

Chapter 6

Skeletal Bone Age Determination Using Deep Learning



Chintamani Dileep Karthik, Chellasami Shrada, and Arjun Krishnamurthy

Abstract In today's clinical diagnosis of dental patients, skeletal maturity assessment is a unique bio-marker. The bone age assessment (BAA) method is used to identify endocrinological as well as growth abnormalities by contrasting the patient's bone age and actual age. A number of approaches for determining skeletal maturity have been devised; however, the two most significant approaches that employ left hand and wrist radiographs are the Tanner-Whitehouse and Greulich-Pyle methods as mentioned in. While these approaches are well known, they are exceedingly time-consuming and need a skilled radiologist who would have to assess the bone age using a hand atlas as a guideline every time. In our paper, we use convolutional neural networks (CNNs) to successfully predict the maturity of bone age from hand X-ray images of patients aged 4–17 years old. The examination effort of radiologists, for example, is a restriction of manual clinical processes. Since the manual methods are subjected to observer variability, developing computer-aided and automated systems for bone age evaluation is advantageous. We compare the existing deep convolutional neural networks (DCNNs) like VGG16, VGG19, inception models to our own custom regression model in this paper.

6.1 Introduction

Medical image analysis is an important component of a diverse variety of diagnostic choices in the healthcare system [1]. In recent decades, digital patient information preservation has allowed machine learning and computer vision to help in the diagnosis and identification of the most serious illnesses. Currently, the majority of healthcare systems rely on radiologists to assess various types of medical imaging. The major issue is the radiologist's limitations in terms of speed and lack of knowledge, resulting in an incorrect diagnosis. Furthermore, financial expenditures become a matter of concern for radiologist training or outsourcing. Following these

C. D. Karthik (✉) · C. Shrada · A. Krishnamurthy
Dayananda Sagar University, Bangalore, India
e-mail: karthikcd7@gmail.com

factors, implementing a trustworthy, precise, scalable, and efficient machine learning technique might considerably enhance radiological image processing [2].

Traditional machine learning approaches, as illustrated in [3], do not give adequate results when dealing with medical imaging, according to experts, as the data must be optimized by a domain expert prior to the deployment of a problem-solving algorithm. However, deep learning, a relatively emerging niche of artificial neural networks, provides for learning as well as the identification of eminent features from an adequate training data set by utilizing the network's increased layers. Deep learning, which includes a multilayered network known as CNN, offers a wide range of applications in evaluating complicated patterns in raw radiological images. Deep CNN (DCNN) frequently necessitates a large amount of terminal resources, therefore, the graphical processing unit (GPU) becomes a handy tool for its execution [4–6]. Furthermore, because of the large number of parameters in the model, effective DCNN training could not be done with a small data set.

Physicians and endocrinologists frequently analyse chronological age and skeletal age as it aids in identification of various disorders which can result in defective development, especially in newborns. The use of bone age assessment (BAA) could be beneficial in the case of predicting the period during which a child will grow, the age when they will reach puberty, and even the maximum height. It is used to track the growth of children undergoing treatment for conditions that impair the same. Following the above assessment of bone age is also very useful for identifying people who do not have proper identification.

6.2 Methodology

The appearance of hand bone radiograph pictures is determined by a variety of factors. In this section, we describe our method for estimating bone age.

6.2.1 Data Set Selection

Although the bulk of studies on this topic used data from the GP digitized atlas of radiographs, the publicly available data set from Kaggle designed for the bone age prediction challenge was used largely because it included up-to-date radiographs. The Radiological Society of North America (RSNA) collected the data, which comprised of approx. 12,000 radiographs of the hand up to the wrist. Each individual radiograph in the collection is labelled with proper bone age by a skilled radiologist. The bones range in age from 1 to 228 months. The data set collection comprises approx. 5000 female and 6500 male radiographs.

