

RESEARCH

Open Access



# Multi Region-Based Feature Connected Layer (RB-FCL) of deep learning models for bone age assessment

Ari Wibisono\*  and Petrus Mursanto

\*Correspondence:  
ari.w@cs.ui.ac.id  
Faculty of Computer Science,  
Universitas Indonesia,  
Kampus UI Depok, Depok  
City 16424, Indonesia

## Abstract

Prediction of bone age from an x-ray is one of the methods in the medical field to support predicting endocrine gland disease, growth abnormalities, and genetic disorders. A decision support system to predict the bone age from the x-ray image has been implemented. It utilizes traditional machine learning methods and deep learning. We propose the Region-Based Feature Connected Layer (RB-FCL) from the essential segmented region of hand x-ray. We treat the deep learning models as the feature extraction for each region of the hand x-ray bone. The Feature Connected Layers are the output from the trained important region, such as 1-radius-ulna, 2-carpal, 3-metacarpal, 4-phalanges, and 5-epiphysis. DenseNet121, InceptionV3, and InceptionResNetV2 are the deep learning models that we used to train the critical region. From the evaluation results, the Mean Absolute Error (MAE) results produced is 6.97. This result is better compared to standard deep learning models, which are 9.41.

**Keywords:** Region-Based Feature Connected Layer, Bone age assessment, Deep learning, X-ray images

## Introduction

A method that is utilized to identify and estimate bone age is called bone age assessment. Bone age from the x-ray pictures can be estimated from the time of little children to youngsters. Bone development is not just impacted by genetic disorders, hormones, and supplements. It is also impacted by disease and mental conditions. Abnormal growth can be caused by several factors, such as genetic disorders, endocrine issues, and pediatric disorders [1–4].

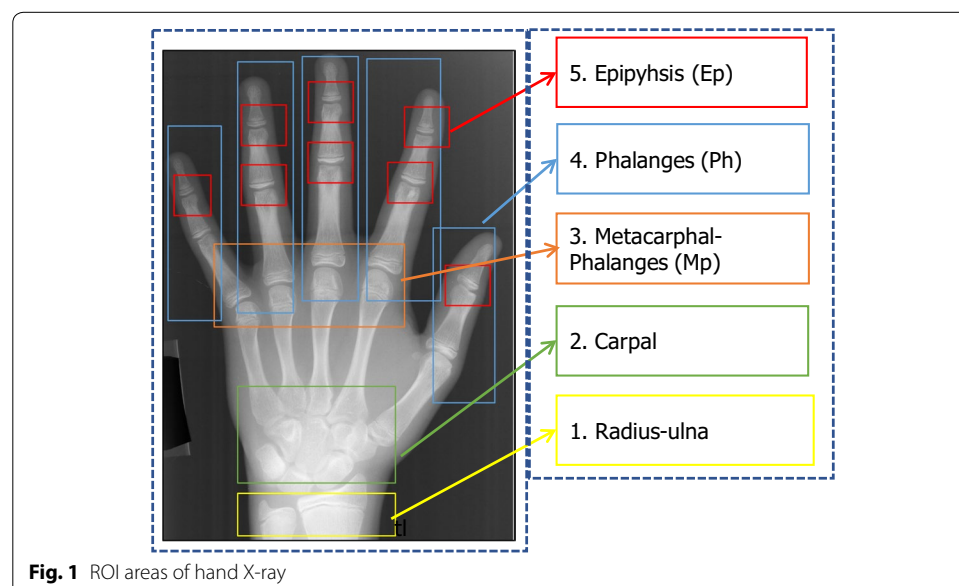
Medical references explain that among several parts of the body, x-ray images of the left wrist can be used to evaluate bone growth. Manually, a radiologist uses two methods to evaluate bone age. These methods are the Greulich–Pyle (GP) and the Tanner–Whitehouse (TW) method [5]. TW uses the scoring method to determine bone age, while the GP method uses the atlas reference from bone age data [6]. Manual assessment of hand radiographs takes a long time and is quite expensive. So we need an automated recognition system that can recognize the age of bone based on the principles of medical science that are studied by radiologists.

In the last decade, evaluation of bone age has become essential to reduce the problems in the manual method for bone age estimation [7]. The main challenge is choosing the most appropriate method for building a bone age prediction system. In general, two methods can be done. The first is the use of image processing to retrieve features that affect bone development. These features will be input for the machine learning algorithm to make predictions. This process is commonly referred to as traditional machine learning or handcrafted method [8]. The second approach is to use deep learning convolutional neural networks. Automated feature extraction has been performed when the convolution occurs, so that prediction of bone age can be directly predicted.

TW method is implemented by Davies et al. which extracted edges, and critical points for local image features [9]. Some local image extraction work to predict bone age is done by Zhang et al. [10]. They implemented fuzzy classification for predicting bone age. Somkantha et al. extracted carpal bones edge and Support Vector Regressor to estimate bone age. The histogram of Oriented Gradient (HOG) and BoG (Bag of Visuals Words) is classified with the Random Forest algorithm [11].

Two cutting edge techniques that are used by radiologists to do Bone Ages Assessment are the Greulich–Pyle (GP) [12] and Tanner–Whitehouse (TW) technique [13]. The GP strategy runs dependent on a current hand atlas. The format incorporates x-ray pictures from 0 to 18 years. The GP strategy works dependent on coordinating the x-ray picture that has been acquired with a current hand atlas reference. This methodology is not challenging to do and can be utilized by many radiologists. However, the GP method has a weakness. The outcomes may vary from one radiologist to the other radiologist.

The TW strategy assesses by evaluating the significant regions of the bone x-ray. Region of Interest (ROI) is utilized to see the significant parts in the bone that decide the bone development. Those parts are Ulna, Epyphysis, Metaphysis, Radius, Phalanx, and Metacarpal which are shown in Fig. 1.



This paper consists of five sections. The first section consists of the introduction and background of this paper. The second section explains our research position and literature review. The third section explains our proposed method. The fourth section is the experiment result, and the last section consists of our discussions.

### Related works

Spampinato et al. has utilized a deep learning approach to predict the bone age of children or teenagers [14]. They experiment with a few deep learning models, for example, Bonet, Googlenet, and Oxford. The BAA result from their experiment can deliver MAE for around 9.6 months. The dataset is assembled from an open dataset got from the Digital hand atlas. The number of datasets utilized was 1391 x-ray pictures [15].

