2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Data Enhancement and Deep Learning for Bone Age Assessment using The Standards of Skeletal Maturity of Hand and Wrist for Chinese

Yu Lu[1] and Xi Zhang[2] and Liwen Jing[1] and Xianghua Fu[1]

Abstract— **Conventional methods for artificial age determination of skeletal bones have several problems, such as strong subjectivity, large random errors, complex evaluation processes, and long evaluation cycles. In this study, an automated age determination of skeletal bones was performed based on Deep Learning. Two methods were used to evaluate bone age, one based on examining all bones in the palm and another based on the deep convolutional neural network (CNN) method. Both methods were evaluated using the same test dataset. Moreover, we can extend the dataset and increase the generalisation ability of the network by data expansion. Consequently, a more accurate bone age can be obtained. This method can reduce the average error of the final bone age evaluation and lower the upper limit of the absolute value of the error of the single bone age. The experiments show the effectiveness of the proposed method, which can provide doctors and users with more stable, efficient and convenient diagnosis support and decision support.**

I. Introduction

Because bones grow and develop in a predictable manner, clinicians can use measurements of bone growth to determine the biological age of individuals. This is an effective indicator that can be used to assess the biological growth and development of individuals. Bone age is determined by examining the developmental status of individual bones, such as their shape, size, position, and degree of closure. As an authoritative assessment index of biological age, bone age is widely used in clinical medicine, athletic competition, and legal practice.

The most common and widely accepted method of assessing bone age is manual assessment using radiographs of the left hand, including the wrist, palm, and fingers. The two leading methods in the world for determining bone age are the mapping method of Greulich and Pyle (GP) [1] and the method of artificial weighted geometric mean maturity, also known as the Tanner and Whitehouse (TW) method [2], [3]. The GP method estimates bone age by comparing radiographs with images in atlases of children at a given age. This method is simple, but it is also subjective and unreliable. The TW method of determining bone age is complicated, time-consuming, and difficult to apply on a large scale [4].

The most appropriate method for Chinese adolescents and children to evaluate skeletal maturity is a geometric mean method, which estimates maturity based on differential analysis and automatic weighting. The method for determining bone age using this weighting method is known as the CHN method (The Standards of Skeletal Development of Hand and Wrist for Chinese). It accurately represents the characteristics of children in a certain population. This data processing method is more scientifically advanced, which improves the accuracy and consistency of the assessment. In addition, it reduces random errors and is simpler and more efficient than the mapping methods TW3 or GP.

Therefore, this method is referred to for bone age evaluation in this paper. In contrast, traditional artificial bone age assessment performed by a physician has two major drawbacks: (1) The assessment is highly subjective, and accuracy is low unless performed by an experienced physician. The assessment may differ if different physicians assess bone age from the same radiograph or if the same physician assesses bone age from different radiographs at different times. (2) Bone age assessment requires a high level of expertise, rigorous long-term training and is time-consuming.

Deep learning is a branch of machine learning [5], [6] in which algorithms are structured into layers to create an "artificial neural network" that can learn independently and make intelligent decisions. Recent studies show that Deep Learning based Convolutional Neural Networks (CNN) are capable of being used in image object recognition and classification, which improves the performance of many recognition tasks in biomedicine [7], [8], [9], [10]. CNN has been successfully applied to many other problems in medical image analysis. In the field of intelligent bone age assessment, image preprocessing was performed to remove background from hand X-ray images, and then deep learning was used to automatically assess bone age [11]. In [12], the authors proposed a deep automatic model for bone age assessment using a region-based convolutional neural network (R-CNN). In [13], an approach is proposed that includes feature extraction and classification methods. In feature extraction, a neural depth network is used to explore the features of the X-ray image, and the features Local Binary Patterns (LBP) [14] and Glutamate Cysteine Ligase Modifier (GCLM) in the image are extracted. Support vector machine based classification method is used to classify the features.

Since previous research is based on evaluating the bone age by examining the whole hand or a specific feature area, the performance should be improved in general. In this study, an intelligent bone age evaluation method based on multiple levels of regions of interest (ROI) is proposed. This method is an intelligent implementation and improvement of the CHN method. In the CHN method, 14 representative hand bones

[1]Yu Lu, Liwen Jing and Xianghua Fu are with College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China lvyu@sztu.edu.cn, jingliwen@sztu.edu.cn, fuxianghua@sztu.edu.cn

[2]Xi Zhang is with the Information Centre, Shenzhen Technology University, Shenzhen, China zhangxi@sztu.edu.cn

are considered individually and the developmental status of each bone is evaluated based on its shape and texture. Then, the bone age is determined based on the total score of all 14 bones and the bone age comparison table.