6.2.2 Data Pre-processing

Despite a large number of samples available, the resolution, orientation, brightness, and contrast among many radiographs vary significantly. In addition, different factors such as timepieces, plaster casts, and surgical screws, as well as the L or R alphabet, are visible (left or right-hand label). Several images feature left and right hands, as well as hands containing missing fingers. This imbalanced data distribution makes the data pre-processing complex and conventional approaches, such as image segmentation, would not yield suitable results. A sample image of data pre-processing result is shown in Fig. 6.1.

6.2.2.1 Frequency Check and Data Augmentation

Certain age groups in the original data set had far fewer images than others. Excess images were removed to achieve uniformity, and ages with fewer images (e.g. 4, 16, 17) were augmented [7]. The augmentation process included brightening and zooming. Overfitting may occur due to a lack of training data, which relates to the poor performance of the network on test data despite acceptable training results. Successful regression needs data augmentation, which is performed by randomly brightening and zooming pictures.

6.2.2.2 Background Noise Removal

Background noise was removed using the following steps:

- Binary thresholding with a threshold value of 20. Python's OpenCV module was used for this purpose.

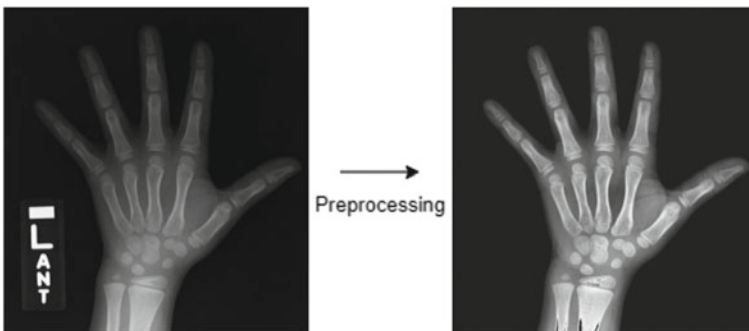


Fig. 6.1 Before and after pre-processing

- Finding contours—which detects colour changes in images and marks them as contours. Placing these contours on plain black image yields new images with less background noise [7].

6.2.2.3 CLAHE

Contrast Limited Adaptive Histogram Equalization (CLAHE) technique employs adaptive histogram equalization (AHE). Due to the black and grey backgrounds of some of the radiographs, normal AHE has a typical problem of producing too much noise in areas that are quite uniform. To prevent this noise, we use CLAHE, which extends standard AHE by preventing over amplification of certain locations as shown in [8]. Python OpenCV package includes this function. Due to the considerable range in brightness and contrast in the radiographs, we applied CLAHE to the training and testing sets to see if it had any positive influence on the model.

6.3 Custom Model

6.3.1 *Model Architecture*

The convolutional layer, pooling layer, dropout layer, and dense layer are the layers in our convolutional neural network. The estimated bone age predicted by the model is the final output. A comprehensive explanation of the layers can be seen in [9]. Refer to Fig. 6.2 for the architecture.

6.3.2 *Approaches*

6.3.2.1 Classification

Our first approach was classification. Each image was assigned to one of the 11 classes (as shown in Table 6.1). When the images were classified, the number of images in each class varied greatly. As discussed earlier, data augmentation was carried out as part of the process of equalizing the number of images in each class. We obtained 2000 images per class after augmentation. As a result, a new data set was created. A custom model was trained on these 22,000 images, yielding a 53.12% accuracy (refer Figs. 6.2, 6.3 and 6.4).

