

Research article

Skeletal bone age assessments for young children based on regression convolutional neural networks

Pengyi Hao¹, Sharon Chokuwa¹, Xuhang Xie¹, Fuli Wu^{1,3}, Jian Wu^{2,3} and Cong Bai^{1*}

¹ College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou, China

² College of Computer Science and Technology, Zhejiang University, Hangzhou, China

³ Real Doctor AI Research Center, Zhejiang University, Hangzhou, China

* Correspondence: Email:congbai@zjut.edu.cn.

Abstract: Pediatricians and pediatric endocrinologists utilize Bone Age Assessment (BAA) for investigations pertaining to genetic disorders, hormonal complications and abnormalities in the skeletal system maturity of children. Conventional methods dating back to 1950 were often tedious and susceptible to inter-observer variability, and preceding attempts to improve these traditional techniques have inadequately addressed the human expert inter-observer variability so as to significantly refine bone age evaluations. In this paper, an automated and efficient approach with regression convolutional neural network is proposed. This approach automatically exploits the carpal bones as the region of interest (ROI) and performs boundary extraction of carpal bones, then based on the regression convolutional neural network it evaluates the skeletal age from the left hand wrist radiograph of young children. Experiments show that the proposed method achieves an average discrepancy of 2.75 months between clinical and automatic bone age evaluations, and achieves 90.15% accuracy within 6 months from the ground truth for male. Further experimental results with test radiographs assigned an accuracy within 1 year achieved 99.43% accuracy.

Keywords: bone age assessment; carpal bones extraction; regression convolutional neural network

1. Introduction

Bone Age Assessment (BAA) is a clinical investigation performed by pediatricians and pediatric endocrinologists to determine the consistency between a child's skeletal bone age and real age [1]. The discrepancy between the two ages will often trigger the likelihood of genetic disorders, hormonal complications and abnormalities in the skeletal system maturity [2]. For over six decades the conventional methods of attaining BAAs have utilized two versions of standardized main atlases; Greulich and Pyle (G&P) [3] atlas and the Tanner-Whitehouse (TW) atlas [1]. These methods use the left hand

X-ray image for BAA and are completely dependent on a human trained expert, who compares the patient's radiograph to the atlases for reference and then assigns a score to the radiograph. On average a well-trained radiologist may take 7.9 min or 1.4 min to perform BAA on a patient using TW or G&P respectively[4]. These manual methods are time consuming and also susceptible to intra- and inter-observer variability.

In recent years, various automatic techniques have been proposed. Most of the approaches were based on some kind of image feature extraction and classification techniques, for example the method in [5] used random forest regression to give a prediction. In these two years, deep learning has proven itself a powerful method for a wide range of computer vision image tasks [6, 7, 8, 9], leading to a growing interest in medical applications [10, 11]. Recently, several deep learning methods were proposed for bone age assessment. Hyunkwang et al. [12] used a deep learning approach on the whole hand. They initially standardized and preprocessed input radiographs by using their proposed preprocessing engine. Then they fine-tuned on an ImageNet to acquire the BAA. Spampinato et al. [13] also used the whole hand and proposed and tested several deep learning methods to evaluate skeletal bone age. They used pre-trained convolutional neural networks (OverFeat, GoogLeNet and OxfordNet) and a custom Bonet trained from scratch. Ren et al. [14] used a regression convolutional neural network (CNN) on the whole hand to give the bone age result. Initially they used an attention module to process all images and then obtain the coarse/fine attention maps as inputs for regression. Then, the regression CNN followed the supervision of the dynamic attention loss during training to give the bone age. Mutasa et al. [15] used a customized convolutional neural network on the whole hand to give the result of the skeletal maturity. Although these methods elevated the BAA to a new height of automation, they were not sufficient since most of these methods were evaluated upon the selected range of ages like 7 to 18. Medical studies [1] verified the value of carpal bone in determining the bone age of young children from 0 to 6 years old before the carpal bones start to overlap. However, these deep learning methods can not extract the powerful features to distinguish young ages.