Castillo et al. estimated bone age by utilizing the VGG-16 model [16]. The dataset that is used is the RSNA dataset. It consists of 12,611 x-ray pictures. The MAE result of their experiment was 9.82 months for male patients and 10.75 months for female patients. Lee et al. contributed to segmenting the standardizing processes, segmenting the Region of Interest, pre-process radiographs, and estimating the bone age assessment. The assessment results have indicated 57.32% and 61.40% precision for the forecast of the age of women and men, respectively [6]. The dataset consists of 4047 for male and 4278 for female x-ray picture.

Wang et al. utilized an alternate methodology in the field of bone age assessment [17, 18]. Given medical references, they categorize bone parts based on the development of the bone components that are appeared in x-ray pictures. It utilized a Faster Region Convolutional Neural Network as the deep learning model [19]. It utilized 600 information for the radius bone and 600 information for the ulna bone. It acquired 92% accuracy for the radius and 90% for the ulna.

Son et al. added to the automatic of the Tanner Whithouse (TW 3) strategy, which is a reference in bone age evaluation [20]. Confinement of the bone epiphysis and metaphysis was done to estimate the age of the bone. The dataset is consists of 3300 x-ray pictures from medical clinics in South Korea. The classification results for the bone area show a precision of 79.6% and 97.2% for top-1 and top-2 accuracy. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are 5.62 and 7.44. Liu et al. did a different method regarding pre-processing to estimate bone age. Non-subsampled Contourlet Transform (NSCT) is done before training with a deep learning model [21]. The dataset utilized is the open digital hand atlas dataset. Generally, the RMSE created from this strategy is 8.28.

Bone information is not just utilized in the medical field. In any case, the bone picture is additionally required in the field of paleontology and taphonomy. Bone information is utilized to get some answers concerning archeological and paleontological locales [22]. Explicit bone age forecast is utilized to discover and investigate historical timelines. Knowing when people begin to eat meat, utilize stone apparatuses, investigate new mainlands, and collaborate with savage creatures. Bone surface alteration is recognized by utilizing a deep learning model. Automatic identification is made by utilizing scratched information on fleshed and defleshed bone.

A few scientists use traditional and deep learning to estimate bone age from x-ray images. The utilization of regression to identify bone age has been utilized by a few

analysts [23–25]. Furthermore, the utilization of random forest [26], K-NN [27], SVM [28–30], ANN [24, 31, 32], and Fuzzy Neural system [33] has been done by a few authors. The utilization of deep learning models also has been contributed by certain scientists to estimate the bone age [6, 34–36, 54].

The other researcher uses the landmark-based multi-region ensemble CNN for bone age assessment [37]. This work differs from our work in terms of the concatenation of layer, the evaluation, and the proportion of data. We combine the connected feature layers of some regions. However, their work directly using input image and segmented to a few regions. The evaluation of this work only uses each of the regions as a comparison. In our research, we evaluate the whole segmented regions that produce Feature Connected Layers. In terms of the dataset, they evaluate the bone age dataset from digital hand atlas with a proportion of 90% training and 10% testing. However, in our work, we evaluate with two public datasets, digital hand atlas dataset (1392 x-ray images) [38] and RSNA dataset (12,814 x-ray images) [39]. The evaluation proportion of our work is 80% training and 20% testing.

Based on previous references, the proportion of training and testing data tested is 90% and 10% [37]. By using many datasets, we can provide the opportunity for the model to test its performance with less training data. We used two datasets; X-Ray digital hand atlas dataset totaling 1392 samples and RSNA dataset 12,814 samples. This large dataset is possible to be tested with a smaller proportion of training compared to the proportion in [37]. With a proportion of 80% training and 20%, we can provide an opportunity for models to be trained with lower training data but resulting in a good performance.

The performance of the bone age assessment method is presented by Dallora et al. [40]. It shows the machine learning algorithm performance result for each dataset. It gives us wholistic information about the current machine learning performance to estimate bone age. Region detection and maturity classification are proposed by Bui et al. [41]. They utilize it to estimate the bone age. Based on the experiment result by using Digital Hand Atlas Dataset, the performance of the MAE is 7.1 months. The performance of deep learning methods to estimate the bone age is presented by Larson et al. [42]. Also, the large Scale Hand X-Ray Dataset bone age estimation is proposed by Pan et al. [43]. The other researcher use two step method in bone age estimation. The authors use deep learning method as a feature extraction then classified it with the age group of the bone [44].

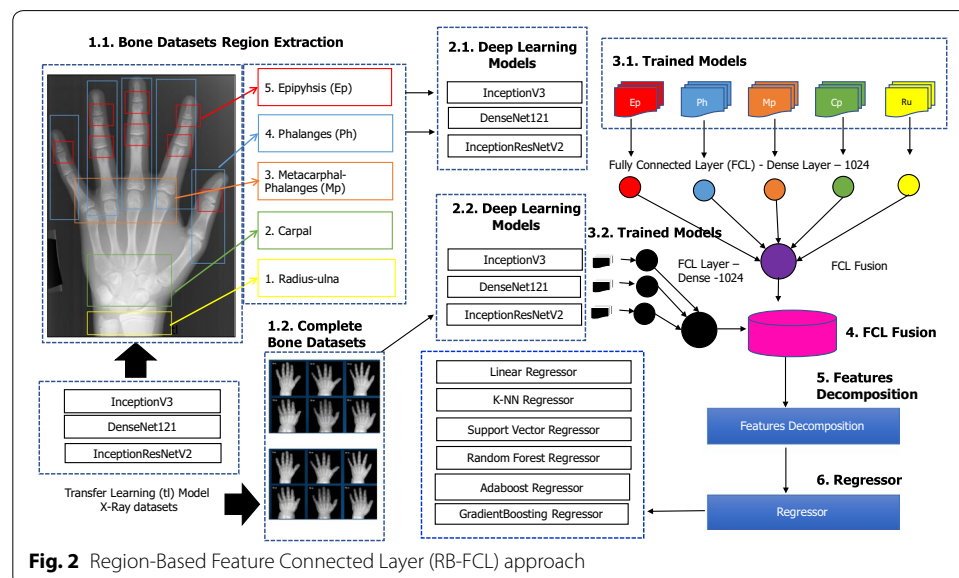
In this research, we try to segment the most important parts of the bone, which is the critical region to estimate bone growth. The baseline method (a manual process) to do bone age assessment is TW and GP, which have been introduced in the introduction sections. We proposed a segmentation in the important area suggested by TW strategy. Those parts are Ulna, Epiphysis, Metaphysis, Radius, Phalanx, and Metacarpal. We choose to follow TW strategy because this method evaluates significant regions of the bone x-ray rather than depend on hand atlas picture as a reference. The essential parts are referred to from the TW method. We use deep learning as a method for extracting Feature Connected Layers (FCL). FCL Concatenation is done to predict the estimated age of bone age using several regressor methods.