Fig. 6.2 Loss for classification

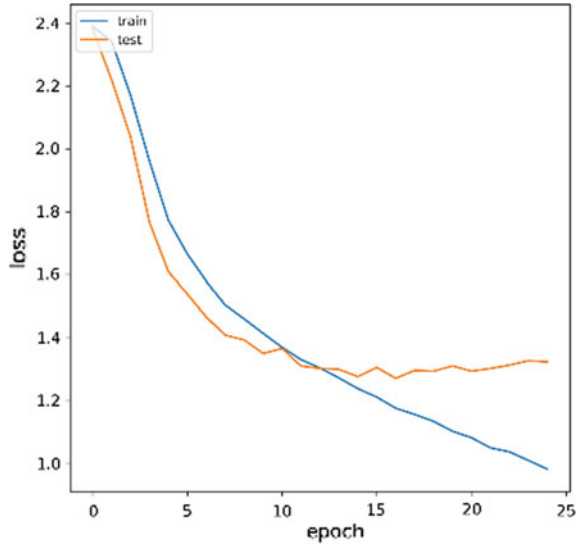


Table 6.1 Segregating into classes

Class name	Age (in years)
c00	4 and below
c01	5 and 6
c02	7
c03	8
c04	9
c05	10
c06	11
c07	12
c08	13
c09	14
c10	15 and above

6.3.2.2 Regression

Regression was our second approach. Regression was a method of predicting a continuous quantity, in which the model predicts a discrete value, but the discrete value in the form of an integer quantity, and the final output of which is the bone age. In this method, 5413 images from the original data set were chosen at random. After pre-processing, the new data set contained 6044 images (5413 images from the original data set + 631 images obtained from augmentation), yielding approximately 450 images for each age. The custom model was trained on the data set with an 85/15 training/validation split. MAPE, MSE, and MAE are evaluation metrics

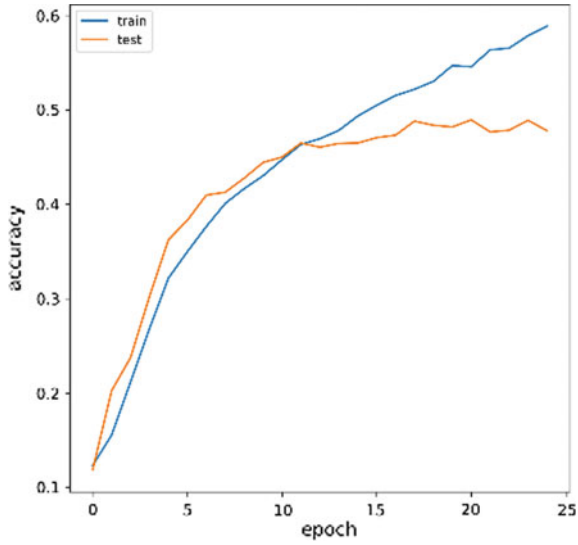


Fig. 6.3 Accuracy for classification

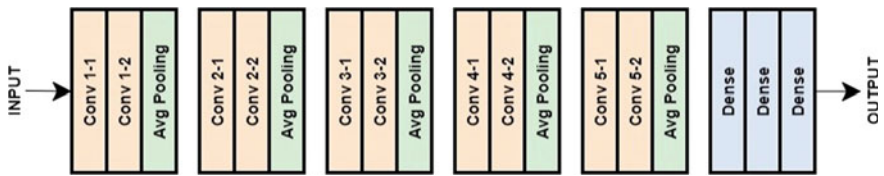


Fig. 6.4 Model architecture

used to predict bone age. This is the approach we have used for our final bone age prediction and in the rest of the paper.

6.4 Experiments

In this section, we test the custom model with various loss functions and compare the custom model with DCNNs such as VGG16, VGG19, and Inception v3 [10].

6.4.1 Loss Functions

As we know, the loss function measures how accurately a model will predict the expected outcome. The loss function outputs the loss, which is a measure of how

