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

Deep Learning Framework for Automatic Bone Age Assessment

Chaitanya Mehta[1], Bibi Ayeesha[1], Ayesha Sotakanal[1], Nirmala S. R[1], Shrinivas D Desai[1], Venkata Suryanarayana K[2], Ashes Dhanna Ganguly[2], and Veerendra Shetty[2]

***Abstract*— Bone age Assessment or the skeletal age is a general clinical practice to detect endocrine and metabolic disarrangement in child development. The bone age indicates the level of structural and biological growth better than chronological age calculated from the birth date. The X-Ray of the wrist and hand is used in common to estimate the bone age of a person. The degree of agreement among the automated methods used to evaluate the X-rays is more than any other manual method. In this work, we propose a fully automated deep learning approach for bone age assessment. The dataset used is from the 2017 Pediatric Bone Age Challenge released by the Radiological Society of North America. Each X-Ray image in this dataset is an image of a left hand tagged with the age and gender of the patient. Transfer learning is employed by using pre-trained neural network architecture. InceptionV3 architecture is used in the present work, and the difference between the actual and predicted age obtained is 5.921 months.**

Clinical Relevance— This provides an AI-based computer assistance system as a supplement tool to help clinicians make bone age predictions.

I. INTRODUCTION

Bone age assessment (BAA) is a procedure for estimating skeletal maturity which is essential for diagnosing and managing endocrine and growth disorders [1]. BAA is the alternative way of finding the age when proper birth records are not maintained. The abnormalities in skeletal development are indicated by the difference between bone age and chronological age.

In traditional BAA practice, radiologists examine a left-hand X-ray image which includes fingers and wrist. The conventional methods, Greulich and Pyle (GP) approach [2] and Tanner Whitehouse method (TW2), are commonly used in practice [3]. In the first method, bone age is calculated by comparing the patient's hand X-ray image with an atlas. The second method considers important regions of interest such as hand bones to estimate the bone age. There are three specific regions in the image; meta-carpals, carpal bones and proximal phalanges which help evaluate the skeletal development stages in a person. But these manual procedures are time-consuming, need the expertise to assess the image and are a tedious task. They also introduce variability in inter and intra observations. These challenges demand an automated and computer-aided method to predict bone age using hand x-ray images. Here, the shape, edges and texture features are considered as parameters to identify and differentiate bone structures. These approaches mainly focus on extracting the image features to implement the assessment process [4].

In recent years, the Artificial Intelligence (AI) technology is growing tremendously and has provided significant benefits in the medical field [5]. This work focuses on applying deep learning for the automatic assessment of X-ray images to determine bone age. Most deep-learning models developed for nonmedical applications are designed to classify a single planar image, enabling rapid re-use and refinement of that foundational work for bone age classification. Therefore, this research work aims to experiment how a neural network can still predict bone age reasonably well.

A. Convolutional Neural Network (CNN)

The basic component of CNN is a convolution layer which acts as a feature extractor. The ReLu layer applies the function f(x) = max (0, x) to the input values, which will change the negative values to 0. The pooling layer, also called as down-sampling layer, takes a sub-region of the image and performs a pooling operation on the numbers. The dropout layer in the network has nodes that will not over-specialize and represent a more generalized model of the data, resulting in less over-fitting.

B. Training

For training the network, a labeled data set is needed. In this research work, the goal is to predict bone age given the bone scan. The input images are labeled with the corresponding bone-age and gender. The training iterations can be separated into 4 steps: the forward pass, the loss function, the backward pass and weights update. In the forward pass, a training image is sent through the network, results in a prediction which is the estimated bone age. This value then gets passed into the loss function, where the difference between the predicted value and the actual value is calculated. Many different loss functions can be applied. The loss function used in this work is, Mean Absolute Error (MAE) computed as in (1)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^P| \quad (1)$$

where y_i is actual age, y_i^P is the predicted age and n is the total number of predictions. When loss is calculated, the weights can be updated accordingly. This is called the backward pass. The goal here is to minimize the loss; in the end, the network should give predictions as close to the actual values as possible. The final step is to perform a weight update.

