



AE-BoNet: A Deep Learning Method for Pediatric Bone Age Estimation using an Unsupervised Pre-Trained Model

Mojtaba Sirati-Amsheh (MSc)¹, Elham Shabaninia (PhD)^{2*},
Ali Chaparian (PhD)¹

¹Department of Medical Physics, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

²Department of Applied Mathematics, Faculty of Sciences and Modern Technologies, Graduate University of Advanced Technology, Kerman, Iran

ABSTRACT

Background: Accurate bone age assessment is essential for determining the actual degree of development and indicating a disorder in growth. While clinical bone age assessment techniques are time-consuming and prone to inter/intra-observer variability, deep learning-based methods are used for automated bone age estimation.

Objective: The current study aimed to develop an unsupervised pre-training approach for automatic bone age estimation, addressing the challenge of limited labeled data and unique features of radiographic images of hand bones. Bone age estimation is complex and usually requires more labeling data. On the other hand, there is no model trained with hand radiographic images, reused for bone age estimation.

Material and Methods: In this fundamental-applied research, the collection of Radiological Society of North America (RSNA) X-ray image collection is used to evaluate the efficiency of the proposed bone age estimation method. An autoencoder is trained to reconstruct the original hand radiography images. Then, a model based on the trained encoder produces the final estimation of bone age.

Results: Experimental results on the Radiological Society of North America (RSNA) X-ray image collection achieve a Mean Absolute Error (MAE) of 9.3 months, which is comparable to state-of-the-art methods.

Conclusion: This study presents an approach to estimating bone age on hand radiographs utilizing unsupervised pre-training with an autoencoder and also highlights the significance of autoencoders and unsupervised learning as efficient substitutes for conventional techniques.

Citation: Sirati-Amsheh M, Shabaninia E, Chaparian A. AE-BoNet: A Deep Learning Method for Pediatric Bone Age Estimation using an Unsupervised Pre-Trained Model. *J Biomed Phys Eng*. 2025;15(3):271-280. doi: 10.31661/jbpe.v0i0.2304-1609.

Keyword

Bone Age Measurement; Radiographs; Pediatrics; Deep learning; Convolutional Neural Networks; Hand Bones

Introduction

In pediatric radiography, bone age assessment is used for diagnostic and therapeutic objectives to investigate endocrine problems, children's growth, and genetic disorders [1]. Hand X-ray images are commonly applied to assess bone maturation due to their availability and the low radiation dose required for capturing images. Currently, two clinical techniques that radiologists use are the Greulich and Pyle (GP) [2] and the Tanner-Whitehouse (TW) techniques (including TW2 [3]

*Corresponding author:
Elham Shabaninia
Department of Applied Mathematics, Faculty of Sciences and Modern Technologies, Graduate University of Advanced Technology, Kerman, Iran
E-mail: e.shabaninia@kgut.ac.ir

Received: 23 April 2023
Accepted: 17 August 2023

and TW3 [4]).

The GP approach is mostly used because of its simplicity and compares the entire X-ray image with a standard reference atlas [1]. On the other hand, the TW method is more accurate with consideration of specific regions of interest from carpal and phalangeal joints and scores each region based on bone morphological features [5]. These two methods usually take considerable amounts of time and are subject to observer variability; two observers may report different scores, or even an observer may have different scores at different times. Therefore, automated and efficient techniques are required for bone age assessment.

Before the advent of deep learning, bone age estimation was performed with traditional machine learning methods; however, deep learning has recently become very popular because of its superior accuracy when trained with large amounts of data, especially for clinical radiological tasks, such as meniscus tears [6], musculoskeletal radiology [7], shoulder pain on radiographs [8], predicting pain progression in knee osteoarthritis [9], and automated detection/classification of shoulder arthroplasty [10]. For bone age estimation, deep learning methods have also achieved attention, due to the ability to automatically realize discriminative features of the images [11]. However, estimating bone age is a challenging supervised task that needs a large amount of labeled data for training (which is usually hard to acquire). In addition, X-ray images of hand bones differ fundamentally from non-radiographic images, and existing models trained on datasets, such as ImageNet, are not suitable for transfer learning. Therefore, in this research, an unsupervised pre-training approach for automated bone age estimation is proposed, which trains an autoencoder to reconstruct original hand radiography images to provide an efficient encoder capable of extracting essential features of the images. Then, the final network uses the trained encoder for bone age assessment using pediatric hand radiographs.