Past research work on carpal bone segmentation has been done by Pietka et al. [16] utilizing dynamic thresholding. However, due to the limitation of the algorithm, the carpal ROI was not taken into consideration in the bone age assessment procedure. Other works like Seok et.al [17], who used the method of Singular Value Decomposition (SVD) fully connected Neural Network utilised fixed-sized feature vectors from SIFT descriptions with SVD, other papers used carpal bone area extraction and calculated the bone age using support vector regression [18]. Somkantha et al. [19] used carpal bone area extraction with edge following technique for carpal bones boundary extraction and finally used five features to calculate the bone age by a support vector regression. Hsieh et al. [20] extracted the morphology of carpal bones and applied the fuzzy theory with principle component analysis to get the bone age. They also used five geometric features of the carpals including the bone area, the area ratio, and the bone contour of the carpals. Zhang et al. [21] used segmentation and morphological feature analysis on the carpal bones then used Fuzzy classification to assess the carpal bones maturity. Giordano et al. [22] used two regions of interest, carpal bones and the second fingers bones (epiphysis/metaphysis). They used segmentation techniques and bone extraction was achieved by integrating anatomical knowledge of the hand and trigonometric concepts. Their method was based on the clinical method of TW2. Hao et al. [23] also mentioned carpal bones, and used carpal bones as ROIs for classification, but since the detection of such ROIs was not robust, the accuracies of bone age assessment were not acceptable.

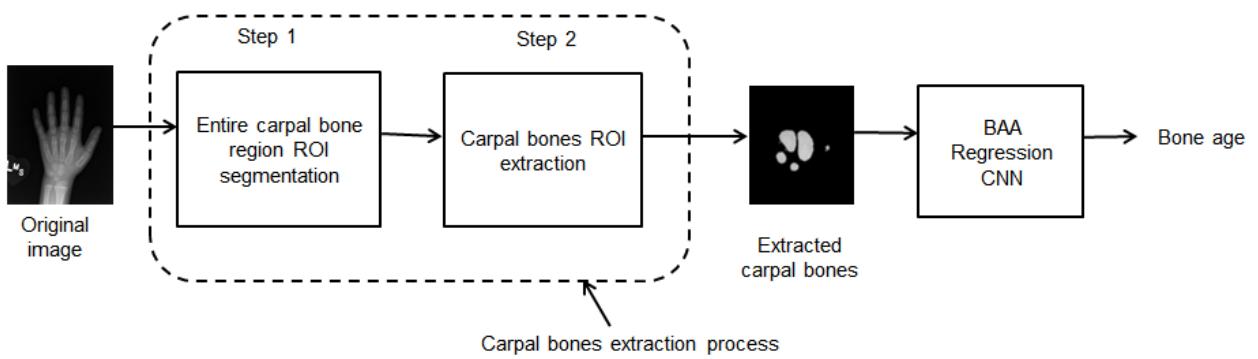


Figure 1. The workflow of our proposed method.

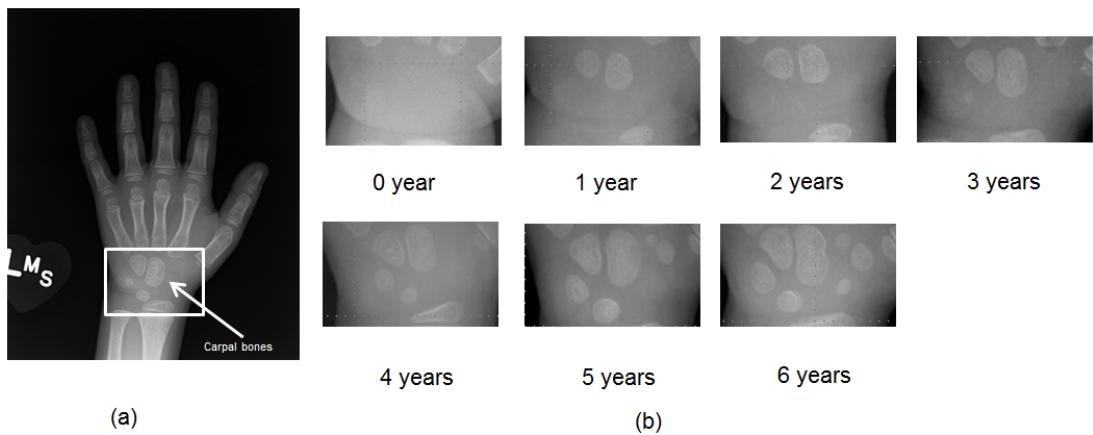


Figure 2. (a) X-ray hand radiograph. The enclosed carpal bones shows the ROI used in this work. (b) The illustration shows the general change of carpals with skeletal maturity from 0 years to 6 years.