Authors[1] are with KLE Technology University, Hubballi, India (corresponding author: +91 9957576638; e-mail: nirmala.s@ kletech.ac.in)
Authors[2], are with Samsung R&D Institute India-Bangalore (e-mail: narayana.kvs@samsung.com)

The recent works in BAA include [6], where several deep neural network architectures are trained end-to-end and use images of the whole hand and specific parts of hand for both training and prediction. The method in [7] uses U-Net to precisely segment hand mask image from a raw X-ray image. To optimize the learning process, they employ six off-the-shell deep Convolutional Neural Networks (CNNs) with pre-trained weights on ImageNet. A deep residual network architecture with 50 layers is used in [8], where the bone age predictions were compared with clinical experts and other CNN models. In this work, we explore a robust model with a novel combination of image masking and properly preprocessed input images to improve the prediction performance.

II. DATA SET

A. Data Description

The network is trained and tested on a data set made available by Kaggle RSNA bone age analysis challenge 2017 [9]. The total data set contains 12611 training DR images and 200 testing DR images of patients aged between 0 to 19 years. The images are in .png format that contains image data and .csv file, which includes patient information such as image id, gender and the actual age. The size of each image is 1514 x 2044. There are 8563 DR images from the ages 1-13 and 2935 DR images from the ages 14-19. The data set is skewed as shown in Fig. 1. Data augmentation is performed to reduce the nonuniformity in data distribution.

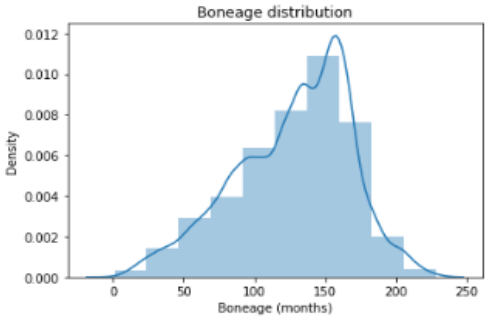
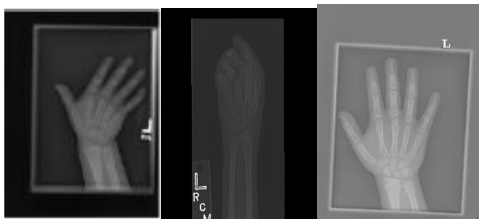


Figure 1. Age distribution of the dataset

Some of the challenges were: the improper orientation of the hand X-rays as shown in Fig. 2(a), limited ROI (metacarpal and phalanx of the thumb) as shown in Fig. 2(b) and around 5% Right Hand DR images were present in the dataset as shown in Fig. 2(c). Besides, there exists variation in contrast and size of the images. The dataset of 12611 bone age images is divided into a training, validation and test set. The dataset is split into 75%/25%, with the 75% parts being the training set. The remaining collection is then again divided into 75%/25%, with the75% being the validation set and the remaining scans being the testing set using k-fold cross validation method. In all three datasets, we have used a nearly equal number of male and female images.

B. Data Preprocessing

While training a neural network using a limited data set, the challenge is that the network can over-fit the data. Image generation is dealing with this by creating more images. The aim is to expose the network with a large collection of input



(a) (b) (c)

Figure 2. Example images of improper orientation (a), limited ROI (b) and right-hand image (c)

images. The following are the various preprocessing operations used in this work:

1. Rotation: random rotation is up to 25 degrees.
2. Horizontal shift: random horizontal shifts are made up to 25% of the original width.
3. Vertical shift: random vertical shifts are made up to 25% of the original height.
4. Shear: random shears are done with a shear factor of up to 0.2.
5. Zoom: up to 20 percent of random zooming is added.
6. Fill mode: Points beyond the input boundary are filled with the nearest ones.
7. Horizontal flip: Randomly flip the input horizontally.