The proposed method first selects 14 bones in the X-ray film of the left hand as ROI. Then, the trained CNN uses the ROI of each bone to make an intelligent decision and obtain the probability that each bone is at a certain stage of development. Considering that bone development is a continuous process, using the traditional hard decision (i.e., the stage considered most probable) leads to some variation in the result. The proposed method uses the most likely two-stage probabilities of the network output to calculate the weighted bone score. Then, the bone age is determined by comparing the total score of the 14 bones with the bone age comparison table to improve the accuracy of the bone age evaluation.

This paper is organised as follows. In Section II, we first introduce the deep convolutional neural network and the proposed method. Then, the method is evaluated and analysed in terms of its accuracy in section III. Finally, in section IV, we conclude the paper.

II. Deep Convolutional Neural Network

A. CNN Framework

A CNN is a type of multilayer neural network that is efficient in processing machine learning problems related to images, especially large images. CNNs successfully reduce the dimensionality of image recognition problems that contain a large amount of data through a variety of methods and eventually allow them to be trained. The most typical CNNs consist of a convolutional layer, a pooling layer, and a fully connected layer [15]. The convolutional layer and the pooling layer work together to form multiple convolutions and extract image features layer by layer, and then full classification is performed using multiple fully connected layers.

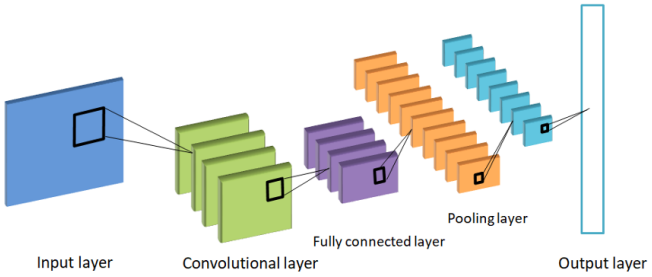


Fig. 1. A typical artificial neural network model.

The operations performed by the convolutional layer can be seen as inspired by the concept of local receptive fields. The main function of the pooling layer is to reduce the data dimension. In practice, a CNN simulates feature differentiation by convolution and reduces the order of magnitude of the network parameters by weight sharing and pooling, and then performs tasks such as classification by conventional neural networks. A CNN usually consists of multiple convolutional layers and pooling layers, with a fully concatenated layer added at the end to form a multilayer artificial neural network, as shown in Figure 1.

TABLE I
Dataset distribution of training and testing sets.

	Children	Adolescents	Middle aged	Elderly
Training set	48	307	680	298
Testing set	12	77	170	74

B. Proposed method

The intelligent method proposed in this study to evaluate bone age based on the CHN method proposed in this study includes three parts: Data preprocessing, data enrichment and bone age evaluation. It is based on training 14 bone-level classifications using the AlexNet framework [16] shown in Figure 2.

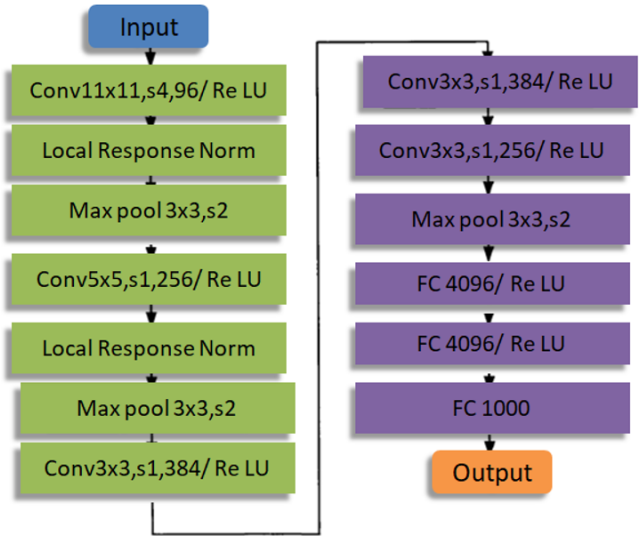
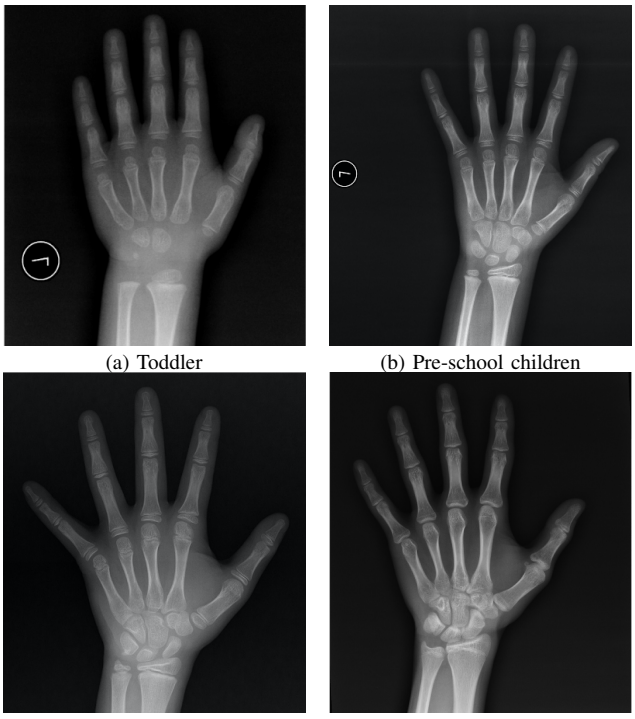


