Research on Radius Grade Optimization Method Based on Deep Clustering

Keji Mao
Zhejiang University of Technology
Hangzhou, China
maokeji@zjut.edu.cn

Jianan Wu
Zhejiang University of Technology
Hangzhou, China
wujianan@zjut.edu.cn

Yan Mao
Zhejiang University of Technology
Hangzhou, China
maoyan@zjut.edu.cn

Yantao Shao
Zhejiang University of Technology
Hangzhou, China
211122120092@zjut.edu.cn

Kai Fang
Zhejiang A&F University
Hangzhou, China
Kaifang@ieee.org

Wei Wang
Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Artificial Intelligence Research Institute
Shenzhen MSU-BIT University
Shenzhen,China
ehomewang@ieee.org

Abstract*—Bone age is an important criterion for evaluating the growth and development of adolescents and children in the medical field. Domestic bone age experts often use the method of first distinguishing bone grades and then calculating bone age when evaluating bone age. However, there is a significant difference in developmental characteristics between adjacent levels of certain hand bones in the established level classification, which can lead to errors in bone age assessment. It is obviously unrealistic to use manual subdivision of levels based on the existing level division. Therefore, this article takes the radius as an example based on the CHN method, which is the standard for evaluating bone age in Chinese children. We optimize the CHN reference map using deep clustering methods based on data science. In this method, we use a joint loss function to jointly optimize deep learning and clustering. In addition, this article uses interpolation method to assign corresponding scores to the newly generated levels after optimization. This article uses a hierarchical recognition model to conduct experimental validation on 1000 adolescents and children. The average difference between the results obtained using the undivided radial atlas in the experiment and the true value was 1.69 months, and the difference after using the subdivided atlas was 0.37 months, which improved the accuracy by 1.32 months.

***Keywords*—Bone age assessment， CHN method， Deep learning， Clustering algorithm， Interpolation method**

I. INTRODUCTION

The most commonly used method for evaluating bone age is to first distinguish the grades of each bone in the fracture area, and then calculate the final bone age based on the grade score. China has proposed the Chinese 05 method [1] and the Chinese wrist bone development standard - the CHN scoring method [2] based on the developmental characteristics of Chinese children and adolescents. The CHN method has more advantages in height prediction compared to other bone age assessment methods, so this article selects the CHN method as the standard for bone age evaluation [3].

The traditional bone age assessment method has problems such as long interpretation process and unstable results[4]. In recent years, research combining medicine and computers has rapidly expanded, and various intelligent bone age assessment methods have emerged in the field of bone age assessment.In 2021, He et al. [5] compressed the image and used a combination of channel attention mechanism and ResNet network for feature extraction and recognition, resulting in an average absolute error of 6.04 months for bone age assessment. In 2022, Hui et al. [6] extracted 18 ROI regions, extracted their features and concatenated them with the features of the entire hand bone, and then trained them in the recognition network. The average error between the final bone age evaluation results and the results provided by experts was 5.33 months. In 2023, Wang et al. [7] cropped the image into small blocks, extracted features from each block, and then fused the most important modules, resulting in a result of 4.17 months.

However, most of these intelligent evaluation methods use average absolute error as the evaluation standard, which cannot guarantee the reliability of the evaluation values in practical clinical applications. In 2021, He et al. [8] segmented 14 ROI regions and performed level recognition. The accuracy of this method's evaluation results and expert results was 96.23% within the 1-year range, and 64.72% within the 0.5 year range. The evaluation standard with a tolerance of 0.5 years old can be used for auxiliary clinical applications, but the accuracy is too low. In 2023, Mao et al. [3] detected and cut 14 ROI regions, and optimized them based on the CHN reference map. This method reduced recognition error by adding a new level to the original hand bone level, and the results showed an accuracy of 94.6% in the 0.5 year old range. This method proves the feasibility of subdividing levels, but its drawback is that the newly added levels are manually selected and do not fundamentally utilize data-driven solutions.

This article addresses the above issues and takes the radius with a large score difference as an example. Based on data science, a hierarchical subdivision method based on deep clustering is proposed. The system flowchart of this article is shown in Figure 1, and the main contributions are as follows:

1) Use Principal Component Analysis (PCA) to reduce the dimensionality of the original data, then use the self encoder in deep learning methods to extract features of the radius, and then classify it using clustering algorithm (K-means) to complete the subdivision of the radius level.