In the next step, we have applied ROI-based masking to extract a region of interest (a hand mask) from the image and remove all foreign objects. Simple methods of removing background do not yield satisfactory results. There is also a strong need for a robust technique of hand segmentation. In this work, SelectROI API is used to create a mask for the region of interest by drawing a bounding box. Fig. 3(a) and 3(b) show the original and masked image, respectively.

Gamma correction is a preprocessing method to improve the contrast of the image. It is a nonlinear operation used to correct the luminance. The gamma correction is a power-law transformation as given in (2),

$$v_{out} = \mathrm{A}\, v_{in}^{\gamma} \qquad (2)$$

where v_{out} is the gamma-corrected output obtained using input value v_{in} elevated to the gamma(γ) power, and multiplied by the constant A.

C. Transfer Learning

It is the method of using an already trained network proven to be effective and fine tune further to enhance the performance. Using the domain knowledge, ROI is defined and initial model is trained with ROI images. Considering this as pretrained network, additional training is done using contrast corrected ROI images and used for bone age prediction.

III. METHODOLOGY

A. InceptionV3

InceptionV3 architecture shown in Fig. 5, is a deep convolutional neural network with a combination of layers (namely,1x1,3x3 and 5x5 convolutional layer) [10]. These layers have filter banks followed by a single output vector, which forms the next layer's input. Both pixel and gender information are used by the InceptionV3 network with an image size of 299x299. An additional dense layer is added to

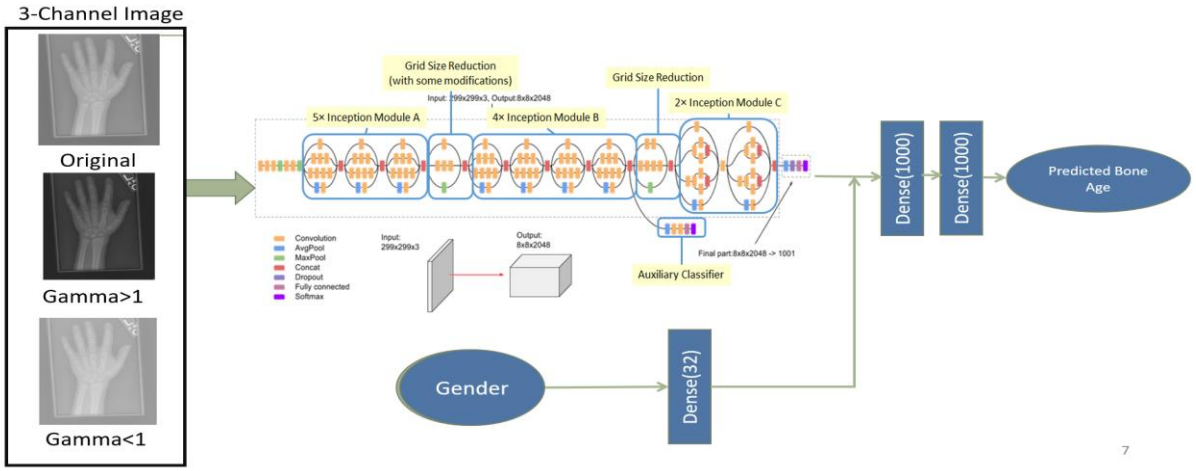


Figure 5: Inception V3 architecture [https://paperswithcode.com/media/methods/inceptionv3onc--view_vjAbOfw.png]