## Proposed method

In this research, we contributed to create an age prediction expert system from hand x-ray. We do the segmentation of the essential parts of bone x-ray. Based on the radiologist's reference, the radius-ulna, carpal, metacarpal, phalanges, and epiphysis sections are parts that can affect the age of the bone. The results of the segmentation of these parts are trained in deep learning to produce a Feature Connected Layer feature (FCL). Several scenarios are carried out to produce the smallest prediction error. Based on the results of the trial, merging some connected layer features from the segmentation section can produce the smallest MAE error with a value of 6.97 months.

There are two flows to do FCL Fusion. In the first flow, the bone dataset is segmented based on the critical regions in determining bone age. The results of each region segmentation are trained using deep learning models and produce FCL with 1024 dense features. Flow 1 has process identifiers 1.1, 1.2, 1.3, and 1.4 in Fig. 2. Figure 2 shows the segmentation results of the hand x-ray. Part 1, which is yellow, is Radius-Ulna. The second part is carpal with green color. The third part is the space between metacarpal and phalanges with people's color. The fourth part is phalanges with blue. The fifth part is the space between phalanges with the name ephypsis in red. In the second flow, the whole hand bone image is trained using several deep learning models. We extracted FCL with 1024 dense features. The results of the dense layer will be combined with the results in the first path. The way strand is identified in process numbers 2.1, 2.2.

Automatic segmentation is done by using the Faster R-CNN standard to separate essential regions from the original image [45, 46]. Region algorithm is implemented in Faster R-CNN. It utilize Region Proposal Network (RPN) to produce the region proposal. It gives around 0.2 s computation time to detect an image. Each of these regions based training is conducted on several deep learning models, namely InceptionV3, Densenet121, and InceptionResnetV2. The selection of this deep learning model is based on the evaluation of FCL results from the deep learning evaluation



shown in Table 2. In both the first and second flows, we use transfer learning from weights derived from the results of x-ray [47, 48].

To predict bone age, researchers use several layers in the deep learning model to predict accurately. In the deep learning model, there are several components, including the input layer, the convolution layer, the pooling layer, and the Feature Connected Layer (FCL). In this research, we treat FCL as a result of feature extraction from bone images. FCL of several deep learning models fusion treated as input features to be included in regressors. Several variations of the integration of FCL are combined to obtain the best accuracy results. FCL layers are taken from the deep learning model DenseNet121 [49], InceptionV3 [50], and InceptionResNetV2 [51, 52].

The first process flow is indicated by the explanation of Eq. 1 through Eq. 9

$$K = [k_0, k_1, k_2, \dots, k_n] \quad (1)$$

$$L = RCNN(k) \quad (2)$$

$$L = \{[L_i]; i = 0, 1, 2, 3, 4\} \quad (3)$$

If K is an image of a hand bone in Eq. 1, X-Ray and L is the result of region segmentation using RCNN in Eq. 2. There are five results of the segmentation matrix derived from the RCCN (K) process with notation;  $i=0, 1, 2, 3, 4$

$$FCLInceptionV3(L) = \{[M_i]; i = 0, 1, 2, 3, 4\} \quad (4)$$

$$FCLDenseNet121(L) = \{[N_i]; i = 0, 1, 2, 3, 4\} \quad (5)$$

$$FCLInceptionResnetV2(L) = \{[O_i]; i = 0, 1, 2, 3, 4\} \quad (6)$$

Each region is generated from RCNN will extract its FCL using several deep learning models with FCL M, N, O results. There are five matrices for each FCL deep learning models result. Each FCL layer result has 1024 dense features.

$$AM = [M_0|M_1|M_2|M_3|M_4] \quad (7)$$

$$AN = [N_0|N_1|N_2|N_3|N_4] \quad (8)$$

$$AO = [O_0|O_1|O_2|O_3|O_4] \quad (9)$$

AM in Eq. 7 is the result of the combined concatenation of the FCL layer matrix results for each region generated by InceptionV3. AN is the concatenation of the combined FCL layer matrix results for each region produced by DenseNet121. AO is the result of the combined FCL layer matrix results for each region produced by InceptionResNetV2.

The second process flow is shown by the explanation of Eq. (10). If K is an image of hand bone X-Ray and W, X, Y is the result of FCL extraction from the whole image.

$$FCLInceptionV3(K) = W \quad (10)$$

$$FCLDenseNet121(K) = X \quad (11)$$

$$FCLInceptionResnetV2(K) = Y \quad (12)$$

$$AZ = [W|X|Y] \quad (13)$$

Suppose  $W$  is the FCL result from InceptionV3,  $X$  is DenseNet121, and  $Y$  is InceptionResNetV2. Each of the FCL output has 1024 output features. The results of combining FCL from three deep learning models are explained in Eq. 13 notation. Concatenation results will be processed by PCA feature decomposition with 50 components, labeled with variable  $P$ , as can be shown in Eq. 6. The scenario is done by combining the matrix between  $AM$ ,  $AN$ ,  $AO$ , and  $AZ$  as  $P$ .  $P$  notation will be included in the PCA Feature Decomposition.

$$P : PCA(AM|AN|AO|AZ) \rightarrow P(p_0, p_1, p_2, \dots, p_{49}) \quad (14)$$

$$G = \{g : g = 1, \text{male}, g = 0, \text{female}\} \quad (15)$$

$$F = [P|G] \quad (16)$$

$$BA = \text{Regressor}(F) \quad (17)$$

Variable  $G$  is a gender variable,  $G$  is 1 for men and 0 for women. The conjugate results of  $P$  and  $G$  are labeled with variable  $F$  in Eq. 16. Bone age prediction is labeled with  $BA$  notation.  $BA$  is generated from the regressor results using the  $F$  features conjugation. We consider using the FCL output from Multi-Path Connectivity and the depth revolution represented by DenseNet and Resnet. In addition, we also tested the output of the deep learning model with Spatial Exploitation, Parallelization, and Inception Block, which is represented by InceptionV3 and InceptionResNetV2. We consider gender as a feature to determine the age of bone images. Feature decomposition is done using Principal Component Analysis (PCA) with a total of 50 components. After that, the gender feature is combined with FCL results from the deep learning model. A complete diagram of the process that we carried out is shown in Fig. 2.