2) In order to extract more useful features, this paper proposes to add the reconstruction loss function and clustering loss function in the autoencoder, and concatenate the two parts through joint loss. Use training to jointly optimize clustering algorithms and feature extraction networks.

3) This article uses the idea of transfer learning to extract features from the subdivided levels using MobileNet-V2, calculates the Euclidean distance between the subdivided levels and adjacent levels, standardizes this distance, and uses interpolation to assign corresponding scores to the new levels generated after subdivision.

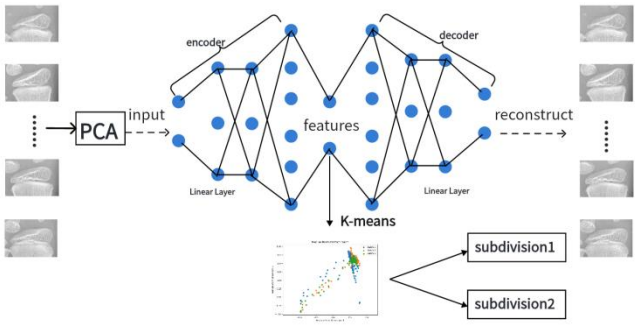

Fig.1. System flow chart

II. Data Introduction

A. Data Sources

The convolutional neural network has been widely used for medical image segmentation because its hierarchical learning and characterization capabilities are better than traditional algorithms[9].The data used in this article was obtained by cutting the region of interest (ROI) of X-ray films using the Faster R-CNN [10] algorithm. Firstly, input the image into the VGG16 network to extract features and generate feature maps. Then, input the feature map into the regional recommendation network to obtain the target score and bounding box position for each anchor point. Filter out a certain number of high-quality ROI regions and map these candidate boxes onto the feature map to generate a feature matrix. Next, the feature matrix is scaled to a uniform size through pooling layers, and the category probability and bounding box position of the ROI region are predicted through a series of fully connected layers. We will input the X-ray films evaluated by experts into the model and classify them based on the developmental stages of the reference bone, providing convenience for the analysis and research of intelligent bone age assessment in the future.

B. Differences in Bone Development

In the CHN method, in order to separate the continuity and uninterrupted development of bones, bones that meet the representative developmental indicators provided by experts are classified into one category, which is the process of grading. However, according to recent research in the fields of medicine and computer science, there is a significant gap in developmental indicators between adjacent levels in this classification standard.

This article takes the radius as an example. When the development level of the radius is 5, the development indication description is that the proximal edge of the epiphysis can distinguish between the palmar and dorsal sides, or the proximal edge is a clear irregular dense white line. The corresponding image description is shown in Figure 2 (left). When the level of the radius is 6, the developmental indication is that the ulnar edge can distinguish between the palmar and dorsal sides (the ulnar notch of the radial epiphysis begins to ossify), and the proximal edge is in an arc or arch shape. The corresponding image description is shown in Figure 2 (right).

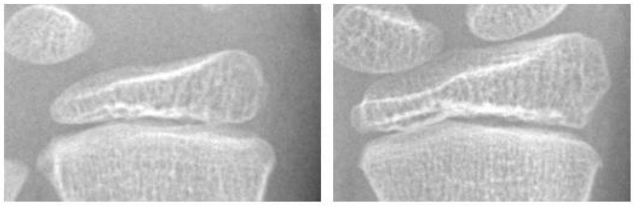

Fig.2. Radial level 5 and level 6

C. Score Difference

The reason for the finer classification of levels in this article is due to the significant difference in scores between adjacent level s of certain bones, which directly affects the accuracy of the final results of bone age assessment. Taking the radius as an example, the grade scores are shown in Table 1. Compared to other levels, there is a significant difference in scores between levels 5, 6, and 7, with a score difference of 12 and 15 points for boys and 12 and 10 points for girls, respectively. Taking male students as an example, the maturity score and bone age control table (partial) of the CHN method were compared, as shown in Table 2. We found that a difference of 4 points corresponds to a bone age difference of 0.1 years, which is 1.2 months. So this article mainly focuses on subdividing new levels between these three levels. Other bones with significant differences in scores also have this problem, as these errors accumulate one by one, leading to significant biases in the final bone age assessment.