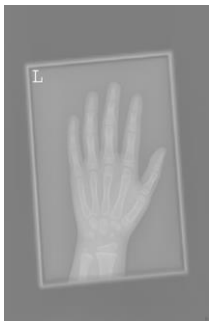


Figure 3(a): Original

Figure 3(b): Masked

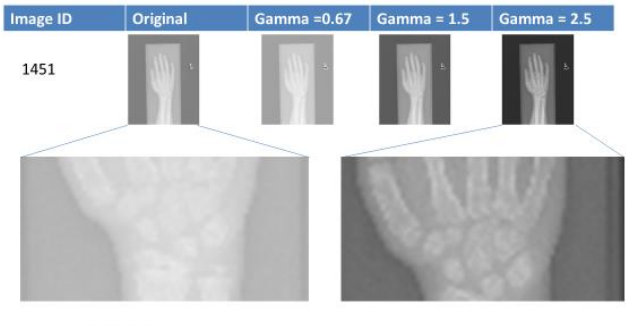


Figure 4: Original and gamma corrected images

get this information and allow the network to know its relationship. Augmentation of data in combination with multiple preprocessing steps has proved to improve the overall performance.

Initially the VGG 16 and Inception V3 networks are trained with different set of preprocessed input images and tested in traditional way. Then transfer learning approach is employed to improve the bone age prediction.

The VGG16 is an excellent neural network architecture, but it may not perform well for complex tasks as it is a simple stack of convolutional and max-pooling layers followed by one another and finally fully connected layers.

On the other hand, Inception nets have inception modules that consist of (1x1) filters, also known as pointwise convolutions, followed by convolutional layers with different filter sizes applied simultaneously. This allows inception nets to learn more complex features. They have more hidden layers when compared to VGG16. Hence, they are used to solve more complex problems.

IV. Results and Discussion

TABLE I. BAA USING DIFFERENT ARCHITECTURES

Architecture	*Images Used*	*MAE in months*
VGG16	Original	8.67
VGG16	Masked	9.67
InceptionV3	Masked	5.94
InceptionV3	Gamma corrected	6.23
InceptionV3	Gamma correction On Masked	6.42
Inception V3	**Transfer Learning with Gamma correction.**	**5.92**

Implementation of different network architectures for BAA is presented in Table I. The original Images were trained using the VGG16 model and an MAE of 8.67 was achieved and masked images gave an MAE of 9.67 using the same model.

In this work, the contrast corrected images are obtained by considering gamma values as 0.67, 1 and 1.5 as shown in Fig.4. Then experiment is conducted by feeding the first channel with gamma=1, second channel with gamma >1 and the third channel with gamma <1 images as shown in Fig. 5.

The InceptionV3 model is trained with original masked images for about 300 epochs and the best MAE of 5.94 was achieved at 240 epochs. After applying gamma correction on

original images, MAE of 6.23 was obtained. Similar experiment is carried out for masked images and MAE of 6.42 is achieved. The next investigation was to apply transfer learning.

The model that gave the MAE of 5.94 on masked images is considered as pre-trained network and training is continued by applying gamma-correction to masked Images. This enabled the model to learn ROI better from contrast-enhanced masked images. Thus, an improved MAE of 5.92 was achieved in this experiment.

Table II demonstrates the performance comparison of present method with several existing BAA approaches. From the table, it is clear that our proposed model outperforms other methods with the lowest MAE of 5.92 months using InceptionV3. The approach by [8], also uses InceptionV3 but the assessment of human reviewers is also considered for evaluating the performance. This makes the method expensive and suffers from inherent variations of human evaluation. But our approach is simple and MAE obtained is also very close to [8].

Further analysis of the prediction errors on the unseen test data for the transfer learning model is shown in Fig. 6. The middle horizontal line represents the median and it is less than 6 months for most of the cases. The lower quartile represents 25% of test data has MAE less than 2.5 months.

TABLE II. COMPARISON TABLE

	Method	*Proposed model*	*MAE in months*
1	Han et al. [11] 2018	Ensemble Learning	8.40
2	Iglovikov et al. [6] without ensemble 2018	U-Net Architecture	8.08
3	Iglovikov et al. [6] with ensemble 2018	VGG Architecture	7.52
4	Wu et al. [12] 2019	Support vector regression	7.38
5	Xiaoying Pan et al. [7] 2020	Inception resnet-V2	7.35
6	Larson et al. [8] 2017	Deep residual network	6
7	Present method	**InceptionV3**	**5.92**

Similarly, the upper quartile represents 75% of test data has MAE less than 9 months for almost all the actual age values. It is observed that age values 11 and 12 years contribute high error due to more outliers, whereas age values 7 to 10 and 13 to 16 years have few outliers and contribute relatively little error. The results can be improved by discarding the outliers and making the data distribution uniform.