Fig. 2. The AlexNet architecture.

1) Datasets and data preprocessing: First, the CNN was trained to estimate bone age using left hand radiographs from individuals of different ages. The dataset DR was collected from Shengjing Hospital of China Medical University and contains 1666 instances of dates and 1666 radiographs. In the data labelling phase, we divided the dataset into four categories according to bone age: Infants, Preschool-aged children, School-aged children and Adolescents.

The post-calibration data set included 60 infants, 384 preschool-aged children, 850 school-aged children, and 372 adolescents. To perform cross-validation during network training, the data set was randomly divided into two parts: 80% of the data were included in the training set and 20% in the test set. The samples of the DR dataset are shown in Figure 3, and the distribution of the dataset can be seen in Table I.

After the artificial brain was successfully trained, we started to use CNNs to predict bone age. Following the CHN



(a) Toddler (b) Pre-school children

(c) School-age children (d) Adolescents

Fig. 3. Example dataset.

method, we first select and calibrate the ROI of 14 bones in each hand slice: Radius, Os metacarpi I, Os metacarpi III, Os metacarpi V, Phalanx proximalis I, Phalanx proximalis III, Phalanx proximalis V, Phalanx media III, Phalanx media V, Phalanx distalis I, Phalanx distalis III, Phalanx distalis V, Os capitatum, and Os hamtum. Each bone is cut with an appropriate fixed size frame. The size of the cutting frame of each bone is set to include the ROI of that bone in almost every hand slice, but also to include as few interfering areas as possible. Calibration of the 14 bones in each hand slice of the dataset is performed by a team of experts who have a high degree of credibility and accuracy.

2) Data augmentation: The distribution of data in the dataset is uneven and prone to overfitting. Therefore, data augmentation technology is used to increase the size of the dataset and increase the generalisation ability of the network. Considering the characteristics of this training dataset, we have used and improved several popular data augmentation techniques. At the same time, we used an online data enrichment method to reduce the pressure on data storage and improve the richness of the dataset. [17]. In general, data enrichment has improved the generalisation ability and test accuracy of the network model.

3) Bone age prediction: Therefore, the process of bone age assessment using the CHN method is shown in Figure 4.

In this study, it is proposed to use the classification probabilities of the two most probable levels output by AlexNet to calculate the weighted score of the head. The weighted score is calculated as in equation 2.

TABLE II

STAGES OF RADIAL DEVELOPMENT STAGING BY THE CHN METHOD.

Radius	0	1	2	3	4	5	6	7	8	9	10
Male	0	15	28	37	48	55	67	80	93	97	106
Female	0	15	28	40	48	50	63	71	86	89	90

$$S = \frac{S_1 P_1 + S_2 P_2}{P_1 + P_2} \tag{1}$$

where S is the final score of the bone, P_1 and S_1 are the maximum output probability and the corresponding score, respectively, and P_2 and S_2 are the second largest output probability and the corresponding score, respectively. The experiments show that the method CHN can reduce the average error in the final bone age score and the upper limit of the absolute value of the error in the single bone age.

III. EVALUATION

Using the CHN method, the radius was divided into levels from 0-10, and each level was assigned a corresponding score, as shown in Table 2. Each of the remaining bones was scored in a similar manner using a similar method. Finally, the scores of 14 bones were summed to obtain a total score, which was compared with the bone age table of the method CHN to determine the bone age.