TABLE I. Scoring Table for Staging of Radial Development

level	1	2	3	4	5	6	7	8	9
male	17	30	39	49	59	71	87	93	94
female	17	30	42	49	57	69	79	83	84

TABLE II. Score and Bone Age Comparison Table

Bone age (year)	score
10.2	657
10.3	661
10.4	665
10.5	669

We use the interpolation method mentioned in the algorithm design below to evaluate the newly added levels 5_5 and 6_5 and obtain corresponding scores. The segmented radial development staging score table is shown in Table 3. The difference in scores between the segmented level and the original level is 6, 6, 8, and 8 points, respectively. This greatly reduces the difference in scores between grades and improves the accuracy of bone age assessment.

TABLE III. Development Staging Score Table (Subdivided)

level	5	5_5	6	6_5	7
male	59	65	71	79	87
female	57	63	69	74	79

III. ALGORITHM DESIGN

A. Algorithm Overview

The algorithm design of this article mainly includes first using Principal Component Analysis (PCA) [11] to reduce the dimensionality of the original data, and then using an autoencoder network to extract features of the radius. Connect the feature extraction network and K-means [12] clustering network through a joint loss function, and optimize the network parameters. Finally, use interpolation to assign values to the newly added levels.

B. Principal Component Analysis (PCA)

PCA (Principal Component Analysis) is a commonly used dimensionality reduction technique used to convert high-dimensional data into low-dimensional representations while preserving the most important features. The implementation process of this method in the text is as follows.

Data standardization: Zero mean is applied to each feature, which means subtracting the average value of that feature. The averaging formula is as follows:

$$(x - \text{mean})/\text{std} \tag{1}$$

Where x is the raw data, mean is the mean, and std is the standard deviation.

Calculate covariance matrix: Calculate covariance matrix from standardized data to describe the relationship between various features. The formula for calculating the covariance matrix is as follows:

$$C = (1/n) * X\hat{}T * X \tag{2}$$

In the formula, C is the covariance matrix, n is the number of samples, X is the zero mean data matrix, and X ^ T represents the transposition of X.

Eigenvalue decomposition: Perform eigenvalue decomposition on the covariance matrix to obtain eigenvalues and corresponding eigenvectors.

$$C = V*D*V\hat{}T \tag{3}$$

Among them, V is the matrix composed of eigenvectors, D is the diagonal matrix, and the elements on the diagonal are the eigenvalues.

Select Principal Component: Based on the size of the eigenvalues, select the largest k eigenvalues and corresponding eigenvectors as the main components, where k is the dimensionality after dimensionality reduction. The value of k here is 4096.

Dimensionality reduction transformation: Projects the original data onto the selected main components to obtain the reduced feature vectors. The dimension transformation formula is as follows:

$$Y = X*W \tag{4}$$

Among them, Y is the dimensionality reduced feature vector matrix, X is the original data matrix, and W is a matrix composed of k selected feature vectors.

C. Self Encoder Network

Autoencoder [13] is an unsupervised learning neural network model aimed at reconstructing input data by learning feature representations of data. It includes encoder and decoder parts. The encoder section consists of three modules, each containing linear layers with different shapes, and a ReLU activation operation is added after each linear layer.As network complexity increases, the network management becomes an extremely challenging task[14]. The output dimensions of these three linear layers are set to 500,500 and 2000, respectively. The features obtained from the image processed by the encoder are the features used for subsequent clustering. The decoder section is also composed of three modules, each containing linear layers with different shapes. After each linear layer, a ReLU activation operation is added. The output dimensions of these three linear layers are set to 2000,500 and 500 respectively, and the final output is the reconstructed image.

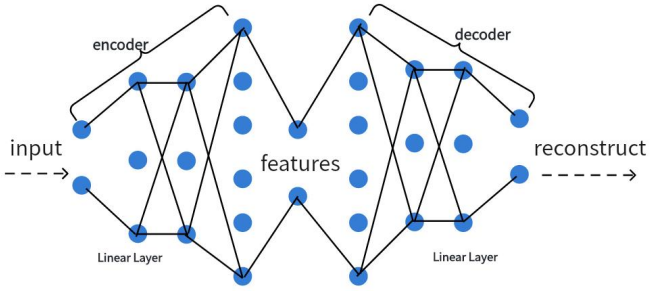

Fig.3. Encoder network structure diagram

The loss function in this section is used to measure the difference between the reconstructed output and the original input. By minimizing the loss function, the autoencoder can learn to extract meaningful features of the input data and achieve data dimensionality reduction and denoising functions. The loss function expression used in this article is as follows:

$$L1 = \sum_{i=1}^{N} \ell\big(g\big(f(x_i)\big), x_i\big) \tag{5}$$

D. K-means Algorithm

The k-means clustering algorithm is an iterative clustering analysis algorithm that involves pre dividing data into K groups, randomly selecting K objects as the initial clustering centers, and then calculating the distance between each object and each seed clustering center. Each object is assigned to the nearest cluster center. The cluster centers and the objects assigned to them represent a cluster. Each time a sample is assigned, the cluster center of the cluster is recalculated based on the existing objects in the cluster. This process will continue to repeat until a certain termination condition is met. The termination condition can be that no (or minimum number of) objects are reassigned to different clusters, no (or minimum number of) cluster centers change again, and the sum of squared errors is locally minimized. The formula commonly used in the K-means algorithm to calculate data points is Euclidean distance [15], which is as follows:

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{6}$$

The x1, x2, y1, and y2 in the formula represent the coordinate positions of the two points, respectively.

The K-means algorithm is not competent when faced with data with more image features. As the dimension of the feature space increases, the distance between data points

becomes more sparse, which may affect the clustering results. Therefore, before clustering high-dimensional data, it is necessary to perform feature selection or dimensionality reduction processing to eliminate redundant features and improve clustering performance. After clustering, in order to combine with the reconstruction function of the autoencoder section for joint loss, we used the loss function shown below in this section:

$$L2 = \sum_{i=1}^{N} \|f(x_i) - Ms_i\|_2^2 \quad (7)$$

Among them is the allocation vector of data point i, where there is only one non-zero element. This means it is a sparse vector used in clustering algorithms to represent which cluster data point i is assigned to. The algorithm uses allocation vectors and centroid matrix M to determine which cluster data point i belongs to.

E. Joint Loss Function

The outstanding contribution of this article is to combine the clustering loss of the clustering algorithm with the reconstruction loss of the autoencoder part, and conduct joint optimization.The joint loss function here fully considers the relationship between multiple tasks or objectives and combines them for optimization. Through joint training, different tasks can jointly utilize these shared features, improving the generalization ability and effectiveness of the model. In addition, the model learns more robust and useful feature representations, thus exhibiting better performance on different tasks. The joint loss function is the weighted addition of the losses in formulas (5) and (7) above, as shown below:

$$\min_{W,Z,M,\{s_i\}} \sum_{i=1}^{N} \left(\ell\big(g\big(f(x_i)\big), x_i\big) + \frac{\lambda}{2}\|f(x_i) - Ms_i\|_2^2 \right) \quad (8)$$

Here is a hyperparameter used to balance reconstruction error and K-means clustering error, with a value greater than or equal to 0. In this article, the value is 1. After the joint loss mentioned above, the parameters in the model will be updated in a direction that is more conducive to clustering.

F. Interpolation Method Evaluation

This method extracts features from the segmented levels using the MobileNet-V2 [16] network. Here, taking level 5 of the radius as an example, the clustered level 5_5. Extract features from images at levels 5 and 6, respectively. The idea of using transfer learning [17] only uses the network structure and parameters of its feature extraction part. After feature extraction for each image, 1000 dimensional data is obtained. The extracted features from image data in different folders are saved in different feature vector lists to obtain three vector lists. First, they are converted into an array form, and then Euclidean distance is calculated. The formula for Euclidean distance is shown in formula (6), and the distances obtained from each list are stored in their respective lists to obtain the distance between these three vectors. Then standardize the obtained Euclidean distance using the following standardization formula:

$$f(x) = (x - xmin)/(xmax - xmin) \quad (9)$$

Find xmin and xmax from the three lists, calculate the difference between the original distance value x and the minimum distance value xmin (x-min), and divide this difference by the difference between the maximum distance value xmax and the minimum distance value xmin (xmax - xmin). f (x) is the normalized distance.