Therefore, in order to improve the accuracy of bone age assessment for children, in this paper we present an automated deep learning approach as illustrated in Figure 1. The ultimate goal of this work is to extract boundaries of carpal bones and apply the results to bone age assessment in young children. First we segment the ROI from the entire left hand wrist radiograph using the proposed image preprocessing techniques, then we extract the carpal bones from the soft tissue using boundary extraction, finally we exploit a regression CNN for predicting the bone age of the patient. The advantages of our proposed method include full automation over manual techniques, the reduction in the time cost of assessment of bone age, and the utilization of carpal bones eliminates redundant features when the whole hand is used, since children's carpal bones mature progressively in discrete phases, as commonly known in pediatrics as illustrated in Figure 2.

2. Proposed method

In this section, the details of the proposed BAA method will be given. We put forward a regression CNN to complete the bone age assessment. Before that, the original hand radiograph is initially sub-

jected for carpal bone ROI segmentation that selects only the carpal bones region, the carpal bones are then extracted through boundary extraction techniques, and the output consists of refined carpal bones without tissue. The extracted carpal bones are used as network input for the BAA regression CNN which extracts feature maps and estimates the bone age of the input radiograph.

2.1. *Carpal bones extraction methodology*

The carpal bones are extracted from the original images using four major steps as illustrated on Figure 3. Detailed explanations of each step are as follows.

2.1.1. Entire carpal bone region ROI segmentation

The region of interest is selected by first cropping the lower half of the image as this portion comprises the carpal bones. It is the first step in Figure 3. The selection of the region of interest using vertical and horizontal projection builds on the work by Krit et al. [19], which is the second step in Figure 3. It involves performing projections on both horizontal and vertical axes, so as to set the lower and upper limit of the ROI. The columns with vertical projection values above 0.8 are considered as part of the region of interest. ROI rows are obtained by performing horizontal projections to set the lower and upper limits of the ROI rows. In this case the lower bound is the row with the minimum value of projection and the upper bound is the first row on the left of the lower bound with the projection value of 0.4. The values 0.8 and 0.4 for vertical and horizontal projection bounds respectively are initially tested and used as prior knowledge according to [19].

2.1.2. Image smoothing

In X-ray radiographs, the contrast between the soft tissue and bony structure is often very poor. Additionally, the radiographs vary extensively in intensity and appearance, this can hinder the distinction of the carpal bones from the irrelevant background. Image smoothing is applied to reduce the noise in homogeneous areas of carpal ROI by manipulating a median filter over the images. Median filter is a non-linear filter, which is able to preserve the edges and contrast associated with bony structures, since this filter replaces the pixel values with the median value available in the local neighborhood.

2.1.3. Canny edge detection

Canny edge detection is employed on the smoothed images to detect the carpal bones edges. A Canny edge detector is applied as it is well-known for dramatically reducing the amount of data to be processed i.e. the tissue around the carpal bones. Since Canny edge detection is well-known and widely available in literature, its algorithm will not be explained here. An example of the result obtained from Canny is presented on the third step in Figure 3.