We use Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |r_i - t_i|, \quad (18)$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \times 100\% \sum_{i=1}^n \left| \frac{r_i - t_i}{t_i} \right|, \quad (19)$$

and Root Mean Squared Error (RMSE).



$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |r_i - t_i|^2} \quad (20)$$

$n$  is the total of data,  $r_i$  is the forecasted value, and  $t_i$  is ground truth value.

## Result

The hardware specifications that we use in this research are Intel(R) Core(TM) i7-6800K CPU @ 3.40 GHz, 32 GB Physical RAM, and 6 GPU NVIDIA GTX 1080 Ti × 11 GB VRAM. We utilized Ubuntu 16.04 as the operating system. We utilized TensorFlow and Keras framework model on top of python programming language to evaluate our proposed method. Keras model application is also be used by Ren et al. to perform the augmentation of the bone [36]. The other researcher uses GoogleNet and ImageNet model to perform the simulation [34]. We use standard input for InceptionV3 and InceptionResNetV2 (299 × 299). DenseNet121 and ResNet50 use 224 × 224 for the size of the input. We have added three dense layers at the end of our network. The standard depth InceptionV3, InceptionResnetV2, DenseNet121 are 159,572, 121 respectively. We consider having 99.99% variety from the principal component from the features. Thus we choose 50 component for the principal components

In general, we conduct four test scenarios. All scenarios are performed on two public datasets, namely digital hand atlas dataset and RSNA dataset. The first test scenario is carried out to evaluate errors in standard deep learning models. The label we give in this scenario is std. The test results are shown in Table 1. The second scenario is the test scenario using a single Feature Connected Layer. The FCL output of scenario 2 is produced by each deep learning model. The label we gave in the second scenario is FCL. The extraction results from 1024 dense feature of FCL will be input to be tested on several regressor algorithms. Table results of the scenario 2 test results are shown in Tables 2 and 3.

The third scenario that we do is to merge the five FCL layers by using Region-Based Feature Connected Layer (RB-FCL). The FCL is produced by each region that has been trained using several deep learning models. The regions are 1-radius-ulna, 2-carpal, 3-metacarpal-phalanges, 4-phalanges, and 5-epiphysis. The results of scenario three are shown in Tables 3 and 4. The label for the third scenario is RB-FCL. The fourth scenario that we do is to do the feature layer concatenation of scenario two (FCL) and scenario three (RB-FCL). In the fourth scenario, we combine the RB-FCL output produced by InceptionV3, DenseNet121, and InceptionResNetV2. We provide IRD labels for the merged features. In general, the smallest MAE value is obtained from the concatenation features merge in the fourth scenario, which is 6.97 months.

**Table 1** Result of standard deep learning models evaluation

Dataset	InceptionV3-std			DenseNet121-std			InceptionResNetV2-std		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
Digital hand atlas	9.41	11.83	10.17	11.76	14.94	12.64	12.71	16.63	15.32
RSNA	10.89	15.07	12.62	12.87	16.24	14.71	12.11	15.68	12.25



**Table 2 Result of single Feature Connected Layer output (FCL) for digital hand atlas dataset**

Regressor	InceptionV3-FCL			DenseNet121-FCL			InceptionResNetV2-FCL		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
LR	10.6	14.0	11.2	10.2	13.7	12.5	10.67	15.19	10.54
KNN-R	10.7	14.4	10.8	10.3	14.2	12.4	10.70	15.60	11.07
SVR	17.9	21.8	31.0	14.2	18.1	28.7	19.20	25.08	38.28
RF-R	10.4	14.0	10.5	10.1	13.7	11.5	9.77	14.02	9.76
DT-R	14.1	19.1	13.0	16.0	21.3	17.0	14.99	20.25	13.65
ADB-R	11.8	15.9	13.2	12.5	16.3	16.0	11.75	16.63	13.54
GB-R	11.3	15.1	11.5	11.3	15.2	15.0	11.25	15.86	12.98

**Table 3 Result of single Feature Connected Layer output (FCL) for RSNA dataset**

Regressor	InceptionV3-FCL			DenseNet121-FCL			InceptionResNetV2-FCL		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
LR	10.22	13.55	11.15	10.14	13.21	11.52	12.23	17.64	14.07
KNN-R	10.45	13.97	10.68	10.38	13.84	10.99	11.80	16.59	12.88
SVR	10.62	14.09	11.22	10.50	14.36	12.12	15.10	23.68	18.32
RF-R	10.00	13.42	10.31	9.78	12.91	10.18	11.46	15.82	12.31
DT-R	13.85	18.60	13.63	13.33	18.19	13.20	15.60	21.84	15.87
ADB-R	12.75	16.30	14.55	12.65	16.51	15.40	15.59	21.07	18.11
GB-R	10.43	14.01	11.21	10.21	13.36	11.14	11.83	16.33	13.17

**Table 4 Result of Region Based Feature Layer (RB-FCL) output for digital hand atlas dataset**

Regressor	InceptionV3-RB-FCL			DenseNet121-RB-FCL			InceptionResNetV2-RB-FCL		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
LR	7.10	10.96	7.23	7.09	10.96	7.19	7.10	10.96	7.20
KNN-R	8.69	12.47	10.30	8.70	12.48	10.30	8.70	12.48	10.33
SVR	11.36	16.07	26.12	11.36	16.07	26.12	11.36	16.07	26.12
RF-R	7.91	11.61	8.41	7.91	11.61	8.48	7.89	11.56	8.50
DT-R	10.45	14.76	11.33	10.89	15.10	11.42	10.10	14.57	11.17
ADB-R	11.34	15.10	13.62	11.31	15.09	13.54	11.29	15.07	13.51
GB-R	9.08	12.91	11.05	9.26	13.00	12.25	9.02	12.50	9.51

Table 1 shows the results of evaluating the deep learning model directly in training into the hand x-ray dataset. From the evaluation results using several deep learning models, for the digital hand dataset, the best MAE value is 9.41, and for the RSNA dataset is 10.89. Experiments from the proposed method that we propose can produce MAE values up to 6.97.

