Dental Hospital of Xi'an Jiaotong University. 8 teeth of the left lower jaw in each panoramic radiograph are staged by professional dentists. Manually dental staging relies on image characteristics of every single tooth, so in the experiments, we need to establish a mapping from a single tooth to its staging result. Based on this requirement, as shown in Fig. 5, each of the 8 teeth of the left lower jaw in a panoramic radiograph is separately cropped and labeled according to the dentists' staging results. After filtering out inappropriate data, 35001 single tooth images with labels are used as the dataset, which is the Dataset-all. The numbers of samples in 8 categories are shown in Fig. 6, 20% of data in each category is used as the testing set, 20% is used as the validating set, and 60% is used as the training set. In the training set, class 1, 2, and 3 that with a small number of samples are classified as Group A, and class 4, 5, 6, 7, and 8 that with a large number of samples are classified as Group B.

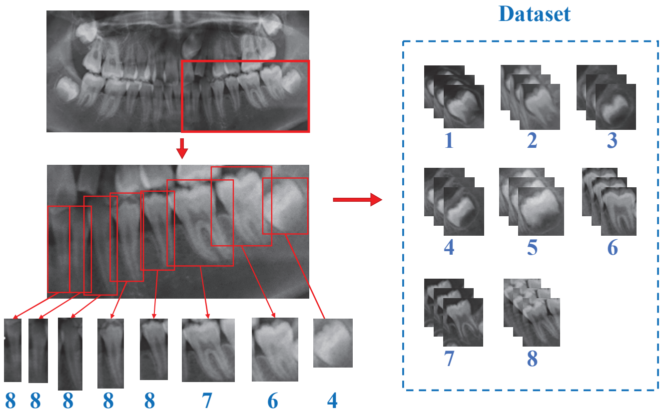

Fig. 5. The process of dataset production

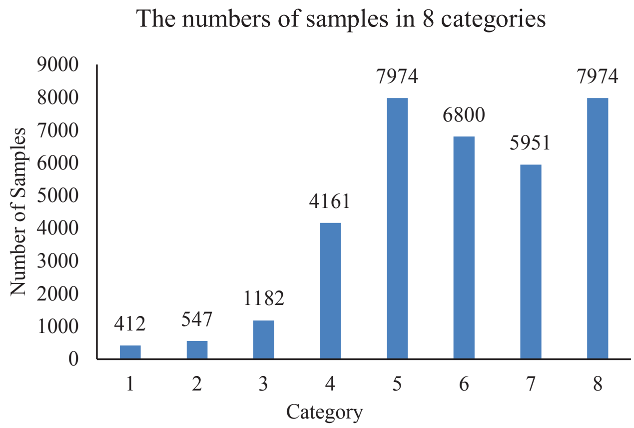

Fig. 6. The numbers of samples in 8 categories in the Dataset-all

Evaluation Metrics: Since it is a multi-classification task, in addition to accuracy, we also calculate the average precision and average recall, which are the average values of precisions and recalls of all categories, as evaluation indicators to analyze the classification results.

The proposed method is implemented by Python 3.6 on both Ubuntu18.04 which has the RTX2080-12GB and CUDA-10.0. All images are resized to 224 × 224. The models use iterations of 50 epochs, the batch size of 32, the learning rate of 0.001, and the learning rate decay of 0.1.

B. *Classification Results*

In the method proposed in this paper, the data augmentation method is used in the fine classification, and the Easy-ensemble is used in the coarse classification to reduce the impacts of sample imbalance. VGG16, ResNet50, and ResNet101 Networks are used as models. In the following, the experimental results are displayed and analyzed from three aspects: fine classification, coarse classification, and the classification effect of the overall classification. Unoptimized method and ours are compared.

First, for the fine classification models trained with and without data augmentation, the comparison of the classification results on the testing set is shown in TABLE I. . It can be seen that in the experiments of the three networks, the classification performance of the model trained with data augmentation improves in the three evaluation indicators of average precision, average recall, and accuracy, compared to the unoptimized one.