2.1.4. Objects refinement

The result from the Canny edge detector still contains some nontrivial objects. This comprises of parts of the metacarpals, radius and ulna as well as other noised edges. For a more effective BAA, these extra noises need to be removed for proper carpal bones identification. The objects refinement builds on the work presented by Zhang et al [21] who proposed using knowledge-based morphological

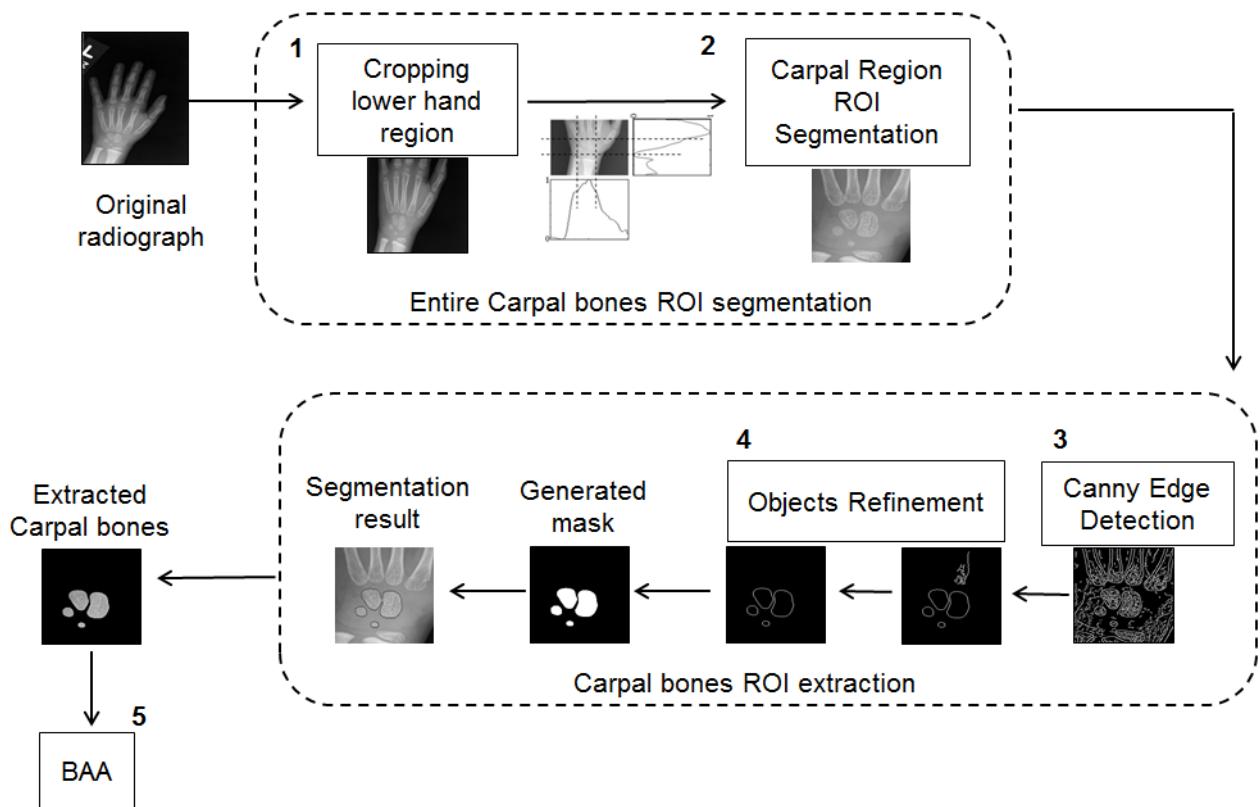


Figure 3. Illustration of the extraction of carpal bones from the entire hand. The original hand is initially subjected for segmentation of the entire carpal bones region and then the extraction of carpal bones, finally the output is fed into the BAA Regression CNN.

operations for cleaning up the extracted bones. Based on this methodology, all the noises that touched the border of the ROI image are removed at first, which includes all the non-carpal bones, some noised lines and other artifacts. Secondly, knowledge-based morphological operations (i.e. eccentricity and solidity) are used to eliminate the remaining noises. Eccentricity, the ratio of the distance between the foci (F1 and F2) to the major axis, is used to measure how far the object deviates from a circle. According to [21], the eccentricity of carpal bone falls between 0.1 and 0.9 and can be used as a prior knowledge. Therefore, objects which have eccentricities below 0.1 or above 0.9 are eliminated in this paper. Thirdly, objects are flood filled and solidity is used to select only closed-contour objects. The first smallest convex polygon that fitted the object is then determined as the convex hull for that object. An object's solidity is then taken as the quantity of pixels in its convex hull. Objects with solidities below 0.5 are eliminated whereas objects, with solidities above 0.5 are taken as the closed-contour objects i.e. the carpal bones. The illustration of objects refinement's result is presented on the fourth step in Figure 3.