Tables 2 and 3 show the test results using a single Feature Connected Layer output (FCL) scenario for the Digital Hand Atlas dataset and RSNA dataset. The FCL scenario uses a whole hand x-ray image for training into each deep learning model (InceptionV3, DenseNet121, and InceptionResNetV2). Models of the training results

are used to produce output layer features that are used as input for the regressor algorithm. The results of the test metric in Table 2 can be seen as the smallest error metric obtained by FCL output from InceptionResNetV2 with MAE values 9.77, RMSE 14.02, and MAPE 9.76. The results were obtained using the Random Forest Regressor.

Table 3 shows the smallest metric errors generated by FCL from DenseNet121 with MAE values of 9.78, RMSE 12.91, and MAPE 10.18. From the test results in Tables 1 and 2, we can see that there is only a small reduction in errors obtained by taking a single feature layer output. For this reason, we try to perform FCL concatenation experiments by using the RB-FCL scenario, which is shown in Tables 4, 5, 6, and 7.

Tables 4 and 5 show the metric error results from the Region-Based Feature Layer Output (RB-FCL) scenario. RB-FCL is done by combining 5 FCL output results from each x-ray hand region in the digital hand atlas dataset. From Table 4 The smallest metric values of MAE, RMSE, and MAPE% are shown by Linear Regression (LR). Either using the RB-FCL feature from InceptionV3, DenseNet121, and InceptionResNetV2, the resulting MAE value is quite small, namely between 7.09 and 7.11. Using the RB-FCL method can minimize the error value from the standard deep learning model evaluation in Table 1 from 9.41 to 7.10, shown in Table 4. Besides, RB-FCL also has a smaller error value when compared to the scenario of using FCL. RB-FCL can produce MAE values up to 7.14 while FCL can only have MAE values of 9.78 on testing using the RSNA dataset

Testing with RB-FCL on the RSNA dataset is shown by Table 5. From the results of testing the MAE value, the Random Forest Regressor has the smallest MAE value, which is around 7.14. The results of this error are relatively the same for the three deep learning models. Similar to testing on a digital hand atlas dataset, testing the MAE value on the RSNA dataset has a smaller value than the MAE value of the test results on the standard deep learning models that are equal to 10.09.

From Tables 4 and 5, merging region-based segmentation on RB-FCL from each hand x-ray region can produce MAE, RMSE, and MAPE values that are smaller than the metrics error values in standard deep learning models. The average value of a successful MAPE reduction is reduced by about 3–4% when compared to the standard deep learning. The region-based segmentation scenario by combining the feature layers of each region (RB-FCL) also has a smaller MAPE metric value compared to the FCL scenario. Decrease in MAPE% by around 2–3%.

**Table 5** Result of Region Based Feature Connected Layer (RB-FCL) output for RSNA dataset

Regressor	InceptionV3-RB-FCL			DenseNet121-RB-FCL			InceptionResNetV2-RB-FCL		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
LR	7.48	10.22	9.415	7.48	10.22	9.413	7.48	10.22	9.413
KNN-R	7.3	10.1	8.809	7.3	10.1	8.809	7.3	10.1	8.808
SVR	7.85	11.64	13.44	7.85	11.64	13.44	7.85	11.64	13.44
RF-R	7.14	10.13	8.252	7.14	10.13	8.284	7.14	10.12	8.36
DT-R	9.67	13.87	10.52	9.57	13.73	10.46	9.67	13.7	10.67
ADB-R	11.1	14.68	12.63	11.1	14.7	12.68	11.1	14.72	12.65
GB-R	7.67	10.63	9.088	7.57	10.49	9.365	7.76	10.67	9.461

Table 6 Result of Feature Connected Layer Output Combination (FCLOMB + IRD) for hand atlas dataset

Regressor	InceptionV3FL + IRD			DenseNet121FL + IRD			InceptionResNetV2FL + IRD			FCLOMB + IRD		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
LR	7.88	11.85	8.216	8.05	11.84	8.623	7.44	11.23	8.496	7.54	11.72	8.342
KNN-R	8.79	12.61	10.29	9.11	12.55	12.23	8.47	12.44	9.46	8.78	12.53	11.03
SVR	11.8	15.78	24.24	14.1	19.85	34.75	11.3	15.44	23.97	14.2	20.64	37.28
RF-R	8.71	12.54	8.918	8.89	12.71	9.818	8.12	11.78	7.895	8.28	12.39	9.57
DT-R	11.9	15.54	12.63	13.1	17.41	14.58	11.9	16.43	11.43	12	16.98	14.92
ADB-R	11.6	15.54	14.65	11.9	15.55	16.08	11.5	15.12	13.54	11.4	14.99	15.3
GB-R	9.75	13.81	13.44	10	13.92	12.97	11.3	16.08	19.01	9.13	12.94	10.94

**Table 7 Result of Feature Connected Layer Output Combination (FCLOMB + IRD) for RSNA dataset**

Regressor	InceptionV3-RB-FCL + IRD			DenseNet121-RB-FCL + IRD			InceptionResNetV2-RB-FCL + IRD			FCLOMB + IRD		
	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%	MAE	RMSE	MAPE%
LR	7.88	10.48	9.524	7.36	9.763	8.609	7.79	10.38	9.734	7.32	9.754	8.627
KNN-R	7.78	10.4	8.74	7.2	9.722	7.92	7.71	10.37	9.031	7.11	9.583	7.912
SVR	8.04	11	11.57	8.22	12.29	14.72	8.01	11.02	11.93	8.17	12.53	15.13
RF-R	7.55	10.08	8.153	7.07	9.497	7.807	7.53	10.25	8.931	6.97	9.346	8.128
DT-R	10.5	14.66	10.87	9.63	13.61	9.819	10.3	14.68	11.06	9.29	12.94	9.515
ADB-R	11.8	15.45	13.55	11.1	14.15	12.72	11.8	15.26	13.23	11.1	14.2	13.59
GB-R	7.97	10.61	9.581	7.61	10.07	8.485	7.74	10.52	9.587	7.35	9.849	8.703

From Tables 4 and 5, the result has shown that RB-FCL resulting from hand x-ray region segmentation can make regressor models to have smaller errors compared to standard deep learning models that use hand x-ray images as a whole. The division of region-based hand x-ray into five parts, namely 1-radius-ulna, 2-carpal, 3-metacarpal-phalanges, 4-phalanges, and 5-epiphysis can make deep learning models to be able to learn only for specific regions. Hence, there is not much general information that models must learn. The deep learning model only studies training data for each region, not the whole. To get global feature information, we combine FCL from each region. Specific information from the highlight region is combined to produce a more representative feature for hand x-ray images.