TABLE I. COMPARISON OF THE FINE CLASSIFICATION RESULTS

Network	Group	Method	Average Precision	Average Recall	Accuracy
ResNet101	A	Unoptimized	0.9006	0.8602	0.8974
		Data Augmentation	**0.9161**	**0.9127**	**0.9127**
	B	Unoptimized	0.8347	0.8023	0.8254
		Data Augmentation	**0.8496**	**0.8489**	**0.8489**
ResNet50	A	Unoptimized	0.8756	0.8583	0.8880
		Data Augmentation	**0.9081**	**0.9061**	**0.9061**
	B	Unoptimized	0.8157	0.7924	0.8213
		Data Augmentation	**0.8398**	**0.8391**	**0.8391**
VGG16	A	Unoptimized	0.8984	0.8843	0.9089
		Data Augmentation	**0.9114**	**0.9089**	**0.9160**
	B	Unoptimized	0.8334	0.8398	0.8438
		Data Augmentation	**0.8446**	**0.8438**	**0.8486**

Second, in the coarse classification, Group B is divided into 10 smaller groups to work the Easy-ensemble method. It can be seen from the experimental results in TABLE II. that the unoptimized method has been able to achieve a good classification effect, and the Easy-ensemble method further improves it.

TABLE II. COMPARISON OF THE COARSE CLASSIFICATION RESULTS

Network	Method	Average Precision	Average Recall	Accuracy
ResNet101	Unoptimized	0.9345	0.9524	0.9826
	Easy-ensemble	**0.9461**	**0.9673**	**0.9866**
ResNet50	Unoptimized	0.9034	0.9525	0.9765
	Easy-ensemble	**0.9039**	**0.9529**	**0.9766**
VGG16	Unoptimized	0.9032	0.9527	0.9763
	Easy-ensemble	**0.9241**	**0.9562**	**0.9811**

Finally, since the classification effects of both the fine classification and the coarse classification are improved, the overall classification effect is unsurprisingly improved. The experimental results of the unoptimized method and ours are shown in TABLE III. In the experiments of the three networks of VGG16, ResNet50, and ResNet101, our method improves the classification effect in three evaluation indicators, precision, recall, and accuracy, which rules out the impacts of the network structure to a certain extent and verifies the effectiveness of our method.

TABLE III. COMPARISON OF THE OVERALL CLASSIFICATION RESULTS

Network	Method	Average Precision	Average Recall	Accuracy
ResNet101	Unoptimized	0.8044	0.7757	0.8157
	Ours	**0.8511**	**0.8316**	**0.8410**
ResNet50	Unoptimized	0.7889	0.7564	0.8043
	Ours	**0.8420**	**0.8213**	**0.8314**
VGG16	Unoptimized	0.7826	0.7658	0.8047
	Ours	**0.8233**	**0.8110**	**0.8291**

In order to further analyze the impacts of our method on reducing sample imbalance and improving the effect of classification, the recall results of 8 categories in the experiment of the ResNet101 network, are analyzed as examples. As shown in Fig. 7, in the 8 categories, our method greatly improves the classification recall rate of class 2, 3, 6, 7. Based on the analysis of the number of samples in each category in Dataset-all, for class 2, 3, 6, 7, there are large gaps with the numbers of their adjacent categories. As a result, when the parameters are updated during training, the networks are more inclined to classify the samples into adjacent categories on the class-to-class boundary[14], resulting in a lower recall rate of the class with a smaller number of samples. Our method is dedicated to reducing the impacts of sample imbalance. Therefore, the classification recall rate of these categories is greatly improved. Although the recalls of class 1 and class 8 decrease slightly, the overall result shows that the classification effect is improved.