Based on the standardized distance mentioned above and using the interpolation method [18] formula, the corresponding score for the new level can be obtained. Taking the first element in the three lists as an example, due to only the newly added level 5_5 does not have a corresponding score, so it is set to x. The first element in the other two lists is set to x1 and x0, and the values of x1 and x0 correspond to the level scores of level 5 and level 6, respectively. The following formula is used:

$$f(x) = f(x0) * (x1 - x)/(x1 - x0) + f(x1) * (x - x0)/(x1 - x0) \quad (10)$$

According to this formula, level 5 can be calculated. The scores corresponding to all elements in the 5 list are averaged to obtain the final level score. The score table after subdivision based on the above algorithm is shown in Table III.

IV. EXPERIMENTAL ANALYSIS

A. Level Subdivision Experiment

*a) Data set:*To ensure that the training data can fully cover the sample space and more easily improve the model's expression and generalization abilities. This article selects 6000 radial level 5 and 6 for level subdivision training. Save the segmented bones according to their levels, and then randomly select 1000 images from each level for interpolation evaluation.

*b) Experimental Result:*The second part of the article mentions that extracting too few features may not fully capture the information of the data and may not achieve the clustering effect we want. If the number of extracted features is too large, the dimensionality of the data space also increases. In high-dimensional space, the distance between data points becomes more sparse, making it difficult to cluster. Therefore, in the experiment, we set features_num. The num parameter is used to control the output of the number of features. In addition, we selected three batch images in each epoch and visualized their clustering situation. For the convenience of observation, only the first two features of each image are used as two dimensions for display. When the features_num is set to 24,the initial distribution before clustering is shown in Figure 4, and the results after clustering are shown in Figure 5.

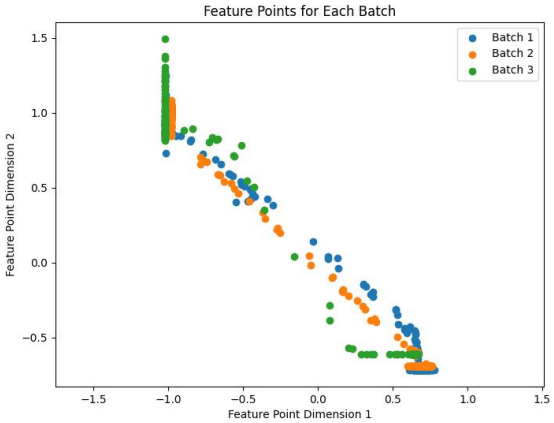

Fig.4. Pre clustering feature distribution map

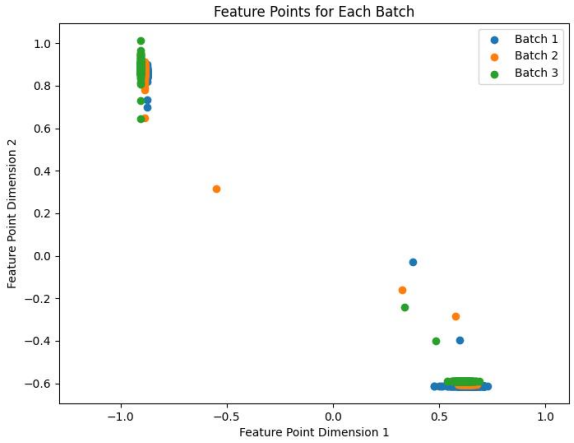

Fig.5. Clustered feature distribution map

From the above figure, it can be concluded that a clear clustering task can be achieved when the number of extracted features is 24. In addition, we validated the clustering performance when the number of features was 10 and 32, respectively, through comparative experiments. When the features_num is set to 10, the distribution of features after clustering is shown in Figure 6.

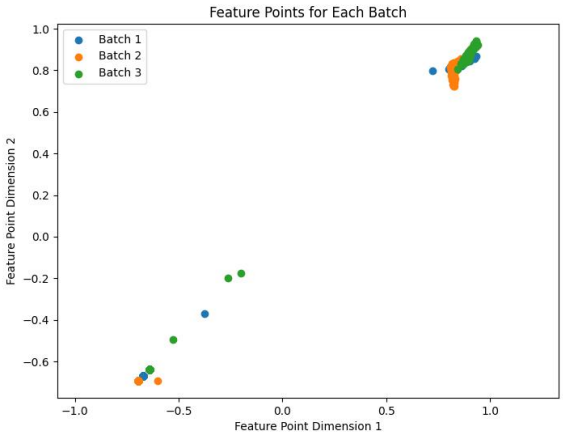

Fig.6. features_ num=10 feature distribution map

There is a significant deviation after clustering, and almost all points are clustered in the same cluster. We consider it as a clustering failure.