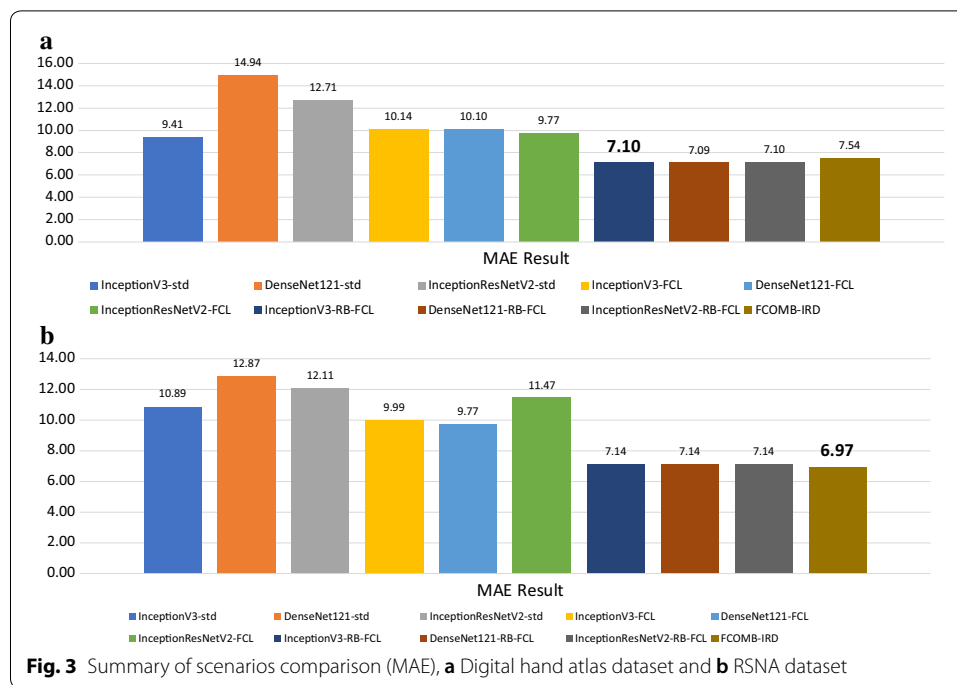
We can compare the overall performance of our proposed RB-FCL method in Tables 4 and 5 compared to Tables 2 and 3. We can see that the overall error performance of the metrics gives a lower error for RB-FCL in Tables 4 and 5 compared to single FCL in Tables 2 and 3. Also, in Tables 4 and 5, the variation of MAE is between 7.10 and 11.34, while in Tables 2 and 3, the variation of MAE is between 9.77 and 17.9. We can see that RB-FCL gives a smaller variation of error compared to single FCL.

Tables 6 and 7 show the results of the combination of RB-FCL with the combined FCL layers of InceptionV3, InceptionResNetV2, and DenseNet121. We combined the label IRD on the Digital Hand Atlas dataset and the RSNA dataset. The FCLOMB label is a combined representation of all feature output results from InceptionV3-RB-FCL, DenseNet121-RB-FCL, and InceptionResNetV2-RB-FCL. The best MAE results from the test scenario are produced by the FCLOMB + IRD scenario with MAE values of 6.97, RMSE of 9346, and MAPE 8128.

## Discussions

In general, the results of all tests can be seen in summary in Fig. 3. Based on the results of all tests, the best metric error results obtained by testing the FCLOMB + IRD feature layer scenario with an MAE value of 6.97. These results are obtained based on a combination of several output layer features from each region. Region segmentation creates a deep learning model to produce models that can specifically study the characteristics of each region used to measure the age of bones. The division of regions based on segmentation of 1-radius-ulna, 2-carpal, 3-metacarpal-phalanges, 4-phalanges, and 5-epiphysis is derived from references used by radiologists to determine the age of bones. Obtaining a specific model for each region can produce representative connected layer output features for each region. The acquisition of features that are more representative makes the regressor model can predict bone age better. This is indicated by the decrease in error value in scenario III (RB-FCL) and scenario 4 (FCLOMB-IRD) when compared to scenario 1 (STD) and scenario 2 (FCL). In scenarios 1 and 2, no region segmentation was performed during the deep learning model training, whereas in scenarios III and IV, segmentation was carried out on the bone age determining region.

Based on the literature. Liu et al. use the same hand, atlas dataset [21]. However, they only use the female data from 2 to 15 years old and male from 2.5 to 17 years old. This dataset also consists of x-ray data from 0 to 2.5 years old and also above 17 years old. In our research, we use all of the data provided by the public dataset. Son et al. used its private dataset. Also, the reproducible code is not available. The other researcher uses



the landmark-based multi-region ensemble CNN for bone age assessment [38]. This work differs from our work in terms of the concatenation of layer, the evaluation, and the proportion of data. We combine the FCL of some regions; however, their work directly using input image and segmented to a few regions. The evaluation of this work only uses each of the regions as a comparison. In our research, we evaluate the whole segmented regions that produce fully connected layers. In terms of the dataset, they evaluate the bone age dataset from digital hand atlas with a proportion of 90% training and 10% testing. However, in our work, we evaluate with two public datasets, digital hand atlas dataset (1392 x-ray images) [38] and RSNA dataset (12,814 x-ray images) [39]. The evaluation proportion of our work is 80% training and 20% testing.

State of the art methods for estimating bone age using Digital Hand Atlas dataset is proposed by Giordano et al. and Spampinato et al. [14, 53]. Giordano et al. produced 21.84 months and Spampinato et al. produced MAE 9.48 months. Our proposed method RB-FCL produce MAE value of 7.1 months. The result of the state of the art method to estimate bone age using the RSNA dataset was produced by Castillo et al. [16]. It produces 9.73 months for MAE. While the MAE that we produce for the RB-HCL method is 6.71 months. Based on the RB-FCL evaluation, our method produces an MAE error that has the same value compared to the state of the art method for digital hand atlas data. For RSNA dataset, our approach produces smaller MAE value compared to the state of the art method.

We compare the results of predictions with MAE of 9.6 months [8] for the digital hand atlas dataset. The use of RB-FCL for digital hand atlas datasets has a smaller value of 7.10 months. We compared the RSNA error results of the comparison dataset with other researchers, Castillo et al. Getting the best MAE value of 9.82 months [16]. While our results have a smaller MAE value of 6.97 months. Based on the comparison of these results, the use of the region-based feature layer RB-FCL method can obtain better bone

age prediction values when compared to standard deep learning procedures. In the next research, we will try to make modifications to the convolution method in order to produce more representative output layer features for each region.