Material and Methods

In this fundamental-applied research for bone age estimation that is a regression-based task, we follow these general steps:

Data preparation: We gathered our dataset consisting of input images and corresponding target values. Also, pre-processing, such as re-sizing and normalization was done on the images appropriately.

Unsupervised pre-training with autoencoder: As a preliminary step, an unsupervised pre-trained model is used based on an autoencoder. The autoencoder is trained on the input images alone, without considering the target values. This pre-training step helps learn meaningful representations from the input data and can serve as a useful initialization for the subsequent regression task.

Network architecture: A CNN-suitable architecture was designed for regression, incorporating the pre-trained autoencoder's learned representations, including convolutional layers, pooling layers, and activation functions. In other words, this architecture has an encoder part of the pre-trained autoencoder to use learned features. Then these features pass through additional layers, such as fully connected layers or global integration layers to capture high-level features.

Loss function: An appropriate loss function was selected for the regression tasks. Common selections, including Mean Squared Error (MSE) loss or Mean Absolute Error (MAE) loss. The loss function quantifies the difference between the predicted regression values and the ground truth values.

Training: The CNN was trained using our prepared dataset. During training, the network learns to minimize the loss function by adjusting the weights of the network through back-propagation. We fine-tuned the pre-trained layers of the autoencoder and updated the weights of the newly added CNN layers. Also, Adam's optimization algorithm was used to update the weights of the network.

Evaluation: The trained CNN model was

evaluated on a separate test dataset. The regression metrics, such as MSE and MAE, were calculated to assess the performance of the model in predicting the target values.

The code is developed in Python using Keras 2.8.0 and TensorFlow as the backend. All the experiments were carried out on a computer equipped with an NVIDIA RTX 2060 graphics processing unit with 6 GB of memory.

Prediction: The model is trained and evaluated to predict new, unseen data. We pass the input images through the trained CNN, and the network will output the predicted regression values.

The use of the unsupervised pre-trained autoencoder enhances the CNN's ability to extract meaningful features from the input images, potentially improving the performance of the regression task. It can help capture relevant patterns and reduce the need for large amounts of labeled training data.

Dataset

This fundamental-applied study uses the collection of Radiological Society of North America (RSNA) X-ray images [12] to evaluate the proposed AE-BoNet method. This collection is labeled by skilled pediatric radiologists using the GP approach. Figure 1 shows a sample image of the dataset with regions

of interest corresponding to the TW method and different bones. RSNA dataset comprises 12,611 images with the age distribution from 1 to 228 months (with a concentration on 5- to 15-year-old children), including 5778 female and 6833 male images. The data is imbalanced for the number of hand images in this dataset for various age groups. Some bone ages include just one sample, while others have as many as 718. The dataset is randomly split into three sets: 200, 2323, and 10088 images for the test set, the validation set, and training, respectively.

Autoencoder architecture

Two neural network models underpin the system (Figure 2): an autoencoder and a regression network for age estimation. The encoder and decoder models are combined to create an autoencoder, which is a particular kind of neural network. The encoder compresses the input, and then the decoder uses this compressed information to reconstruct the original input. Eventually, the decoder model is discarded, and the encoder model is used in a new model for the regression task of bone age estimation.

Autoencoders are models based on neural networks used for unsupervised learning. An autoencoder consists of two basic

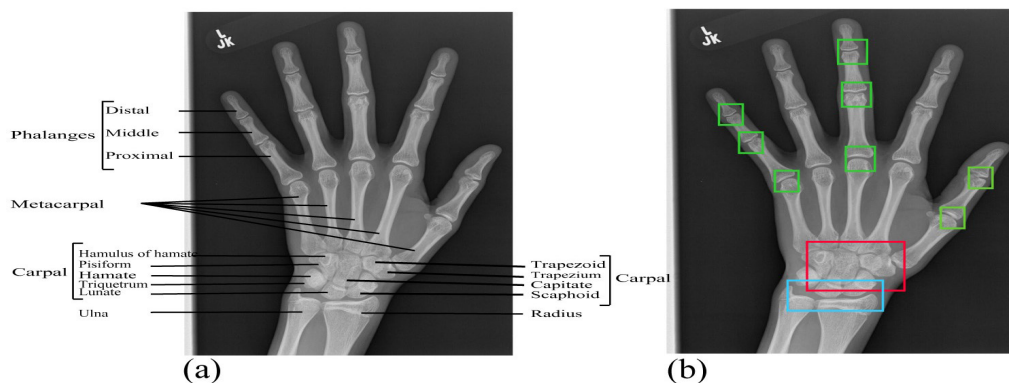


Figure 1: (a) Different bones in a sample image of hand radiograph (image from the Radiological Society of North America (RSNA) dataset [12]), (b) Regions of interest in the Tanner-Whitehouse (TW) approach [1]. The green, red, and blue boxes refer to the phalangeal joints, carpal, and radius/ulna regions, respectively