2.2. Image preprocessing pipeline

For comparison of our proposed work with earlier research which utilize the whole hand, the entire hand radiograph is also fed as input into the BAA Regression CNN. Prior to regression these

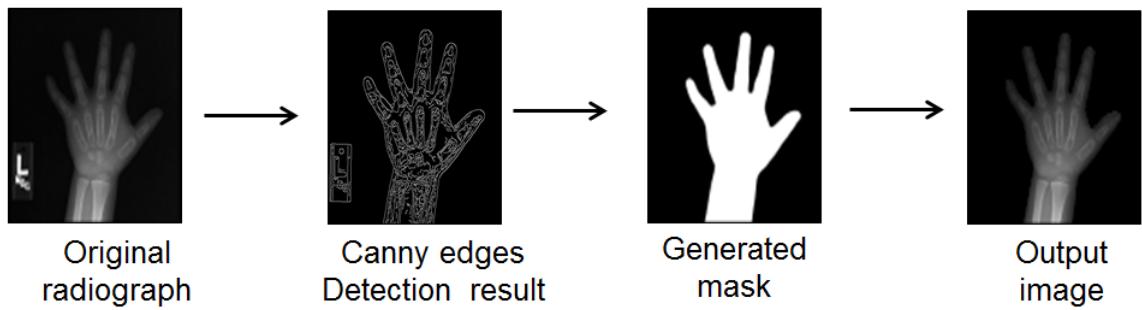


Figure 4. Image preprocessing pipeline workflow for the entire hand region.

images are subject to preprocessing (identifying and extracting the entire hand) as shown on Figure 4. Original radiographs have non-uniform hospital annotations and varying orientations, intensity, and contrast. These variations greatly hinder deep learning algorithm from learning important features, therefore standardizing the images is vital. BAA primarily focuses only on the bony area hence an image preprocessing pipeline is implemented to clean up the images for BAA.

2.3. BAA Regression CNN

We propose a BAA regression CNN with a structure illustrated in Figure 5. All the three Conv Blocks are similar, with a convolutional 2D layer and ReLU activation layer, Batch Normalization, MaxPooling 2D layer and Dropout layer. The Fully-connected block includes a Flatten layer, one Dense layer with ReLU activation and Batch Normalization. Then through a Dropout layer and a sigmoid Dense layer which gives a regression output, we can get the prediction for the input. The convolution kernels are 32, 64, 128 respectively, and the Dense layer has 1024 filters. A dropout value of 0.5 was used across all layers for enhanced generalization. We implemented these networks under the open source deep learning library Keras. The model was trained for 200 epochs using Mean Squared Error loss function (MSE) with Adam optimizer and a learning rate of 0.001, the decay was learning rate / number of epochs. The image input was resized to 96 x 96 spatial domain, so as to reduce computational costs.

2.3.1. Data augmentation

Deep learning techniques require vast amounts of data to yield high accuracy results. However, medical images are hard to obtain due to patients' privacy constraints and image annotation requires an onerous and time-consuming effort of highly trained human experts. Therefore in order to increase the dataset for improved performance various augmentations were implemented including rotations, altering the brightness and shearing.

3. Experiments

3.1. Dataset and evaluations

In this section, we present the various experiments carried out, the results obtained and finally the discussion of the results. The Children's Hospital, Zhejing University School of Medicine of China,

