Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2s (2026) | Pages: 205-210

Bone Mature: A Deep Learning Framework for Bone Age Detection

Sanika Bhatkar[1], Kush Makwana[2], Manjiri Kadam[3], Jayashree Khanapuri[4]

[1,2,3,4]Department of Electronics and Telecommunications, KJSIT, Mumbai

[1]*sanika.bhatkar@somaiya.edu*, [2]*kush.makwana@somaiya.edu*, [3]*manjiri.k@somaiya.edu*

Peer Review Information

Type: *Article*
Received: *27 March 2026*
Revised: *12 April 2026*
Accepted: *26 May 2026*
Published: *16 June 2026*

Abstract

Bone age estimation is one of the most important processes in paediatric radiology which is used to determine the level of skeletal development and identify various developmental problems. Hand radiograph interpretation is manually used in the estimation of bone age, but this method is rather subjective and inconsistent. In this paper, the author presents the Bone Mature, a two-step deep learning system to estimate bone age without the need of manually annotated region of interest. During the first step, an annotation-free cascaded critical bone region localization model is trained using InceptionV3 with an extra convolutional block attention module (CBAM) and gradient-weighted class activation mapping (Grad-CAM). In particular, carpal and metacarpal-phalanx bones, which are anatomically important, are automatically detected in hand radiographs by the proposed cascaded critical bone region localization model. The second step involves the localized areas being inputted into ResNet50 and Xception models with mid-level feature fusion and gender-assisted estimation.

Keywords: Bone Age Assessment; Deep Learning; Annotation-Free ROI Extraction; Grad-CAM; CBAM; InceptionV3; ResNet50.

How to Cite This Article

Bhatkar, S., Makwana, K., Kadam, M., & Khanapuri, J. (2026). Bone Mature: A deep learning framework for bone age detection. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2s), 205-210.

Introduction

A bone age assessment (BAA) is an important diagnostic tool in children health care. It offers helpful knowledge of skeletal maturity, development issues, and developmental disorders. BAA has a great impact on clinical decision-making in the field ofitoring, and treatment planning. The classic systems, such as the Greulich-Pyle (G&P) and Tanner-Whitehouse (TW) systems, are based on hand X-rays being examined by human experts. This is because process is subjective, slow and inconsistent as there are different opinions among observers.

Recent advances in deep learning have resulted in automated age estimation of bones. These models are learnt to acquire unique ossification patterns by simply looking at hand X-rays. Although they are more accurate and efficient, a considerable number of existing models lack focus on significant bone regions or require manual delineation of the regions of interest (ROIs). To address these problems, novel models identifying key bone regions automatically have become popular.

In the current work, we present Bone Mature, a two-step deep learning system that classifies the bone age based on automated methods. Our methodology involves a cascading method to obtain important bone parts. It uses a backbone of InceptionV3 with Convolutional Block Attention Module (CBAM) and localization using Grad-Cam.

to find the key locations in the anatomy, namely, the carpal and metacarpal-phalanx bones, without requiring manual ROI annotations. The second stage involves processing the extracted regions using ResNet50 and Xception networks, and combining mid-level features and gender-prediction to estimate bone age.

Experiments on RSNA bone age dataset demonstrate that the system can capture significant skeletal features and give accurate bone age estimates. Bone Mature provides a viable and scalable solution to automated bone age determination in clinical settings by eliminating manual processes and emphasizing on explainability

Review Of Literature

The GreulichPyle (G&P) and TannerWhitehouse (TW) standards have long been regarded as the masterpiece-based methods of bone age assessment, the visual analysis of hand radiographs and the application of these as references in comparison with atlases. Despites are subjective, time-consuming and lack inter-observer variability, which inspires the creation of automated methods. Early automated techniques, like BoneXpert method, utilized Active Appearance Models to examine single hand bones and transfer intrinsic age of the bone to GP/TW standards, with a clinically acceptable accuracy without manual intervention [1]. Although efficient, these model-based methods need well-considered hand-crafted representations and lack flexibility.