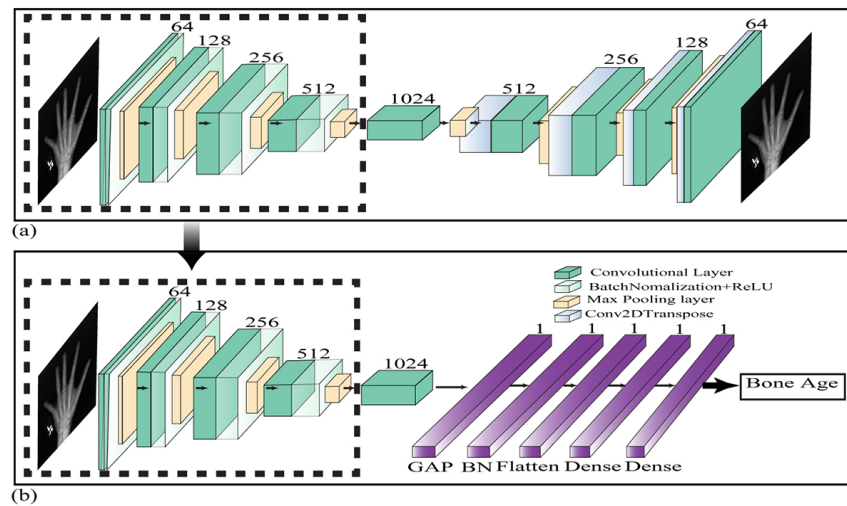


Figure 2: An outline of the proposed approach. (a) an autoencoder is trained to extract useful spatial features for hand bone images and to reconstruct original images (b) a regression network to estimate the final bone age using the encoder component of the autoencoder in (a)

components: an encoder converting input into code and a decoder reconstructing the input from the code. Reconstruction would be carried out by an optimum autoencoder as nearly as possible. The model frequently learns valuable features from the data as it needs to prioritize, which input aspects [13]. The autoencoder model incorporates down- and up-sampling. Shortly, an autoencoder consists of the following components:

Encoder: to extract the proper features from the input and compress them to a bottleneck layer-defined internal representation.

Decoder: to receive the bottleneck as input and reconstruct the original image using the encoder.

Convolutional neural networks (convents) are appropriate for both encoding, and decoding as inputs are images. The original data χ is mapped by the encoder function ϕ to a bottleneck's latent space f . The latent space f at the bottleneck is mapped to the output via the decoder function, represented by φ . In this circumstance, the output function is the same as the input function. Thus, the original image is reconstructed after generalized non-linear compression, and minimizing reconstruction error is the learning aim of an autoencoder

[14].

$$\begin{aligned}\phi: \chi &\rightarrow f \\ \varphi: f &\rightarrow \chi\end{aligned}\quad (\text{Equation 1})$$

$$\phi.\varphi = \arg \min_{\phi.\varphi} \|\chi - (\varphi \circ \phi)\chi\|^2$$

The loss function utilized to train the neural network through the standard backpropagation procedure is defined as follows:

$$\phi.\varphi = \arg \min_{\phi.\varphi} \|\chi - (\varphi \circ \phi)\chi\|^2 \quad (\text{Equation 2})$$

Regression model

When the autoencoder is trained, the decoder portion is discarded, and the model up to the bottleneck is utilized. Then, the Global Average Pooling (GAP) layer is used to lower the number of parameters in the model and help to reduce overfitting. The subsequent flattened layer flattens the input (without altering the information in the preceding GAP layer). Also, batch normalization is used to establish high learning rates [15], leading to the learning process more stable and significantly reducing the required number of training epochs. Following the flattened layer, two dense layers were added for further feature extraction and mapping. Fully connected layers connect each neuron to every neuron in the previous

layer. These dense layers will learn complex relationships in the data and cause the model to estimate the bone age based on the learned features.