As can be seen from Table II, the values between the levels are discontinuous, and the differences are large. However, bone development is a continuous process, and it is not appropriate to determine which of the two adjacent levels is more accurate. The test results show that the bone development level (the level with the highest output probability of the AlexNet) is difficult to determine using the conventional method and has a certain margin of error.

A. Accuracy

In this study, two methods of evaluating bone age, one based on the whole palm bone and the other using the CHN method, are used to train and test the network. The results of all methods are then tested using the same test data set. The bone age assessment method, which uses the whole palm bone, is further divided into the two categories of test classification and regression, which are called full hand classification and full hand regression, respectively. The network model for full hand classification outputs 18 levels corresponding to ages 1-18. When using the full hand classification method to evaluate bone age, the bone age is specified with an accuracy of 0.1 years. The bone age label is approximated to the integer bone age so that the prediction result of the model has an error of ± 0.5 years.

From the literature [18], [19], there is a large margin of error in the manual assessment of bone age. If the same evaluator estimates the same hand bone age at different times, the results will be different. In addition, different evaluators may not evaluate the same hand bone age in the same way. In practice, the statistical estimation error of experienced experts is about 0.25 years. For some younger

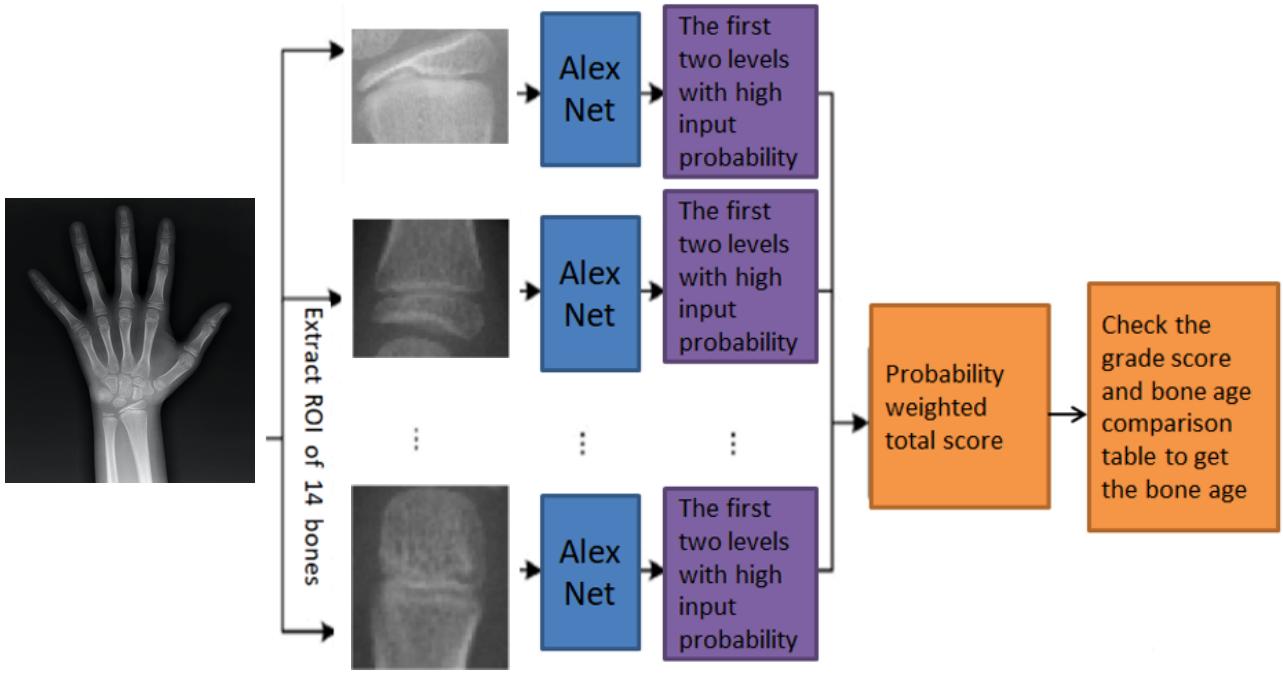


Fig. 4. Flowchart of the process of bone age assessment using the CHN method.

TABLE III
COMPARISON OF THE ACCURACY OF BONE AGE ASSESSMENT USING DIFFERENT METHODS.