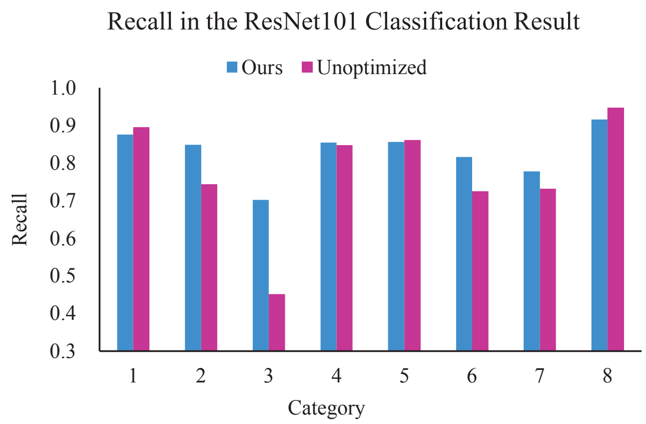

Fig. 7. Comparison of unoptimized method and ours in the recall of ResNet101 classification Result

C. Conclusion

When the total amount of data is limited, the problem of sample imbalance will negatively affect the classification results. The development of teeth is continuous, and the image features of the teeth of adjacent categories are relatively similar, and when the number of samples in an adjacent category is large, the classification of the small number of samples results in a low recall and negatively affects image classification effect.

Aiming at the problem of dental staging, the method proposed in this paper effectively improves the classification effect. Using a coarse-to-fine classification method, according to the number of samples, categories are divided into 2 groups, and the overall classification is achieved through coarse classification and fine classification, which can effectively reduce the impacts of sample imbalance. The data augmentation and Easy-ensemble method can improve the classification effect in the case of sample imbalance to a certain extent. The method proposed in this paper is of constructive significance to solve the other sample imbalanced multi-classification problems with limited data, which are similar to the dental staging problem in this paper.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant Nos. 61627811 and 61971343, the Key Research and Development Program of Shaanxi Province of China under Grant No. 2020GXLH-Y-008, and the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2020JM-012.

REFERENCES

[1] Arany S, Iino M, Yoshioka N. Radiographic survey of third molar development in relation to chronological age among Japanese juveniles[J]. Journal of Forensic Science, 49(3): JFS2003372-5 (2004).

[2] Litsas G, Lucchese A. Dental and chronological ages as determinants of peak growth period and its relationship with dental calcification stages[J]. The open dentistry journal, 10: 99 (2016).

[3] Demirjian A, Goldstein H, Tanner J M. A new system of dental age assessment[J]. Human biology, 211-227 (1973).

[4] Djukic K, Zelic K, Milenkovic P, et al. Dental age assessment validity of radiographic methods on Serbian children population[J]. Forensic science international, 231(1-3): 398. e1-398. e5 (2013).

[5] McKenna C J, James H, Taylor J A, et al. Tooth development standards for South Australia[J]. Australian dental journal, 47(3): 223-227 (2002).

[6] Qudeimat M A, Behbehani F. Dental age assessment for Kuwaiti children using Demirjian's method[J]. Annals of human biology, 36(6): 695-704 (2009).

[7] Štern D, Payer C, Urschler M. Automated age estimation from MRI volumes of the hand[J]. Medical image analysis, 58: 101538 (2019).

[8] Celikoglu M, Erdem A, Dane A, et al. Dental age assessment in orthodontic patients with and without skeletal malocclusions[J]. Orthodontics & craniofacial research, 14(2): 58-62 (2011).

[9] Mack K B, Phillips C, Jain N, et al. Relationship between body mass index percentile and skeletal maturation and dental development in orthodontic patients[J]. American Journal of Orthodontics and Dentofacial Orthopedics, 143(2): 228-234 (2013).

[10] Beucher A, Møller AB, Greve MH. Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark[J]. Geoderma, 2019, 352(1): 52-58.

[11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132-7141 (2018).

[12] Tan M, Le Q V. Efficientnet: Rethinking model scaling for convolutional neural networks[J]. arXiv preprint arXiv:1905.11946, (2019).

[13] X. Y. Liu, J. Wu and Z. H. Zhou. Exploratory Undersampling for Class-Imbalance Learning[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2009, 39(2): 539-550.

[14] Cao Q, Shen L, Xie W, et al. Vggface2: A dataset for recognising faces across pose and age[C]. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, pp. 67-74 (2018).