The first dense layer used ten neurons with the Rectified Linear Unit (ReLU) activation function. Also, the L_2 regularization technique is applied to avoid overfitting and prevent learning complex models [16]. This is achieved by adding a penalty to the loss function during training based on the magnitude of the weights in the neural network. The value of 0.0001 was set as the strength of the penalty. This approach enhances the model's ability to generalize to unseen data. A single output neuron with a linear activation function was employed in the final dense layer. This type of setup is used for regression tasks, where the network's output is directly predicting a continuous value.

Results

Evaluation metric

Mean Square Error (MSE) and Mean Absolute Error (MAE) are taken into consideration as the metrics for assessing the final result:

$$MAE = \frac{1}{N} \sum_{i=1}^N |f_i - y_i| \quad (\text{Equation 3})$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (\text{Equation 4})$$

In this equation, N is the total number of test points, f_i is the actual bone age, and y_i is the anticipated age.

Autoencoder results

The original X-ray image size for encoder input is 256×256. The encoder consists of four convolutional blocks (convolution, batch normalization, and pooling layer). The filters gradually increase (64,128, 256,512, and 1024). Furthermore, the size of feature maps is progressively reduced as they pass through the convolutional block. At the decoder, the

bottleneck is fed to complete the model symmetrically (the number of filters is 1024, 512, 256, 128, and 64, respectively). The Adam optimizer [17] and mean squared error are used as the loss function. Table 1 indicates the parameters used for training the autoencoder.

The autoencoder trained for 900 epochs in 5 stages. As can be seen in Table 2, both the loss value and metric value improved at each training step. This indicates the autoencoder model successfully reconstructed the input image with minimal error. Therefore, we utilized the trained weights of the encoder part for the subsequent model.

Bone age estimation results

Table 3 demonstrates the parameters used

Table 1: Autoencoder parameters

Parameters	Value
Activation function of CNNs	Relu
Activation function of output	Sigmoid
Optimizer	Adam
Loss	MSE
Metrics	Accuracy
Batch size	8
Learning rate	0.001
Image size	256×256

Relu: Rectified Linear Unit, CNNs: Convolutional Neural Networks, MSE: Mean Square Error

Table 2: Autoencoder results

Epoch	Loss	Metric
	MSE	Accuracy
Step1 = 300 epoch	0.000206	0.4866
Step2 = 200 epoch	0.000172	0.6169
Step3 = 100 epoch	0.000166	0.6819
Step4 = 150 epoch	0.000156	0.6830
Step5 = 150 epoch	0.000148	0.6839

MSE: Mean Square Error

for training the regression model for bone age estimation. Accordingly, the Adam optimizer is used to train our model with an initialization learning rate of 0.001. Again, a batch size of 16 and the L_2 regularization weight of 0.0001 are selected for this phase.

Figure 3 provides the curves for the mean absolute error and mean square error of the final model during training and validation. With the increase in the number of epochs, a significant reduction is observed in both the mean absolute error and mean square error values. Eventually, both curves stabilize after about

ten epochs.

In order to visually assess the performance of the proposed model for bone age estimation, the outcomes of the model's predictions for the test set are displayed in Figure 4. The blue dot points show the anticipated values, while the green line indicates the real bone age. This 45-degree diagonal green line is used as a reference to gauge the accuracy of the model predictions, which represents the ideal case where the predicted ages perfectly match the actual ages. As can be observed, the predicted values are closely aligned with the green line, showing that the suggested approach performs well across all age ranges.

Comparing various bone age estimation techniques in related works is challenging since they employ various datasets with diverse evaluation protocols. Thus, conducting a fair comparison would be difficult. Here proposed AE-BoNet method is compared with the two most related works (Wibisono et al. [18] and Gao et al. [19]) that use RSNA for evaluation (Table 4).

In addition, results are compared with four well-known CNN networks (VGG16 [20], InceptionV3 [21], ResNet50 [22], Xception [23], and MobileNet [24]) trained on the ImageNet dataset using a transfer-learning approach (i.e., for the backbone network) to investigate the effectiveness of the proposed

Table 3: Final proposed model parameters

Parameters	Value
Activation function of CNNs	Relu
Activation function of output	Linear
Optimizer	Adam
Loss	MSE
Metrics	MAE
Batch size	16
Learning rate	0.001
L_2 regularization weight	0.0001
Image size	256×256

CNNs: Convolutional Neural Networks, Relu: Rectified Linear Unit, MSE: Mean Square Error, MAE: Mean Absolute Error