Error interval	±0.5	±1.0
CHN method	54.16	85.21
TOP2-CHN method	64.72	96.23
Full hand regression	48.24	78.46
Whole-hand taxonomy	45.78	72.45

evaluators, the error may be as high as 0.82 years. Therefore, in the field of bone age assessment, an error of 1.0 years is often used to measure the performance of automated assessments. If the average error of automated bone age assessment is within 0.5 years, it can be used clinically as a complementary assessment method. The assessment accuracy of each method is listed in Table III.

In general, four different network models have achieved good accuracy rates. Compared with bone age assessment based on the whole hand bone, bone age assessment based on the improved AlexNet model proposed in this study was much more accurate. Compared with the whole hand classification method and the whole hand regression method, the CHN method improves the accuracy rate by about 10% points, and the TOP2-CHN method improves the accuracy rate by about 10% points compared with the CHN classification method.

The accuracy of the methods CHN and TOP2-CHN is higher than that of the all-hands classification method [20] and the regression method [21], and the method TOP2-CHN is the most accurate, especially within the [-1,0, 1,0] age error interval, where the classification model TOP2-CHN achieves an accuracy rate of 96.23%. Since 14 classification models need to be trained in the method used in this study, the model training time increases compared to whole hand bone based age estimation.

However, the results obtained are sufficient to demonstrate the superiority of the TOP2 CHN method. This shows that the CHN method is feasible for segmenting ROI the hand bone image compared to determining the bone age based on the whole hand bone and can significantly improve the accuracy. Most previous work has used whole hand images to assess bone age. Our whole hand regression method is similar to the method recently proposed in the paper IRK+18. In this method, the radiographs of the hand are preprocessed and the AlexNet is used for training. Whole hand classification is also used.

B. Error rate

In addition, the average errors in estimating the bone age of the methods CHN and TOP2-CHN are determined for the 1194 test sets. The average error in estimating bone age using the TOP2-CHN model is 16.95% lower than that of the CHN method. In general, the relevant experts consider that a model system with an average error in bone age estimation of less than 0.5 years can be used as an expert assistance system.

In addition, if the sample size is large enough, the ROC curve can be used in the TOP2-CHN method to determine whether the prediction of bone age is correct. For each data set, there are two categories (yes or no), and the ROC curve is plotted with the false and true positive rates on the horizontal and vertical axes, respectively, as shown in Figure 5.

The false-positive rate is mostly between 0.0-0.4, the true-positive rate of the TOP2-CHN method is mostly between

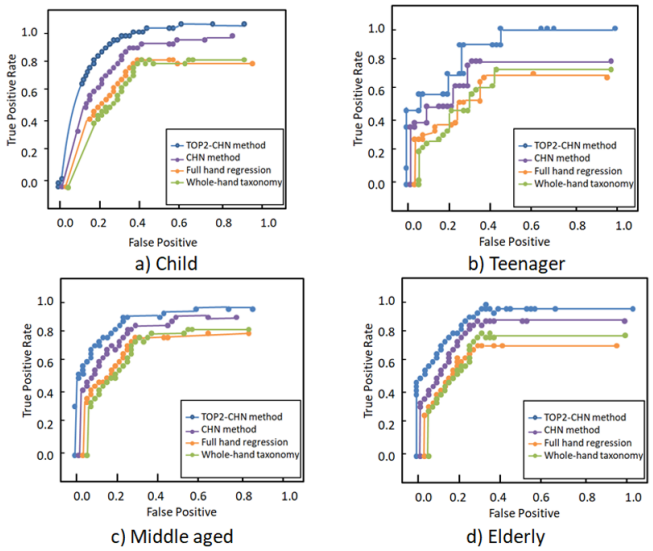


Fig. 5. ROC curves based on various bone age prediction methods.

0.6 and 1.0, and the true-positive rates of the whole hand classification and regression methods are between 0.2 and 0.8, indicating that the TOP2-CHN method can predict bone age more accurately.

IV. Conclusions

A TOP2-CHN bone age estimation is proposed, which combines the CNN model with the features of the output of the classification network model. The experimental results show that the method proposed in this study achieves better results on various indicators, such as the accuracy rate of the absolute error of bone age determination within 1.0 years and the average bone age error, compared with intelligent whole palm bone-based bone age determination. Therefore, it can be developed as an auxiliary system for expert bone age assessment. In this study, a set of automatic bone age detection systems based on a Deep CNN was developed and implemented. The method of using the deep learning model was studied, and the trained bone age detection model was used online to provide assistance to more doctors and users. This method provides stable, efficient and convenient diagnosis and decision support services. The accuracy depends on exact landmark detection for measuring the length. For future work, the use of rations and angles between the landmarks is likely to give better results.