We believe that the reason for this result may be that the number of extracted features is too small, as a small number of features can make it difficult for the network to form clusters, resulting in scattered points with no rules to follow.This problem can be solved by increasing the number of features.

When the features_num parameter is set to 32, the distribution of clustered features is shown in Figure 7. The clustering effect is quite impressive, similar to when the value is 24, but considering that lower dimensions can reduce computational and storage costs, lower dimensions can reduce redundant information and improve the model's generalization ability. Taking into account the above factors, we ultimately chose to set the features_num value of num to 24.

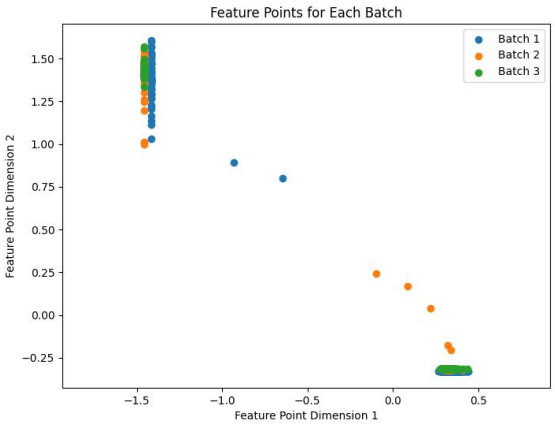

Fig.7. features_ num=32 feature distribution map

B. *Bone Age Assessment Experiment*

*a) Data set:*In order to avoid interference caused by the use of duplicate data during model evaluation, we randomly selected the hand bones of 1000 children and adolescents aged 3.8 to 14 from the database as the experimental dataset. The radii of these samples have not been used in the clustering and subdivision experiments mentioned above, ensuring the rigor of data selection. This includes 621 female students and 379 male students.The distribution of radial grades in male and female students is shown in Figure 8(a) and Figure 8(b).

*b) Experimental Result:*In this experiment, we borrowed the level recognition method proposed by Mao et al. [3]. After fine-tuning the method, train the non segmented and segmented radial data separately. The recognition accuracy obtained through training is 87% and 83%, respectively. The experimental results are shown in Figure 8(a) and Figure 8(b).

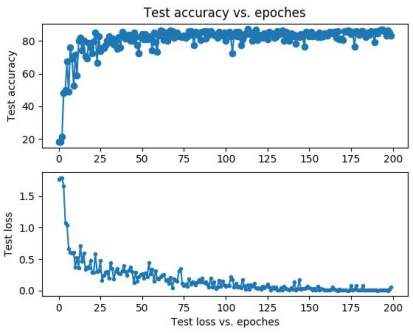

Fig.8(a). Experimental result graph

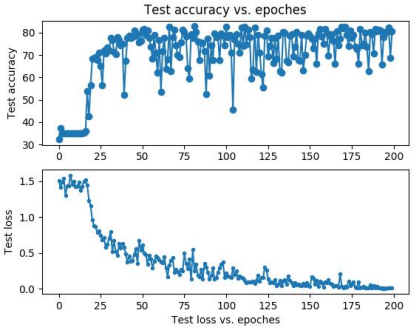

Fig.8(b). Experimental result graph

The trained model was used to recognize the radius of the 1000 children mentioned above. The experimental results showed that when a certain radius in the dataset was misjudged, the level recognized by the segmented model was closer to its true value.

Compare the undivided recognition results obtained with Table 1, and the segmented recognition results with Table 3 to obtain the corresponding scores for the radius in these 1000 samples. Then calculate the sum of the difference between the scores obtained by these two methods and the scores corresponding to the true level of the sample. Divide the sum of score differences by the number of samples to obtain the average error score of the samples. From Table 2, it can be seen that a difference of 4 points corresponds to a bone age difference of 0.1 years, which is 1.2 months. Calculate the average difference in bone age corresponding to the average error score of the sample. The average difference before subdivision was 1.69 months, and after using the subdivision graph, the average difference was 0.37 months, which improved the accuracy by 1.32 months. The experimental results are shown in Table 4.