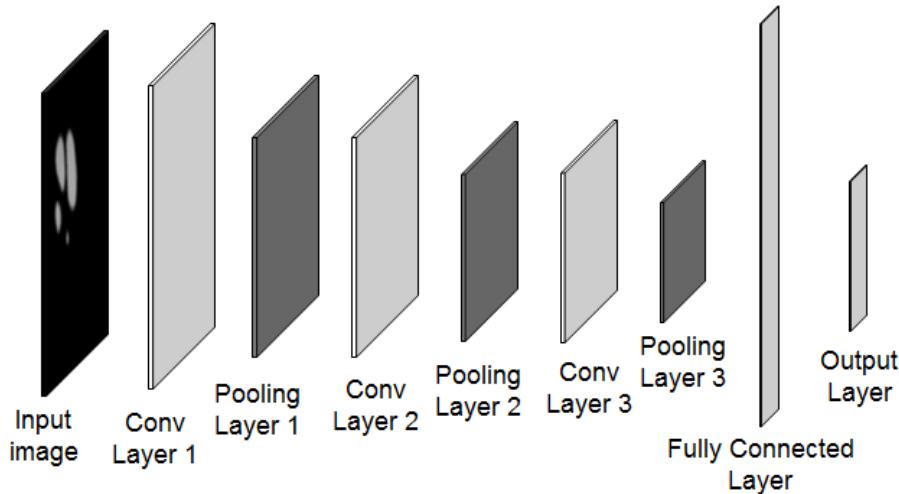


Figure 5. BAA Regression CNN architecture.

provided us with 432 radiographs with chronological age of 0 to 6 years. The dataset comprises of 156 male radiographs and 276 female radiographs. Deep learning methods require vast amounts of data for training, hence we first sort a cross validation method that would effectively make use of our dataset. For evaluations, we compared two different cross validation techniques so as to measure their performances on our dataset.

Validation set approach The validation set approach involves randomly dividing the available set of observations into two parts, a training set and a validation set. The model is fit on the training set, and the fitted model is used to predict the responses for the observations in the validation set. The dataset is split as training and test sets with a proportion 0.8 and 0.2 respectively. All the classes are represented in the test set according to the same proportion they appeared in the training set.

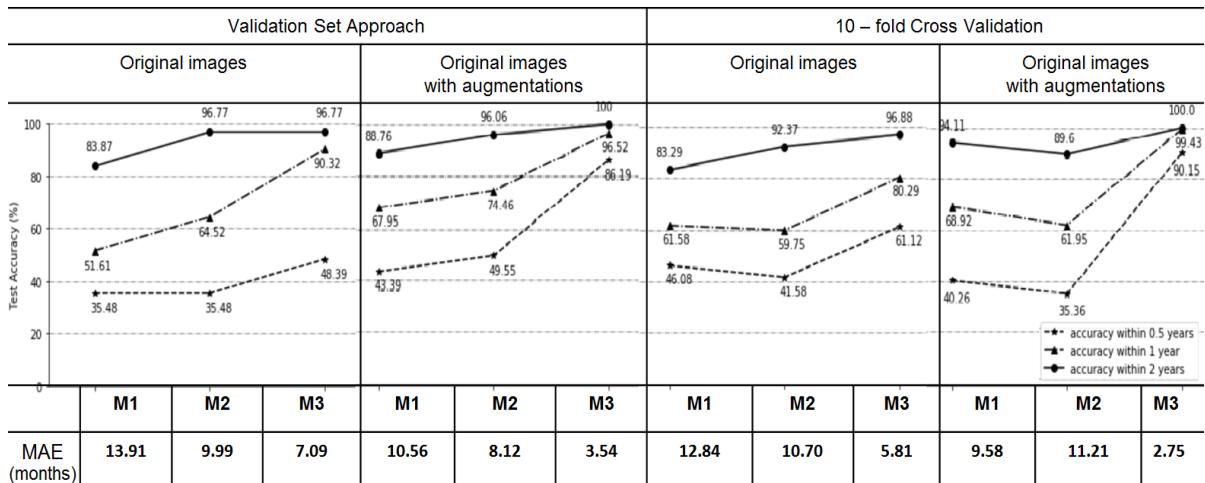
K-fold cross-validation This approach involves randomly dividing the set of observations into K groups (where K was chosen as 10 in our experiments), or folds, of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining folds [24]. The accuracy is calculated as the average of the K folds.