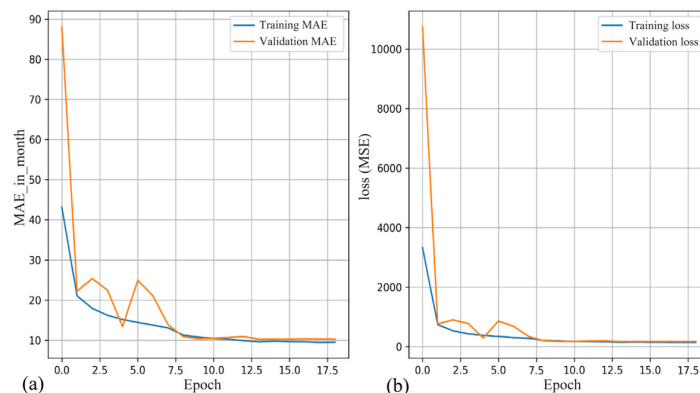


Figure 3: Training curves for the suggested final model. (a) shows mean absolute error curves for the train set and validation set (b) loss curves for the train set and validation set considering mean square error as loss function (MSE: Mean Square Error, MAE: Mean Absolute Error)

AE-BoNet method. Table 5 shows the results for the RSNA dataset using two different image sizes of 256×256 and 450×450 . Therefore, these results are on the same data with different sizes.

Discussion

BoNet [1] was one of the first studies to

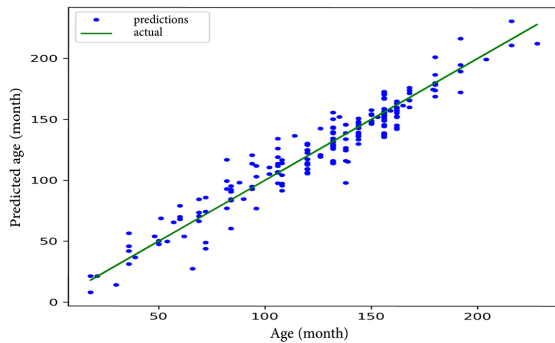


Figure 4: Prediction on 200 test set images. The green line displays the actual bone age, and the blue dots show the predicted values

Table 4: Comparisons with related works

Method	MAE (month)	Dataset	Year
Wibisono et al. [18]	14.78	RSNA	2019
Gao et al. [19]	9.997	RSNA	2020
AE-BoNet	9.3	RSNA	2023

MAE: Mean Absolute Error, RSNA: Radiological Society of North America, AE-BoNet: Autoencoder Bone Age Network

create an end-to-end convolutional neural network for deep learning-based bone age assessments. Convolutional neural networks automatically learn the hierarchies of discriminative features. Consequently, the requirement is eliminated for feature engineering [11, 25]. Supervised learning of deep networks requires a substantial amount of labeled data. However, annotating a large dataset in the medical field can be extremely time-consuming and expensive. A typical approach is to use pre-trained networks and fine-tune them to the available data [26].

The proposed method in the study of Koitka et al. [27] includes regression and detection networks; they used the Faster-RCNN architecture [28] with Inception-ResNet-V2 [29] as a feature extractor and developed a network, trained using 240 manually labeled photographs. Gao Y et al. [19] manually segmented one thousand samples from a dataset to create the mask image before training U-Net. The trained U-Net was then used for segmenting the whole images of the original dataset. Pan et al. [30] also utilized transfer learning with various networks, including the Xception [23], the ResNet-50 [22], the Inception-ResNet-V2 [29], VGG-16-19 [20], and Inception-V3 [21], with weights already trained on ImageNet. Furthermore, they employed active learning (AL) [30] to segment images using 300 manually annotated images. Ren et al. [31] used transfer learning and

Table 5: Comparisons results of the proposed method with conventional pre-trained Convolutional Neural networks on RSNA (Radiological Society of North America) data

Pretrained Network	Image size= 450×450		Image size= 256×256	
	MAE	MSE	MAE	MSE
Xception	14.7	352.8	16.1	413.6
VGG16	15.8	427.6	19.1	661.2
InceptionV3	13.6	295.5	16.1	416.3
MobileNet	12.5	268.8	12.9	296.8
AE-BoNet	9.9	161.1	9.3	144.6