As deep learning has emerged, convolutional neural networks (CNNs) have shown to be more effective at learning discriminative features directly off of medical images. Kalejahi et al. demonstrated that CNN-based regression and classification models are much more effective in estimating bone age than the conventional manual techniques, not to mention that the former minimizes observer variability [2]. On the same note, Ren et al. suggested a regression CNN with coarse-to-fine attention, which delivers similar performance to professional radiologists and is entirely automated [3].

Multi-stage and region-centric deep learning pipelines have been studied on a number of studies. Chu et al. proposed a two-step deep neural network in which hand segments were initially divided with the help of U-Net and VGG16, and bone age prediction was formulated as an ordinal regression problem, which performed better on the RSNA dataset [4]. Nevertheless, these techniques are very sensitive to accurate segmentation and manual region marking, which makes them more expensive to annotate and reduces their scalability.

In a bid to address the problem of annotation dependency, weakly supervised and annotation-free methods have become more and more popular. Escobar et al. published the Radiological Hand Pose Estimation (RHPE) dataset and proposed BoNet, which uses local region features with global hand context to enhance robustness in bone age prediction [5]. Liu et al. also introduced PEAR-Net, an unsupervised attention based model that learns to find discriminative bone areas without the use of ROI labels, and had low mean absolute error on the RSNA dataset [6].

Recent research has focused on the attention processes to increase interpretability and concentration on clinically significant areas. Attention modules incorporated into CNNs like Grad-CAM and CBAM have proven to be more localizing to important skeletal features, which allows explainable bone age prediction [7]. Cascaded methods of critical bone region extraction without annotation have also been suggested, in which several passes are performed to localize anatomically important regions without the need to label them manually and with good accuracy and applicability to clinical use [8].

The use of multimodal models based on the clinic metadata (e.g. gender and chronological age) has further enhanced the accuracy of prediction. These techniques help in overcoming biological variations in skeletal development and minimise prediction bias in case of data imbalance by combining image-based characteristics with auxiliary data [9].

In spite of these developments, there are still difficulties in finding a balance between accuracy, interpretability and practical application. Most of the existing methods are based on manual annotations, are not explainable, or do not have explicit attention to multiple important bone regions. Inspired by these constraints, the proposed Bone Mature framework uses an annotation-free, cascaded critical bone region detection approach with attention mechanisms and gender-assisted estimation, an effective and explainable way to evaluate the age of the bones Automatically.