The results obtained from these two cross validation methods are illustrated in Figure 6. Our evaluations also compare the use of the different radiograph regions for effectively performing BAA. Samples of radiograph regions used as input are shown in Figure 7. The methods that use different regions for BAA are: Method 1 (M1) uses the whole hand for BAA; Method 2 (M2) uses the entire carpal region for BAA; Method 3 (M3) uses the extracted carpal bones for BAA.

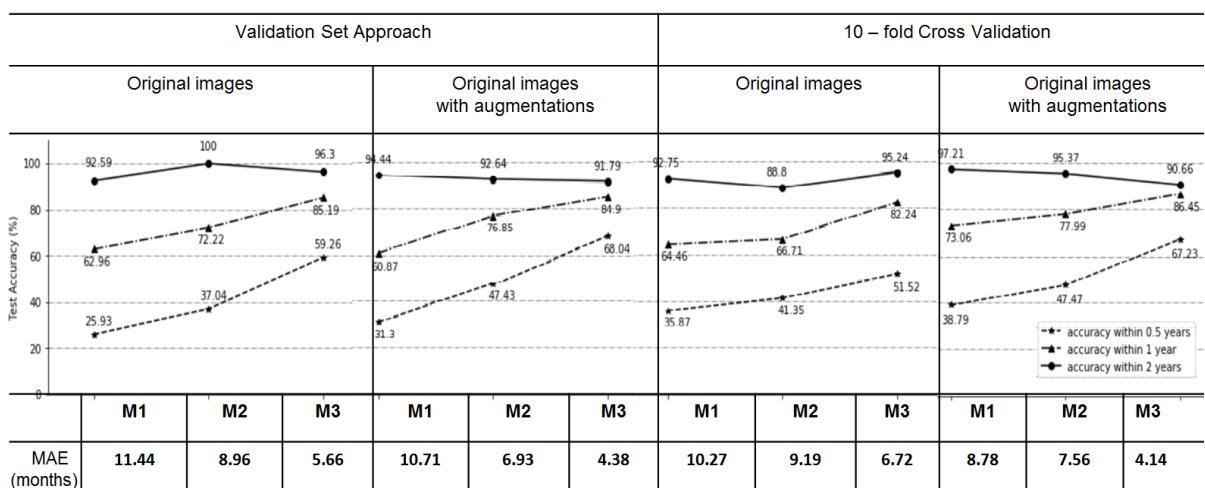
Accuracies are assessed with 4 measures; first, if the difference between prediction and the ground truth is smaller or equal to 0.5 years, then it is accurate within 0.5 years. Accuracy within 1 and 2 years are cases where the network's prediction is within 1 and 2 years from the ground truth, respectively. In addition, mean absolute error (MAE) between the predictions and the ground truths is also used. The smaller MAE means the predictions are much closer with the ground truth.

3.2. Accuracy on the male radiographs

Validation Set approach: The results obtained on the male hand test radiographs are reported on Figure 6 (a). Upon using the validation approach on original test radiographs with augmentations for extracted carpal bones (M3), the proposed method achieves 86.19%, 96.52%, 100% accuracies within 0.5 years, 1 year and 2 years respectively. Original test radiographs with augmentations for carpal region (M2), accuracies decrease to 49.55%, 74.46% and 96.06% within 0.5 years, 1 year and 2 years respectively. While original test radiographs with augmentations for the whole hand (M1), accuracies further decrease to 43.39%, 67.95% and 88.76% within 0.5 years, 1 year and 2 years respectively.



(a) Male



(b) Female

Figure 6. This illustration shows the performance of the three primary methods of bone age assessments (M1, M2 and M3) for (a) male and (b) female. M1 trains a CNN with whole hand radiographs. M2 contains images with only the carpal bones region and M3 contains our proposed work which uses extracted carpal bones.

Whereas original test radiographs without augmentations for extracted carpal bones (M3) achieve

48.39%, 90.32% and 96.77% accuracies within 0.5 years, 1 year and 2 years respectively. Consequently, original test radiographs without augmentations for carpal region (M2), accuracies decrease to 35.48%, 64.52% and 96.77% within 0.5 years, 1 year and 2 years respectively. And original test radiographs without augmentations for the whole hand (M1), accuracies further decrease to 35.48%, 51.61% and 83.87% within 0.5 years, 1 year and 2 years respectively.