Fig. 6.5 Actual age versus predicted age plot with MAE as loss function

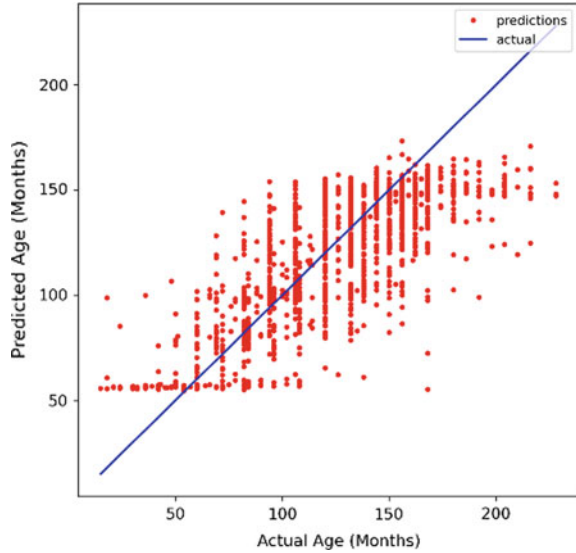


Table 6.2 Metric values for each loss function on the custom model

Loss functions	Custom model		
	MAE	MAPE	MSE
MAE	20.24	17.61	629.87
MSE	17.31	18.62	548.90
Huber	13.89	13.16	349.52

accurately the model predicts final bone age. The selection of an efficient loss function was important for training our custom model.

6.4.1.1 MAE

MAE will never be negative because we are always considering the absolute value of the errors. MAE will be less beneficial if we are concerned about our model's outlier predictions. The large errors caused by outliers are weighted the same as the smaller errors. As a result, we get a few catastrophic predictions (refer Fig. 6.5). The metric values for this loss function on the custom models are presented in Table 6.2.

6.4.1.2 MSE

To compute the MSE, the difference is squared among the model's predictions and true values and finally averaged over the whole data set. Since the errors are always

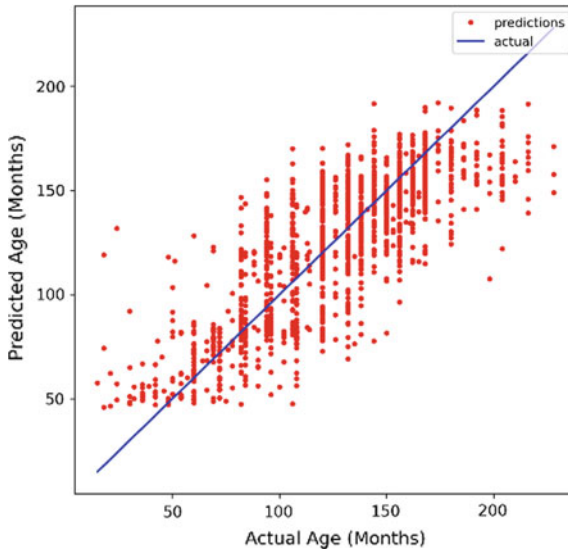


Fig. 6.6 Actual age versus predicted age plot with MSE as loss function

squared, MSE will never be negative (Fig. 6.6). The metric values for this loss function on the custom models are presented in Table 6.2.

6.4.1.3 Huber

The best of both worlds, MSE and MAE are offered by Huber Loss which acts by balancing the MSE and MAE together. The Huber loss function was proved to give the best results for our custom model. When compared Huber loss function with other loss functions such as MAE and MSE, Huber loss function proved to be more effective (refer Fig. 6.7). The Metric values for this loss function on the custom models are presented in Table 6.2.

6.4.2 Comparing with Pre-trained Models

6.4.2.1 VGG16

VGG16 is a CNN model which consists of 16 layers. It was trained on the ImageNet data set comprising 14 million images divided over 1000 classes. The model achieved an accuracy of 92.7% for the top-5 test set. Figure 6.8 is the scatter plot obtained after applying VGG16. Using this model, we achieved an MAE of 16.77 (refer Table 6.2).