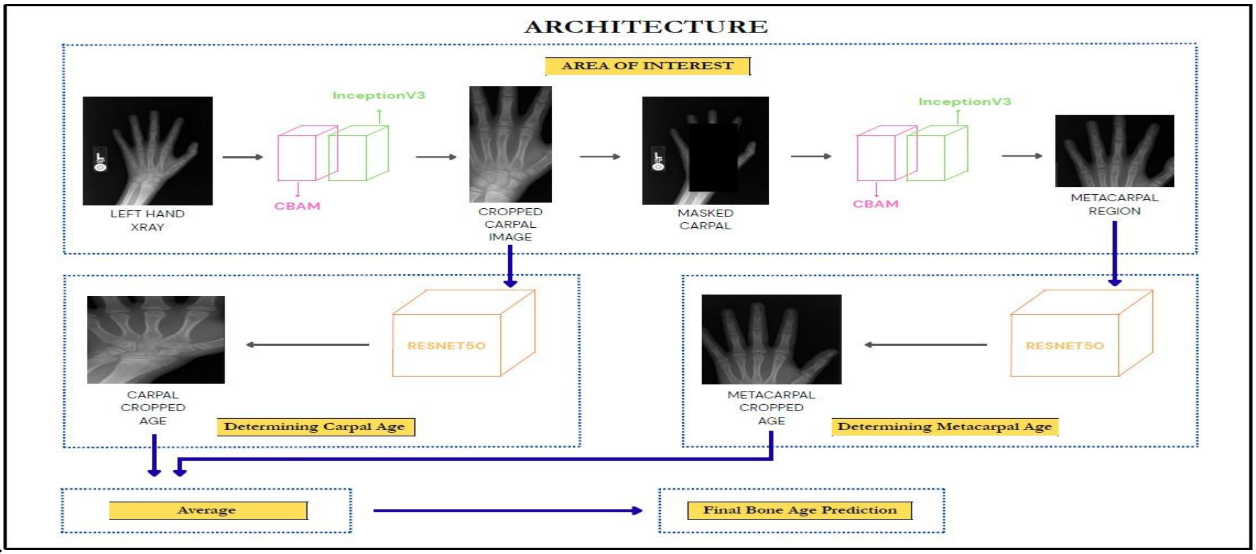

Fig. 1. *Deep Learning-Based Architecture for Bone Age Prediction from Hand X-Ray Images*

Methodology

This part is associated with the method that we are using to estimate bone ages using hand radiographs. We use a two stage deep learning architecture. This model comprises two parts:

- Locating the bone parts.
- Age estimations of the bone through deep convolutional neural networks.

The key objective is to target the sections of the bones which physicians believe to be vital without having to mark those parts by hand.

Overview Of the System

Our system will help to estimate the bone age using a left-hand X-ray image. It does this by examining the sections of the bones. Of making a prediction based on the entire hand image the system initially identifies the important bone areas. Then it determines the age using these regions. By doing this it can disregard the portions of the image that are not important and make a more accurate prediction.

The method has a main step:

- Preparing the hand radiographs.
- Intelligently locating the bone critical areas.
- Using deep neural networks to get features
- A regression-based model to predict the bone age.

Description Of the Dataset

- We are testing our method with the Radiological society of North America Pediatric Bone Age dataset.
- This data can be accessed publicly on Kaggle.
- It contains left-hand X-ray pictures with the age of the bones written. The pictures show children of different ages and are both boys and girls.
- The RSNA data is solely being used to train, validate and test our system.
- We are not employing any datasets so that anyone can re-do our results to obtain the same responses.C. Finding The Critical Bone Regions

Finding The Critical Bone Regions

We must examine some portions of the bones in order to determine the age of the bones. These are the carpal bones and the bones of the metacarpus-phalanx.

They change in a way as we grow.

To find these regions automatically we are using a method that does not need any annotations.

The type of convolutional neural network we are employing is able to outline the most significant parts of the image. Here are the places that assist the network in making predictions.

According to these areas the system discovers two bone regions:

- The carpal bone area.
- The phalanx and metacarpus area.

This method of determining the areas does not require one to mark the bones manually and is consistent with the procedure used by doctors when they determine the age of bones.

The Bone Age Estimation Network.
Once we locate the bone areas which we use in estimating the bone age.

There are approaches where different regions are operated with the use of networks, but we are employing the same ResNet-based architecture in both regions.The features are obtained with the help of a ResNet model per region.

The features from both regions are then combined to make a feature vector.In this manner, the model is able to capture information on significant areas of the bones.Age estimation of the bone is then done by using the combined features.This is achieved through a regression task, with the network giving an age of a bone.E. Training The Model And Loss Function

Training The Model and Loss Function

- The model is being trained on the learning of the bone age labels of the RSNA dataset.
- We are employing the data preprocessing methods to ensure that the images are equal in size.
- As we are considering bone age estimation as a regression problem we are minimizing the squared error loss function to minimize the difference between the estimated and the real values of the bone age.
- We are judging the model's performance based on absolute error which is a typical measure of bone age evaluation studies.

Results

The performance of the proposed bone age assessment framework was looked at using the RSNA Pediatric Bone Age dataset. This information was to be employed as training and validation. This model was trained over 30 epochs. The loss function was mean squared error and evaluation metric was mean absolute error. These are very popular indicators of bone age estimation since they indicate the extent to which the predictions are accurate in years with respect to months.

The performance of training and validation was monitored in the training process to check the learning performance of the model. The extent to which it would be applicable to new data. The training loss decreased gradually in epoch. This indicates that optimization of network parameters was carried out effectively. The validation loss also decreased. With some fluctuations. This is typical in imaging since the skeletal development of peoples may be different. The final training loss of the model was 249.88 and validation loss was 157.51. These findings indicate that the network did not overfit on hand radiograph features and learned them.