## Conclusion

Bone age assessment is one way to estimate the age of a human bone. The use of image procession and deep learning techniques has been widely used to conduct bone age assessment procedures. In this study, we propose the Region-Based Feature Connected Layer output (RB-FCL) segmentation of several deep learning models to be able to predict bone age. Region-based is divided according to regions recommended by radiologists when they do the manual assessment on hand x-ray. The regions are 1-radius-ulna, 2-carpal, 3-metacarpal-phalanges, 4-phalanges, and 5-epiphysis. From the results of testing using the proposed method (RB-FCL), the best error results obtained were 6.97 months for the MAE, RMSE 9346, and MAPE 8128. The results are obtained from the merging of the output layer features for each region. These results are better than the test results using a standard deep learning procedure that has an MAE value of 9.41 months.

## Abbreviations

ANN: Artificial Neural Network; BAA: Bone age assessment; Fast R-CNN: Fast Region-based Convolutional Network; FCL: Feature Connected Layer; GP: Greulich–Pyle; IRD: Combination of single FCL and RB-FCL; MAE: Mean average error; MAPE: Mean average percentage error; RB-FCL: Region-Based Feature Connected Layer; RMSE: Root mean square error; ROI: Region of Interest; TW: Tanner–Whitehouse; The FLOMB: Combined representation of all feature output results from InceptionV3-RB-FCL, DenseNet121-RB-FCL, and InceptionResNetV2-RB-FCL.

## Acknowledgements

We want to express our gratitude for the grant received from Universitas Indonesia. PUTI Q1 Grant No NKB-1279/UN2. RST/HKP05.00/2020.

## Authors' contributions

AW: Propose RB-FCL, coding implementation, create simulation scenarios, and doing simulation measurement for two public datasets. Revise the introduction, methods, add datasets, and revise results & discussions. PM: Verify the experiment process, data compilation, and the consistency of derived formula application. Revise the results, analysis, and discussion sections. Both authors read and approved the final manuscript.

## Funding

Universitas Indonesia (2020).

## Availability of data and materials

<https://github.com/arizbw/rb-fcl>.

## Competing interests

Not applicable.

Received: 16 March 2020 Accepted: 14 August 2020

Published online: 28 August 2020

## References

1. Poznanski AK, Hernandez RJ, Guire KE, Bereza UL, Garn SM. Carpal length in children—a useful measurement in the diagnosis of rheumatoid arthritis and some congenital malformation syndromes. *Radiology*. 1978;129(3):661–8.
2. Bull RK, Edwards PD, Kemp PM, et al. Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. *Arch Dis Childhood*. 1999;81:172–3.
3. White H. Radiography of infants and children. *JAMA*. 1963;185:223.
4. Gilsanz V, Ratib O. Hand bone age: a digital atlas of skeletal maturity. Berlin: Springer; 2005. <https://books.google.co.id/books?id=SRwFrRszloC>. Accessed 29 Dec 2019.
5. Satoh M. Bone age: assessment methods and clinical applications. *Clin Pediatr Endocrinol*. 2015;24(4):143–52.
6. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Choy G, Do S. Fully automated deep learning system for bone age assessment. *J Digit Imaging*. 2017;30(4):427–41.