References

[1] W. W. Greulich and S. I. Pyle, *Radiograph Atlas of Skeletal Development of the Hand and Wrist*, 2nd ed. Stanford University Press, 1959.

[2] J. M. Tanner, R. H. Whitehouse, N. Cameron, W. A. Marshall, M. J. R. Healy, and H. Goldstein, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*, 2nd ed. Academic Press, 1983.

[3] G. Tanner, M. J. R. Healy, H. Goldstein, N. Cameron, N. Cameron, and J. M. Tanner, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*, 3rd ed. W.B. Saunders, 2001.

[4] T. Widek, P. Genet, T. Ehammer, T. Schwark, M. Urschler, and E. Scheurer, "Bone age estimation with the Greulich-Pyle atlas using 3T MR images of hand and wrist," *Forensic Science International*, vol. 319, p. 110654, February 2021.

[5] Y. Lu, X. Zhang, X. Fu, F. Chen, and K. K. L. Wong, "Ensemble machine learning for estimating fetal weight at varying gestational age," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI 2019)*, 2019, pp. 9522–9527.

[6] Y. Lu, X. Fu, F. Chen, and K. K. L. Wong, "Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning," *Artificial Intelligence in Medicine*, vol. 102, p. 101748, January 2020.

[7] M. Zhao, Y. Wei, Y. Lu, and K. K. Wong, "A novel u-net approach to segment the cardiac chamber in magnetic resonance images with ghost artifacts," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105623, November 2020.

[8] Y. Lu, X. Fu, X. Li, and Y. Qi, "Cardiac Chamber Segmentation Using Deep Learning on Magnetic Resonance Images from Patients Before and After Atrial Septal Occlusion Surgery," in *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2020)*. IEEE, 2020, pp. 1211–1216.

[9] X. Zhu, Y. Wei, Y. Lu, M. Zhao, K. Yang, S. Wu, Z. Hui, and K. K. Wong, "Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging," *Computer Methods and Programs in Biomedicine*, vol. 199, p. 105914, February 2021.

[10] Y. Lu, H. Liang, S. Shi, and X. Fu, "Lung Cancer Detection using a Dilated CNN with VGG16," in *Proceedings of the 4th International Conference on Signal Processing and Machine Learning (SPML 2021)*. ACM, 2021.

[11] V. I. Iglovikov, A. Rakhlin, A. A. Kalinin, and A. A. Shvets, "Paediatric Bone Age Assessment Using Deep Convolutional Neural Networks," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, ser. Lecture Notes in Computer Science. Springer, 2018, vol. 11045, pp. 300–308.

[12] B. Liang, Y. Zhai, C. Tong, J. Zhao, J. Li, X. He, and Q. Ma, "A deep automated skeletal bone age assessment model via region-based convolutional neural network," *Future Generation Computer Systems*, vol. 98, pp. 54–59, September 2019.

[13] X. Chen, J. Li, Y. Zhang, Y. Lu, and S. Liu, "Automatic feature extraction in x-ray image based on deep learning approach for determination of bone age," *Future Generation Computer Systems*, vol. 110, pp. 795–801, September 2020.

[14] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, March 2009.

[15] S. Lawrence, C. Giles, A. C. Tsoi, and A. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, January 1997.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[17] A. Torralba, R. Fergus, and W. T. Freeman, "80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1958–1970, November 2008.

[18] A. Tristán-Vega and J. I. Arribas, "A Radius and Ulna TW3 Bone Age Assessment System," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 5, pp. 1463–1476, May 2008.

[19] H. H. Thodberg, S. Kreiborg, A. Juul, and K. D. Pedersen, "The BoneXpert Method for Automated Determination of Skeletal Maturity," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 52–66, January 2009.

[20] T. Liang, X. Xu, and P. Xiao, "A new image classification method based on modified condensed nearest neighbor and convolutional neural networks," *Pattern Recognition Letters*, vol. 94, pp. 105–111, July 2017.

[21] M. Ma, Z. Chen, and J. Wu, "A Recognition Method of Hand Gesture with CNN-SVM Model," in *Bio-inspired Computing – Theories and Applications*, ser. Communications in Computer and Information Science. Springer, 2016, vol. 681, pp. 399–404.