V. CONCLUSION

A deep learning approach is employed for automated bone age assessment. Improved prediction was obtained by transfer learning when gamma correction is applied to masked input images and then fed to the Inception v3 network. The proposed method resulted in a better MAE compared to other existing models. But the performance of the BAA model needs improvement to reduce the MAE further. The skeletal maturity values obtained from the present method need to be validated by a medical expert to ensure accuracy. The scope for future work is focused on using different local regions of the hand separately, then weighted combination of them according to their significance to compute the bone age.

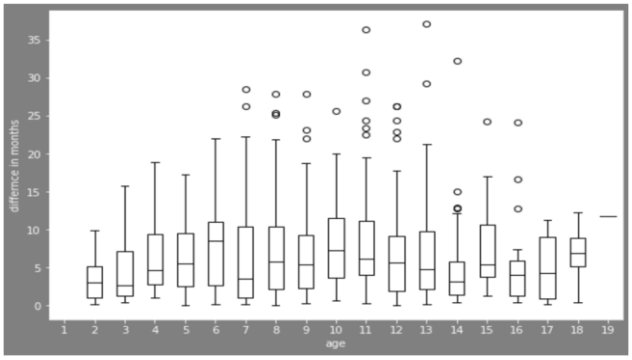


Figure 6. Boxplot of true age and MAE

REFERENCES

[1] R. Vanderwilde, L. T. Staheli, D. E. Chew, and V. Malagon, "Measurements on radiographs of the foot in normal infants and children," The Journal of Bone and Joint Surgery, vol. 70, no. 3, pp. 407–415, 1988.

[2] E. Reynolds, "Radiographic atlas of skeletal development of the hand and wrist," Am J Phys Anthropol 8(4):518–520,1950

[3] J. M. Tanner, R. H. Whitehouse, N. Cameron, W. A. Marshall, M. J. Healy, and H. Goldstein, "Assessment of skeletal maturity and prediction of adult height (TW2 method)", London: Academic Press, 1975.

[4] Thodberg H. H, Kreiborg S, Juul A, Pedersen KD (2009), "The BoneXpert method for automated determination of skeletal maturity," IEEE Trans Med Imaging 28:52–66.

[5] S. H. Tajmir, H. Lee, R. Shailam et al., "Artificial intelligence assisted interpretation of bone age radiographs improves accuracy and decreases variability," Skeletal Radiology, vol. 48, no. 2, pp. 275–283, 2019.

[6] V. I. Iglovikov, A. Rakhlin, A. A. Kalinin, A. A. Shvets (2018), "Paediatric Bone Age Assessment Using Deep Convolutional Neural Networks", In: Stoyanov D. et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science, vol. 11045. Springer, Cham.

[7] Xiaoying Pan, Yizhe Zhao, Hao Chen, De Wei, Chen Zhao, Zhi Wei, "Fully Automated Bone Age Assessment on Large-Scale Hand X-Ray Dataset", International Journal of Biomedical Imaging, vol. 2020, Article ID 8460493, 12 pages, 2020.

[8] Larson, David B., et al. "Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs." Radiology 287.1 (2017).

[9] Halabi, Safwan S., et al. "The RSNA pediatric bone age machine learning challenge." *Radiology* 290.2 (2019): 498-503.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826.

[11] J. Han, Y. Jia, C. Zhao, and F. Gou, "Automatic bone age assessment combined with transfer learning and support vector regression," 9th International Conference on Information Technology in Medicine and Education (ITME), pp. 61–66, Hangzhou, China, 2018.

[12] E. Wu, B. Kong, X. Wang et al., "Residual attention-based network for hand bone age assessment", 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1158–1161, Venice, Italy, 2019.