The proposed approach works well in terms of the mean absolute error curves indicating that the predictions were accurate. The mean error decreased to 12.43 months during the training. The validation MAE was 9.76 months. The smallest validation mean absolute error was 9.48 months. This shows that the model can be used for samples. The mean error on the validation was less than the training mean error. This indicates that the examination of bone areas and a ResNet-based architecture can assist the model to be effective. The average values of the error abs were decreasing with the epochs. This indicates that the model improved in predicting bone age as it was trained.

The findings also indicate that the examination of bone areas is superior to that of the entire hand. The carpal and metacarpus-phalanx were examined in the model. This assisted the model in determining patterns that are significant in estimating bone age. The results are similar to methods that do not use annotations. This demonstrates that the suggested framework is a compromise of accuracy, ease and reproducibility. The bone age assessment framework fared well on the bone age assessment task. The bone age evaluation platform was based on the RSNA Pediatric Bone Age dataset. The bone age evaluation model was trained for 30 epochs. The mean squared error and mean absolute error were used as the bone age assessment framework. The bone age assessment system was reliable and fixed. The bone-age assessment model acquired is an attribute of hand radiographs. The bone age assessment framework was good at predicting bone age. Bone age assessment framework is an approach, to the estimation of bone age. The analysis is made on one publicly available dataset, and performance can be different when used on data collected with varying imaging conditions or across different populations. Moreover, small variability in the validation curves implies that additional gains may be obtained by refusing to use more sophisticated data augmentation methods or more sophisticated feature fusion methods. However, the given findings prove that the offered approach is a valid and effective way to evaluate bone age with the help of publicly available data and is appropriate to use in scholarly research and baseline clinical decision-support systems.

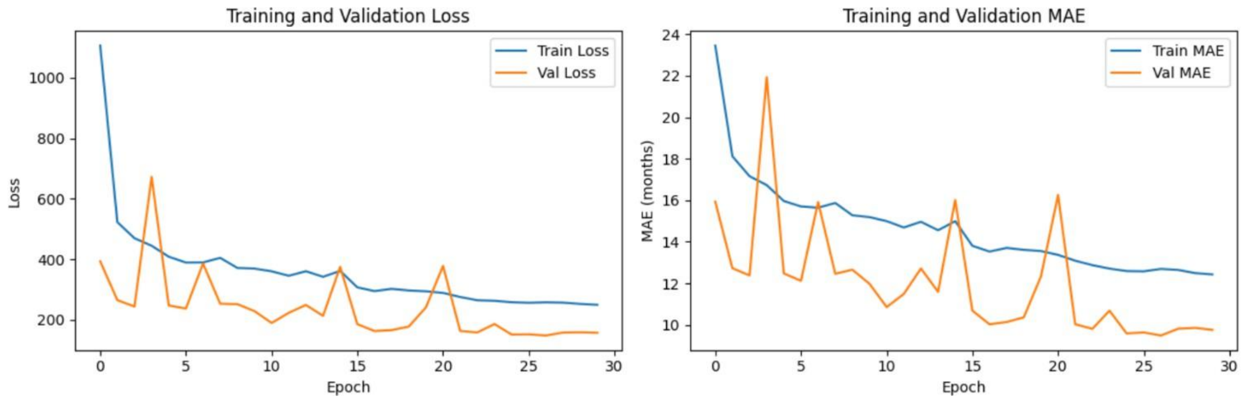

Fig. 2. *Epoch-Wise Variation of Loss and Mean Absolute Error for Bone Age Prediction*

Table 1. *Predicted Bone Age Results Obtained from Carpal and Metacarpal Region Analysis*

Sr.No	Metacarpal Predicted Age	Carpal Predicted Age	Average	Predicted Bone Age
1	53.84 months	173.88 months	113.86 months	9.49 years
2	90.61 months	132.92 months	111.77 months	9.31 years
3	38.28 months	32.24 months	35.26 months	2.94 years
4	141.27 months	140.74 months	141.00 months	11.75 years
5	153.63 months	126.27 months	139.95 months	11.66 years

Conclusion

The paper concerns itself with one such application of deep learning to determine the age of bones in kids based on their X-rays in the hands. The technique which they invented consists of two processes. It automatically locates the regions of the bones in the X-ray. It then determines the bone age based on those parts. They examined various sections of the hand using the basic network thus simplified and more stable system.