MAE: Mean Absolute Error, MSE: Mean Square Error, VGG: Visual Geometry Group, AE-BoNet: Autoencoder Bone Age Network

deployed the Inception-V3 network for feature extraction. Additionally, rectangular bounding boxes were created to trim and localize the foreground for one thousand images. Salim et al. [32] initially employed mask R-CNN [33] for instance, segmentation, and background removal. Then, a regression network architecture was employed with a pre-trained VGG-19 convolutional neural network model to estimate bone age. Liu B et al. [34] applied a CNN network (VGG-U-Net) previously trained on ImageNet, to segment the hand and wrist from the X-ray images. A conditional GAN network was established to determine bone age. Chen et al. [35] employed a localization network with the InceptionV3 backbone and a regression network with the Xception backbone. Zulkifley et al. [36] presented a method for bone age assessment using image registration. The hand image was segmented based on DeepLab V3+ architecture [37], which uses the Xception as its backbone, and then MobileNetV1 [24] was used for the key points regressor for angle alignment and angle calculating from the four key points of interest. Hao et al. [38] suggested RT-FuseNet, in which residual block in ResNet was used to extract coarse feature maps with an attention mechanism. Three fully connected layers were utilized to extract textual features, and also the role of transfer learning was investigated. Hao et al. [39] used the transfer learning approach and used EfficientNet for feature extraction.

The primary issue with transfer learning using pre-trained networks like VGG, Xception, and MobileNet is that these models were not specifically trained on hand radiographic images. Instead, they were trained on large-scale datasets like ImageNet. To address this limitation, this paper introduces the concept of unsupervised pre-training using a CNN autoencoder trained specifically on hand radiography images, which directly relates to the bone age problem.

To investigate the effectiveness of the number of epochs, the training of the autoencoder

was performed on the RSANA dataset in five steps. According to Table 2, the value of the loss decreased steadily at each stage of the training process, and as a result, the reconstructed image would be highly comparable to the original. Therefore, by raising the number of epochs, the autoencoder model was able to learn the useful features from the input, leading to improved performance of the transferred weights for the final model. Consequently, the encoder portion can be reused for subsequent networks. The final model for estimating bone age reuses the extracted features previously learned by the encoder. Thus, this method allows for a more efficient training process with less training data.

A comparison between this paper and other related works is shown in Table 4. Wibisono et al. [18] utilized pre-trained VGG16 and MobileNet for a deep learning approach, in which the mean absolute error achieved by pre-trained VGG16 was 14.78 months. Gao et al. [19] used U-Net for hand bone region segmentation and pre-trained VGG16 as a model backbone, and they reached the mean absolute error of 9.997 months. Therefore, the approach adopted in this article performed better than other works mentioned, as shown in Table 4. Furthermore, Table 5 demonstrates that the suggested approach in this study achieved better outcomes than other pre-trained models.

Conclusion

This study highlights the significance of employing autoencoders and unsupervised learning as alternatives to conventional methods in the domain of medical image analysis. The proposed AE-BoNet model has shown encouraging outcomes, achieving a Mean Absolute Error score of 9.3 on the RSNA test set. Moreover, the experimental results suggest that this novel approach can be effectively utilized in various medical image decision-making scenarios, especially when there is a scarcity of labeled data or no similar pre-trained model is

accessible.

Acknowledgment

The authors thank Dr. Raheleh Kafieh, Assistant Professor of Durham University, for her valuable comments.

Authors' Contribution

A. Chaparian conceived the idea. The draft of the paper was written by M. Sirati-Amsheh. The final review and editing of the article were done by M. Sirati-Amsheh, E. Shabaninia, and A. Chaparian. The method implementation was carried out by M. Sirati-Amsheh. Results and Analysis were carried out by M. Sirati-Amsheh and E. Shabaninia. The research work was proofread and supervised by A. Chaparian and E. Shabaninia. All the authors read, modified, and approved the final version of the manuscript.

Ethical Approval

This article does not contain any studies with human participants or animals performed.

Funding

This study was funded by the Isfahan University of Medical Sciences with the approved project number: 340104.