TABLE IV. COMPARISON TABLE OF AVERAGE DIFFERENCE

	Sum of differences	Mean Error Score	Average difference (month)
Before subdivision	5632	5.632	1.69
After subdivision	1233	1.233	0.37

V. CONCLUSION

This article takes the radius as an example in the CHN method of the Chinese children's bone age assessment standard, and subdivides the grades 5 and 6 in the radius that have a significant difference in maturity and scores between adjacent grades. Using a combination of deep learning and clustering algorithms to segment adjacent bone grades with significant differences, a more accurate atlas was obtained. The experimental results showed that the average difference between the results obtained using the undivided radial atlas and the true value was 1.69 months, and the difference after using the subdivided atlas was 0.37 months. This result verifies the effectiveness of the method and can be used to assist clinical diagnosis, which has significant implications for improving the accuracy of bone age assessment.

References

[1] Zhang Shaoyan et al., Standard for the Development of Wrist Bones in Chinese People - Zhonghua 05. III. Long term Trends in Bone Development in Chinese Children[J].Chinese Journal of Sports Medicine, 26(02): 149-153(2007).

[2] Zhang Shaoyan et al.,Standard for the development of wrist bones in Chinese people - CHN method[J].Sports Science, 1993(06): 33-39(1993).

[3] Mao Keji,Wu Kunxiu et al., Research on the CHN Intelligent Bone Age Assessment Method for Optimizing Reference Graph Development Indications[J].Journal of Electronics and Information Technology,45(03):958-967(2023).

[4] Yi Zhang et al., Progress in the application of artificial intelligence in the evaluation of bone age imaging in children[J], Chinese Journal of Radiology, 57(4) : 420-423(2023).

[5] J. He and D. Jiang, “Fully Automatic Model Based on SE-ResNet for Bone Age Assessment,” in IEEE Access, vol. 9, pp. 62460-62466, (2021).

[6] Hui Q, Wang C, Weng J, Chen M, Kong D. A Global-Local Feature Fusion Convolutional Neural Network for Bone Age Assessment of Hand X-ray Images. Applied Sciences.12(14):7218(2022).

[7] Wang Chong, et al."Attention-based multiple-instance learning for Pediatric bone age assessment with efficient and interpretable." Biomedical Signal Processing and Control 79.P1(2023).

[8] He Ming,Zhao Xudong,Lu Yu,Hu Yi. An improved AlexNet model for automated skeletal maturity assessment using hand X-ray images[J]. Future Generation Computer Systems,121(2021).

[9] W. Wang et al., "Cross-Modality LGE-CMR Segmentation Using Image-to-Image Translation Based Data Augmentation," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 20, no. 4, pp. 2367-2375, 1 July-Aug. (2023).

[10] Li W .Analysis of Object Detection Performance Based on Faster R-CNN[J].Journal of Physics: Conference Series, 1827(1):012085(2021).

[11] SatoIlic Mika."Cluster‐scaled principal component analysis." Wiley Interdisciplinary Reviews: Computational Statistics 14.3(2021).

[12] Zhao Huiling."Design and Implementation of an Improved K-Means Clustering Algorithm." Mobile Information Systems 2022.(2022).

[13] Tripathy, S., Tabasum, M. Autoencoder: An Unsupervised Deep Learning Approach. In: Dutta, P., Chakrabarti, S., Bhattacharya, A., Dutta, S., Shahnaz, C. (eds) Emerging Technologies in Data Mining and Information Security. Lecture Notes in Networks and Systems, vol 490. Springer, Singapore. (2023).

[14] T. Wang, J. Li, W. Wei, W. Wang and K. Fang, "Deep-Learning-Based Weak Electromagnetic Intrusion Detection Method for Zero Touch Networks on Industrial IoT," in IEEE Network, vol. 36, no. 6, pp. 236-242(2022).

[15] Bingcai Zhang et al., Application and Research of Data Mining Method Based on Euclidean Distance Outliers in Audit[J]. China's Management Informatization. 011(013):48-50(2008).

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 4510-4520(2018).

[17] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," in Proceedings of the IEEE, Jan. ,vol. 109, no. 1, pp. 43-76(2021).

[18] Mitra, P.P. Fitting elephants in modern machine learning by statistically consistent interpolation. Nat Mach Intell 3, 378–386 (2021).