7. Mughal AM, Hassan N, Ahmed A. Bone age assessment methods: a critical review. *Pak J Med Sci*. 2014;30(1):211–5. <https://doi.org/10.12669/pjms.301.4295>.
8. Gabryel M, Dama'sević R. The image classification with different types of image features. In: Rutkowski L, Korytkowski M, Scherer R, Tadeusiewicz R, Zadeh LA, Zurada JM, editors. *Artificial intelligence and soft computing*. Cham: Springer International Publishing; 2017. p. 497–506.
9. Davis LM, Theobald BJ, Bagnall A. Automated bone age assessment using feature extraction. In: Yin H, Costa JAF, Barreto G, editors. *Intelligent data engineering and automated learning—IDEAL 2012*. Berlin: Springer; 2012. p. 43–51.
10. Zhang A, Gertych A, Liu BJ, Huang HK. Bone age assessment for young children from newborn to 7-year-old using carpal bones. vol. 6516; 2007, pp. 6516–6516. <https://doi.org/10.1117/12.709710>.
11. Somkantha K, Theera-Umporn N, Auephanwiriyakul S. Bone age assessment in young children using automatic carpal bone feature extraction and support vector regression. *J Digit Imaging*. 2011;24:1044–58.
12. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. *Am J Med Sci*. 1959;238(3):393.
13. Goldstein H, Tanner JM, Healy M, Cameron N. Assessment of skeletal maturity and prediction of adult height (TW3 method). 3rd ed. London: Saunders; 2001.
14. 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. <https://doi.org/10.1016/j.media.2016.10.010>.
15. Gertych A, Zhang A, Sayre J, Pospiech-Kurkowska S, Huang HK. Bone age assessment of children using a digital hand atlas. *Comput Med Imaging Graph*. 2007;31:322–31.
16. Castillo JC, et al. RSNA bone-age detection using transfer learning and attention mapping; 2017. [http://noiselab.ucsd.edu/ECE228\\_2018/Reports/Report6.pdf](http://noiselab.ucsd.edu/ECE228_2018/Reports/Report6.pdf). Accessed 20 June 2019.
17. Wang S, Shen Y, Shi C, Yin P, Wang Z, Cheung PWH, et al. Skeletal maturity recognition using a fully automated system with convolutional neural networks. *IEEE Access*. 2018;6:29979–92.
18. Wang S, Shen Y, Zeng D, Hu Y. Bone age assessment using convolutional neural networks. In: 2018 international conference on artificial intelligence and big data, ICAIBD 2018; 2018, pp. 175–8.
19. RSNA Dataset, <https://www.rsna.org/en/education/ai-resources-and-training/ai-image-challenge>. Accessed 20 Nov 2019.
20. Son SJ, Song Y, Kim N, Do Y, Kwak N, Lee MS, Lee BD. TW3-based fully automated bone age assessment system using deep neural networks. *IEEE Access*. 2019;7:33346–58. <https://doi.org/10.1109/ACCESS.2019.2903131>.
21. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
22. Liu Y, Zhang C, Cheng J, Chen X, Wang ZJ. A multi-scale data fusion framework for bone age assessment with convolutional neural networks. *Comput Biol Med*. 2019;108(March):161–73. <https://doi.org/10.1016/j.compbiomed.2019.03.015>.
23. Cifuentes-Alcobendas G, Domínguez-Rodrigo M. Deep learning and taphonomy: high accuracy in the classification of cut marks made on fleshed and defleshed bones using convolutional neural networks. *Sci Rep*. 2019;9(1):18933. <https://doi.org/10.1038/s41598-019-55439-6>.
24. Cunha P, Moura DC, Guevara López MA, Guerra C, Pinto D, Ramos I. Impact of ensemble learning in the assessment of skeletal maturity. *J Med Syst*. 2014;38(9):87. <https://doi.org/10.1007/s10916-014-0087-0>.
25. O'Connor JE, Coyle J, Bogue C, Spence LD, Last J. Age prediction formulae from radiographic assessment of skeletal maturation at the knee in an Irish population. *Forensic Sci Int*. 2014;234(188):e1–8.
26. Davies C, Hackman L, Black S. The persistence of epiphyseal scars in the distal radius in adult individuals. *Int J Legal Med*. 2016;130(1):199–206. <https://doi.org/10.1007/s00414-015-1192-4>.
27. Urschler M, Grassegger S, Stern D. What automated age estimation of hand and wrist MRI data tells us about skeletal maturation in male adolescents. *Ann Hum Biol*. 2015;42(4):358–67. <https://doi.org/10.3109/03014460.2015.1043945>.
28. Harmsen M, Fischer B, Schramm H, Seidl T, Deserno TM. Support vector machine classification based on correlation prototypes applied to bone age assessment. *IEEE J Biomed Health Inform*. 2013;17(1):190–7. <https://doi.org/10.1109/TITB.2012.2228211>.
29. Haak D, Yu J, Simon H, Schramm H, Seidl T, Deserno TM. Bone age assessment using support vector regression with smart class mapping. In: Novak CL, Aylward S, editors. *Lake Buena Vista (Orlando Area), Florida, USA; 2013*. p. 86700A.
30. Kashif M, Deserno TM, Haak D, Jonas S. Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment. *Comput Biol Med*. 2016;1(68):67–75. <https://doi.org/10.1016/j.compbiomed.2015.11.006>.
31. Wang L, Xie X, Bian G, Hou Z, Cheng X, Prasong P. Guidewire detection using region proposal network for x-ray image-guided navigation. In: 2017 international joint conference on neural networks (IJCNN), Anchorage, AK, 2017, pp. 3169–75.
32. Tang FH, Chan JLC, Chan BKL. Accurate age determination for adolescents using magnetic resonance imaging of the hand and wrist with an artificial neural network-based approach. *J Digit Imaging*. 2018;32:283–9.
33. Liu J, Qi J, Liu Z, Ning Q, Luo X. Automatic bone age assessment based on intelligent algorithms and comparison with TW3 method. *Comput Med Imaging Graph*. 2008;32(8):678–84. <https://doi.org/10.1016/j.compmedimag.2008.08.005>.
34. Lin H-H, Shu S-G, Lin Y-H, Yu S-S. Bone age cluster assessment and feature clustering analysis based on phalangeal image rough segmentation. *Pattern Recognit*. 2012;45(1):322–32.
35. Zhao C, Han J, Jia Y, Fan L, Gou F. Versatile framework for medical image processing and analysis with application to automatic bone age assessment. *J Electr Comput Eng*. 2018;2018:13.
36. Iglovikov VI, Rakhlin A, Kalinin AA, Shvets AA. Paediatric bone age assessment using deep convolutional neural networks. In: Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, et al., editors. *Deep learning in medical image analysis and multimodal learning for clinical decision support (Lecture notes in computer science)*. Berlin: Springer International Publishing; 2018. p. 300–8.
37. Wang X, Peng Y, Lu L, Lu Z, Summers RM. Tienet: text-image embedding network for common thorax disease classification and reporting in chest X-rays. *arXiv preprint*; 2018. [arXiv:1801.04334](https://arxiv.org/abs/1801.04334).

38. Shaomeng C, et al. Landmark-based multi-region ensemble convolutional neural networks for bone age assessment. *Int J Imaging Syst Technol*. 2019;29(4):457–64.
39. Digital Hand Atlas Database System. <https://ipilab.usc.edu/research/baaweb/>. Accessed 15 Dec 2019.
40. Dallora AL, et al. Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS ONE*. 2019;14(7):e0220242.
41. Bui TD, Lee JJ, Shin J. Incorporated region detection and classification using deep convolutional networks for bone age assessment. *Artif Intell Med*. 2019;97:1–8.
42. Larson DB, et al. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology*. 2018;287(1):313–22.
43. Pan X, et al. Fully automated bone age assessment on large-scale hand x-ray dataset. *Int J Biomed Imaging*. 2020;2020:12.
44. Chen X, et al. Automatic feature extraction in X-ray image based on deep learning approach for determination of bone age. *Future Gen Comput Syst*. 2020;110:795–801.
45. Shaoqing R et al. Faster R-CNN: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst*; 2015.
46. Wan Shaohua, Goudos Sotirios. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput Netw*. 2020;168:107036.
47. Wibisono A et al. Deep learning and classic machine learning approach for automatic bone age assessment. In: 2019 4th Asia-Pacific conference on intelligent robot systems (ACIRS), Nagoya, Japan; 2019, pp. 235–40.
48. Saputri MS, Wibisono A, Mursanto P, Rachmad J. Comparative analysis of automated bone age assessment techniques. In: 2019 IEEE international conference on systems, man and cybernetics (SMC), Bari, Italy; 2019, pp. 3567–72.
49. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: The IEEE conference on computer vision and pattern recognition (CVPR), 2017, pp. 4700–8.
50. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: The IEEE conference on computer vision and pattern recognition (CVPR); 2016, pp. 2818–26.
51. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: The IEEE conference on computer vision and pattern recognition (CVPR); 2016, pp. 770.
52. Nazir U, Khurshid N, Bhimra MA, Taj M. Tiny-Inception-ResNet-v2: using deep learning for eliminating bonded labors of brick kilns in South Asia. In: The IEEE conference on computer vision and pattern recognition (CVPR) workshops; 2019, pp. 39–43.
53. Giordano Daniela, Kavasidis Isaak, Spampinato Concetto. Modeling skeletal bone development with hidden Markov models. *Comput Methods Programs Biomed*. 2016;124:138–47.
54. Ren X, Li T, Yang X, Wang S, Ahmad S, Xiang L, et al. Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph. *IEEE J Biomed Health Inform*. 2018;23:2030–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)