Conflict of Interest

None

References

1. Spampinato C, Palazzo S, Giordano D, Aldinucci M, Leonardi R. Deep learning for automated skeletal bone age assessment in X-ray images. *Med Image Anal.* 2017;**36**:41-51. doi: 10.1016/j.media.2016.10.010. PubMed PMID: 27816861.
2. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. *The American Journal of the Medical Sciences.* 1959;**238**(3):393.
3. Tanner JM. Assessment of Skeletal Maturity and Predicting of Adult Height (TW2 Method). *Prediction of Adult Height.* 1983:22-37.
4. Tanner J, Healy M, Goldstein H, Cameron N. Prediction of Adult Height TW3 equations. In: Assessment of Skeletal Maturity and Prediction of adult height by TE3 method, 3rd ed. London, WB Saunders; 2001. p. 26-43.
5. Li K, Zhang J, Sun Y, Huang X, Sun C, Xie Q, Cong S. Automatic bone age assessment of adolescents based on weakly-supervised deep convolutional neural networks. *IEEE Access.* 2021;**9**:120078-87. doi: 10.1109/ACCESS.2021.3108219.
6. Fritz B, Marbach G, Civardi F, Fucentese SF, Pfirrmann CWA. Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference. *Skeletal Radiol.* 2020;**49**(8):1207-17. doi: 10.1007/s00256-020-03410-2. PubMed PMID: 32170334. PubMed PMCID: PMC7299917.
7. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal Radiol.* 2020;**49**(2):183-97. doi: 10.1007/s00256-019-03284-z. PubMed PMID: 31377836.
8. Grauhan NF, Niehues SM, Gaudin RA, Keller S, Vahldiek JL, Adams LC, Bressemer KK. Deep learning for accurately recognizing common causes of shoulder pain on radiographs. *Skeletal Radiol.* 2022;**51**(2):355-62. doi: 10.1007/s00256-021-03740-9. PubMed PMID: 33611622. PubMed PMCID: PMC8692302.
9. Guan B, Liu F, Mizaian AH, Demehri S, Samsonov A, Guermazi A, Kijowski R. Deep learning approach to predict pain progression in knee osteoarthritis. *Skeletal Radiol.* 2022;**51**(2):363-73. doi: 10.1007/s00256-021-03773-0. PubMed PMID: 33835240. PubMed PMCID: PMC9232386.
10. Yi PH, Kim TK, Wei J, Li X, Hager GD, Sair HI, Fritz J. Automated detection and classification of shoulder arthroplasty models using deep learning. *Skeletal Radiol.* 2020;**49**(10):1623-32. doi: 10.1007/s00256-020-03463-3. PubMed PMID: 32415371.
11. Zulkifley MA, Mohamed NA, Abdani SR, Kamari NAM, Moubark AM, Ibrahim AA. Intelligent Bone Age Assessment: An Automated System to Detect a Bone Growth Problem Using Convolutional Neural Networks with Attention Mechanism. *Diagnostics (Basel).* 2021;**11**(5):765. doi: 10.3390/diagnostics11050765. PubMed PMID: 33923215. PubMed PMCID: PMC8146101.
12. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, et al. The RSNA Pediatric Bone Age Machine Learning Challenge. *Radiology.* 2019;**290**(2):498-503. doi: 10.1148/radiol.2018180736. PubMed PMID: 30480490. PubMed PMCID: PMC6358027.
13. Oh M, Zhang L. DeepMicro: deep representation learning for disease prediction based on microbiome data. *Sci Rep.* 2020;**10**(1):6026. doi: 10.1038/s41598-020-63159-5. PubMed PMID: 32265477. PubMed PMCID: PMC7138789.
14. Yang Y, Wu QJ, Wang Y. Autoencoder with invertible functions for dimension reduction and image reconstruction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems.* 2016;**48**(7):1065-79. doi: 10.1109/TSMC.2016.2637279.

15. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32 nd International Conference on Machine Learning; Lille, France: JMLR; 2015.
16. Van Laarhoven T. L2 regularization versus batch and weight normalization [Internet]. arXiv [Preprint]. 2017 [cited 2017 Jun 16]. Available from: <https://arxiv.org/abs/1706.05350>.
17. Kingma DP, Ba J. Adam: A method for stochastic optimization [Internet]. arXiv [Preprint]. 2014 [cited 2014 Dec 22]. Available from: <https://arxiv.org/abs/1412.6980>.
18. Wibisono A, Saputri MS, Mursanto P, Rachmad J, Yudasubrata AT, Rizki F, Anderson E. Deep learning and classic machine learning approach for automatic bone age assessment. In 4th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS); Nagoya, Japan: IEEE; 2019. p. 235-40.
19. Gao Y, Zhu T, Xu X. Bone age assessment based on deep convolution neural network incorporated with segmentation. *Int J Comput Assist Radiol Surg*. 2020;**15**(12):1951-62. doi: 10.1007/s11548-020-02266-0. PubMed PMID: 32986142.
20. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. [Internet]. arXiv [Preprint]. 2014 [cited 2014 Sep 4]. Available from: <https://arxiv.org/abs/1409.1556>.
21. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); Las Vegas, NV, USA: CVPR; 2016. p. 2818-26.
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); Las Vegas, NV, USA: CVPR; 2016. p. 770-8.
23. Fran C. Xception: Deep learning with depthwise separable convolutions. In IEEE conference on computer vision and pattern recognition (CVPR); Honolulu, HI, USA: CVPR; 2017.
24. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [Internet]. arXiv [Preprint]. 2017 [cited 2017 Apr 17]. Available from: <https://arxiv.org/abs/1704.04861>.
25. He J, Jiang D. Fully automatic model based on se-resnet for bone age assessment. *IEEE Access*. 2021;**9**:62460-6. doi: 10.1109/ACCESS.2021.3074713.
26. Koitka S, Demircioglu A, Kim MS, Friedrich CM, Nensa F. Ossification area localization in pediatric hand radiographs using deep neural networks for object detection. *PLoS One*. 2018;**13**(11):e0207496. doi: 10.1371/journal.pone.0207496. PubMed PMID: 30444906. PubMed PMCID: PMC6239319.
27. Koitka S, Kim MS, Qu M, Fischer A, Friedrich CM, Nensa F. Mimicking the radiologists' workflow: Estimating pediatric hand bone age with stacked deep neural networks. *Med Image Anal*. 2020;**64**:101743. doi: 10.1016/j.media.2020.101743. PubMed PMID: 32540698.
28. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems 28 (NIPS 2015). NeurIPS Proceedings; 2015.
29. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-First AAAI Conference on Artificial Intelligence; California USA: AAAI Press; 2017.
30. Pan X, Zhao Y, Chen H, Wei D, Zhao C, Wei Z. Fully Automated Bone Age Assessment on Large-Scale Hand X-Ray Dataset. *Int J Biomed Imaging*. 2020;**2020**:8460493. doi: 10.1155/2020/8460493. PubMed PMID: 32190035. PubMed PMCID: PMC7072110.
31. Ren X, Li T, Yang X, Wang S, Ahmad S, Xiang L, Stone SR, Li L, Zhan Y, Shen D, Wang Q. Regression Convolutional Neural Network for Automated Pediatric Bone Age Assessment From Hand Radiograph. *IEEE J Biomed Health Inform*. 2019;**23**(5):2030-8. doi: 10.1109/JBHI.2018.2876916. PubMed PMID: 30346295.
32. Salim I, Hamza AB. Ridge regression neural network for pediatric bone age assessment. *Multimedia Tools and Applications*. 2021;**80**(20):30461-78. doi: 10.1007/s11042-021-10935-8.
33. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (ICCV); Venice, Italy: ICCV; 2017. p. 2961-9.
34. Liu B, Zhang Y, Chu M, Bai X, Zhou F. Bone age assessment based on rank-monotonicity enhanced ranking CNN. *IEEE Access*. 2019;**7**:120976-83. doi: 10.1109/ACCESS.2019.2937341.
35. Chen C, Chen Z, Jin X, Li L, Speier W, Arnold CW. Attention-Guided Discriminative Region Localization and Label Distribution Learning for Bone Age Assessment. *IEEE J Biomed Health Inform*. 2022;**26**(3):1208-18. doi: 10.1109/JBHI.2021.3095128. PubMed PMID: 34232898.
36. Zulkifley MA, Abdani SR, Zulkifley NH. Automated bone age assessment with image registration using hand X-ray images. *Applied Sciences*. 2020;**10**(20):7233. doi: 10.3390/app10207233.
37. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV); Munich, Germany: 2018. p. 801-18.
38. Hao P, Ye T, Xie X, Wu F, Ding W, Zuo W, Chen W, Wu J, Luo X. Radiographs and texts fusion learning based deep networks for skeletal bone age assessment. *Multimedia Tools and Applications*. 2021;**80**:16347-66. doi: 10.1007/s11042-020-08943-1.
39. Hao G, Li Y. Bone age estimation with x-ray images based on efficientnet pre-training model. *Journal of Physics: Conference Series*. 2021;**1827**:012082. doi: 10.1088/1742-6596/1827/1/012082.