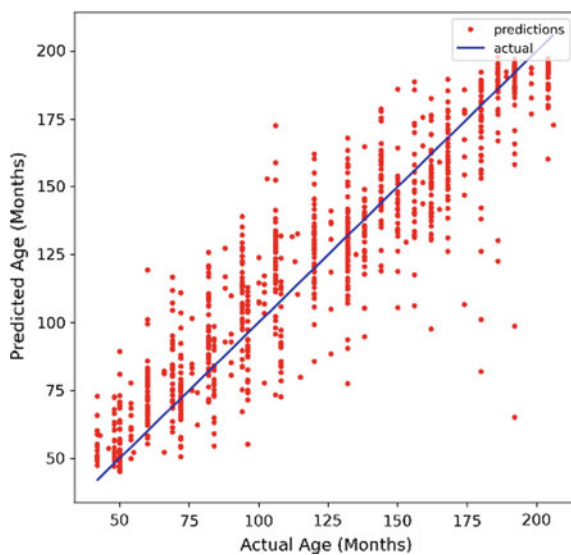


Fig. 6.7 Actual age versus predicted age plot with Huber as loss function

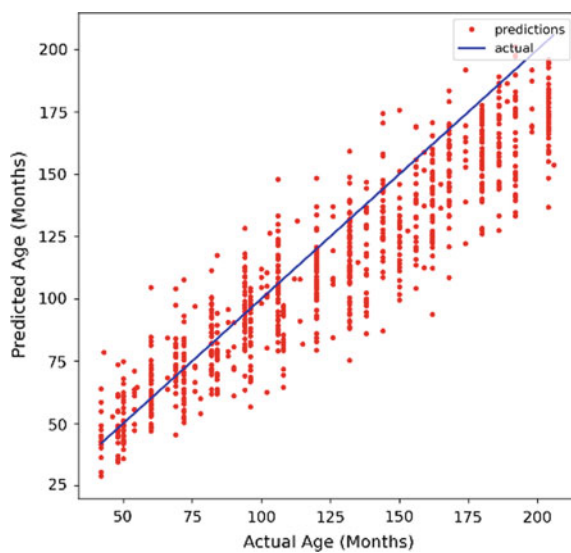


Fig. 6.8 Actual age versus predicted age plot for VGG16

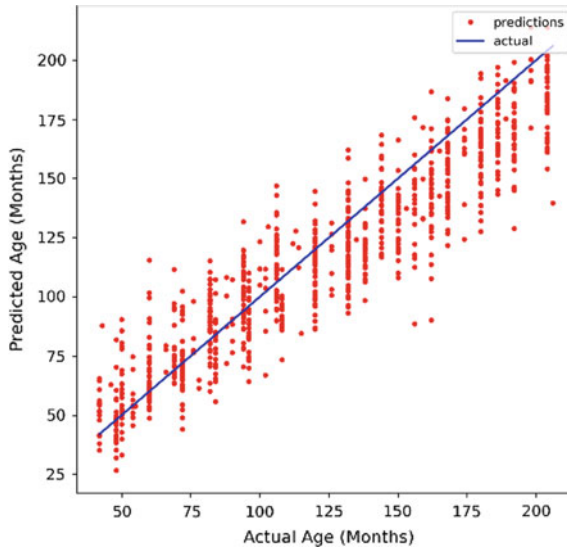


Fig. 6.9 Actual age versus predicted age plot for VGG19

6.4.2.2 VGG19

VGG19 is a CNN model which consists of 19 layers, and it is a variant of VGG. Figure 6.9 is the scatter plot obtained after applying VGG19. We achieved an MAE value of 14.80 with VGG19 (refer Table 6.2).

6.4.2.3 Inception V3

Inception is a CNN model developed by Google. Inception v3 is Google Inception's third version. Figure 6.10 is the scatter plot obtained after applying Inception v3. Using this model, we achieved an MAE of 35.66 (refer Table 6.3).

6.5 Conclusion

Bone age has already been utilized as a diagnostic and therapeutic indication. Moreover, bone age may be used to predict pubertal peak height velocity and menarche timing. In this study, we created a unique DNN model for determining bone age automatically. In summary, we used a set of X-ray pictures to perform classification and regression to estimate bone age. Using 6044 pictures, the MAE value for bone ages using our own model was 13.89, VGG-16 was 16.769, VGG-19 was 14.80, and Inception v3 was 35.66.