They applied this approach to a collection of X-rays of the RSNA Pediatric Bone Age data. The outcomes were fairly satisfactory. The method was able to learn a lot from the X-rays and make estimates of the bone age. Their lowest error was 9.48 months and that is quite good. This demonstrates that emphasis can be placed on the

this can be effectively done by just looking at the parts of the bones even without any manual annotations and additional clinical information. The outcomes are the same as those recorded previously by other individuals and this is why this technique is practical in life. Another factor that is essential is the fact that this approach is clear and simple to comprehend. They have utilized a dataset, implying that anybody can use it and compare their findings. The manner in which they were examining sections of the bones is also the same manner in which doctors examine them in real life. This renders the technique not only good in research, but also in real-life clinical applications. However there are still some limitations, to this method. They tested it on just a single dataset, hence they are unaware of its effectiveness with X-rays of individuals or captured differently. In the future they could try using datasets or coming up with new ways to combine the different parts of the bones. They could also try to make the method more automated. This study as a whole demonstrates that learning is a promising method to estimate bone age by looking at specific sections of the bones and it could result in more effective and efficient diagnostic systems.

References

1. Li, Z., Chen, W., Ju, Y., Chen, Y., Hou, Z., Li, X., & Jiang, Y. “Bone age assessment based on deep neural networks with annotation-free cascaded critical bone region extraction,” Frontiers in Artificial Intelligence, vol. 6, Art. no. 1142895, Mar. 2023, doi:10.3389/frai.2023.1142895.
2. Halabi, S.S., et al. “The RSNA Pediatric Bone Age Machine Learning Challenge,” Radiology, vol. 290, no. 2, pp. 498–503, 2019, doi:10.1148/radiol.2018180736.

3. He, K., Zhang, X., Ren, S., & Sun, J. “Deep residual learning for image recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778, 2016, doi:10.1109/CVPR.2016.90.
4. Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., & Leonardi, R. “Deep learning for automated skeletal bone age assessment in X-ray images,” Medical Image Analysis, vol. 36, pp. 41–51, 2017, doi:10.1016/j.media.2016.10.010.
5. Larson, D.B., Chen, M.C., Lungren, M.P., Halabi, S.S., Stence, N.V., & Langlotz, C.P. “Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs,” Radiology, vol. 287, no. 1, pp. 313–322, 2018, doi:10.1148/radiol.2017170236.
6. Pan, X., Zhao, Y., Chen, H., Wei, D., Zhao, C., & Wei, Z. “Fully automated bone age assessment on large-scale hand X-ray dataset,” International Journal of Biomedical Imaging, vol. 2020, Art. no. 8460493, pp. 1–12, 2020, doi:10.1155/2020/8460493.
7. Iglovikov, V., Rakhlin, A., Kalinin, A., & Shvets, A. “Pediatric bone age assessment using deep convolutional neural networks,” Proc. Int. Workshop Deep Learning in Medical Image Analysis, Granada, Spain, pp. 300–308, 2017.
8. Ren, X., Li, T., Yang, X., Wang, S., Ahmad, S., & Xiang, L. “Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph,” IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 5, pp. 2030–2038, Sep. 2019, doi:10.1109/JBHI.2018.2876916.
9. Woo, S., Park, J., Lee, J.-Y., & Kweon, I.-S. “CBAM: Convolutional block attention module,” Proc. European Conf. Computer Vision (ECCV), Munich, Germany, pp. 3–19, 2018, doi:10.1007/978-3-030-01234-2_1.
10. Lee, B.D., & Lee, M.S. “Automated bone age assessment using artificial intelligence: The future of bone age assessment,” Korean Journal of Radiology, vol. 22, no. 5, pp. 792–800, 2021, doi:10.3348/kjr.2020.094.