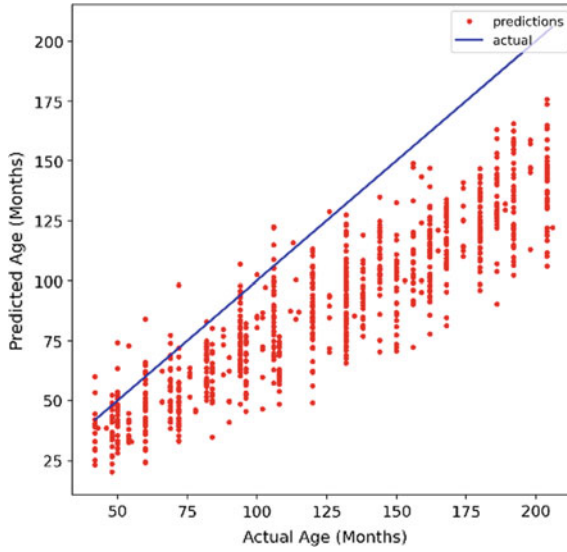


Fig. 6.10 Actual age versus predicted age plot for Inception v3

Table 6.3 Comparison of metrics between our model and various state-of-the-art models

Models	MAE	MSE	MAPE
Inception v3	35.66	1706.74	27.80
VGG16	16.77	460.69	13.65
VGG19	14.80	355.1	13.00
Our custom model	13.89	349.52	13.16

References

1. Mohapatra, S., Nayak, J., Mishra, M., Pati, G.K., Naik, B., Swarnkar, T.: Wavelet transform and deep convolutional neural network-based smart healthcare system for gastrointestinal disease detection. *Interdiscip. Sci.: Comput. Life Sci.* **13**(2), 212–228 (2021)
2. Satoh, M.: Bone age: assessment methods and clinical applications. *Clin. Pediatr. Endocrinol.* **24**(4), 143–152 (2015)
3. Prokop-Piotrkowska, M., Marszałek-Dziuba, K., Moszczyńska, E., Szalecki, M., Jurkiewicz, E.: Traditional and new methods of bone age assessment—an overview. *J. Clin. Res. Pediatr. Endocrinol.* **13**(3), 251 (2021)
4. Mukhopadhyay, A., Mukherjee, I., Biswas, P.: Comparing CNNs for non-conventional traffic participants. In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, pp. 171–175 (2019)
5. Mohapatra, S., Swarnkar, T., Mishra, M., Al-Dabass, D., Mascella, R.: Deep learning in gastroenterology: a brief review. *Handbook of Computational Intelligence in Biomedical Engineering and Healthcare*, pp. 121–149 (2021)
6. Mohapatra, S., Pati, G.K., Swarnkar, T.: Efficiency of transfer learning for abnormality detection using colonoscopy images: a critical analysis. In: *2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAIECC)*, pp. 1–6. IEEE (2022)

7. Sarić, R., Kevrić, J., Čustović, E., Jokić, D., Beganović, N.: Evaluation of skeletal gender and maturity for hand radiographs using deep convolutional neural networks. In: 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 1115–1120. IEEE (2019)
8. Kushol, R., Raihan, M., Salekin, M.S., Rahman, A.B.M.: Contrast enhancement of medical X-ray image using morphological operators with optimal structuring element. arXiv preprint [arXiv:1905.08545](https://arxiv.org/abs/1905.08545) (2019). Westerberg, E.: Ai-based age estimation using X-ray hand images (2020)
9. Westerberg, E.: AI-based Age Estimation using X-ray Hand Images. Doctoral dissertation, Thesis, Faculty of Computing Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden (2020)
10. Tanner, J.M.: Assessment of skeletal maturity and prediction of adult height. TW2 Method, pp. 50–106 (1983)