K-fold cross validation approach: Three accuracy measures are also evaluated for the k-fold approach i.e. accuracy within 0.5 years, 1 year and 2 years from the ground truth. The results obtained on the male hand test radiographs are reported on Figure 6. Upon using the K-fold cross validation approach on original test radiographs with augmentations for extracted carpal bones (M3), the proposed method achieves $90.15 \pm 1.29\%$, $99.43 \pm 0.04\%$ and 100% accuracies within 0.5 years, 1 year and 2 years respectively. Original test radiographs with augmentations for carpal region (M2), accuracies decrease to $35.36 \pm 4.52\%$, $61.95 \pm 4.02\%$ and $89.60 \pm 2.79\%$ within 0.5 years, 1 year and 2 years respectively. While original test radiographs with augmentations for the whole hand (M1), accuracies slightly increase to $40.26 \pm 4.78\%$, $68.92 \pm 3.27\%$ and $94.11 \pm 0.84\%$ within 0.5 years, 1 year and 2 years respectively.

Whereas original test radiographs without augmentations for extracted carpal bones (M3) achieve $61.12 \pm 12.97\%$, $80.29 \pm 12.51\%$ and $96.88 \pm 0.64\%$ accuracies within 0.5 years, 1 year and 2 years respectively. Consequently, original test radiographs without augmentations for carpal region (M2), accuracies decrease to $41.58 \pm 0.87\%$, $59.75 \pm 1.08\%$ and $92.37 \pm 0.62\%$ within 0.5 years, 1 year and 2 years respectively. And original test radiographs without augmentations for the whole hand (M1), accuracies slightly increase to $46.08 \pm 1.14\%$, $61.58 \pm 1.03\%$ and $83.29 \pm 0.77\%$ within 0.5 years, 1 year and 2 years respectively.

Most interestingly, all the experiments including validation set approach and k-fold cross validation with or without augmentations, show that using extracted carpal bones achieves lower MAE. M3 obtained the lowest MAE, the mean absolute error between prediction and ground truth is only 2.75 months. Another point which also needs to be mentioned is that augmentations decrease the MAE. In validation set approach, the MAE is decreased from 7.09 months to 3.54 months for M3 without or with augmentations. In k-fold validation, the MAE of M3 with augmentations is decreased by 3.06 months.

3.3. Accuracy on the female radiographs

Validation Set approach: The results obtained on the female hand test radiographs are reported on Figure 6 (b). Upon using the validation approach on original test radiographs with augmentations for extracted carpal bones (M3), the proposed method achieves 68.04%, 84.90%, 91.79% accuracies within 0.5 years, 1 year and 2 years respectively.

Original test radiographs with augmentations for carpal region (M2), accuracies decrease to 47.43%, 76.85% and 92.64% within 0.5 years, 1 year and 2 years respectively. While original test radiographs with augmentations for the whole hand (M1), accuracies further decrease to 31.30%, 60.87% and 94.44% within 0.5 years, 1 year and 2 years respectively.

Whereas original test radiographs without augmentations for extracted carpal bones (M3) achieve 59.26%, 85.19% and 96.30% accuracies within 0.5 years, 1 year and 2 years respectively. Consequently, original test radiographs without augmentations for carpal region (M2), accuracies decrease to 37.04%, 72.22% and 100% within 0.5 years, 1 year and 2 years respectively. And original test

radiographs without augmentations for the whole hand (M1), accuracies further decrease to 25.93%, 62.96% and 92.59% within 0.5 years, 1 year and 2 years respectively.

K-fold cross validation approach: Three accuracy measures are also evaluated for the k-fold approach i.e. accuracy within 0.5 years, 1 year and 2 years from the ground truth. The results obtained on the male hand test radiographs are reported on Figure 6. Upon using the K-fold cross validation approach on original test radiographs with augmentations for extracted carpal bones (M3), the proposed method achieve $67.23 \pm 2.86\%$, $86.45 \pm 2.38\%$ and $90.66 \pm 2.24\%$ accuracies within 0.5 years, 1 year and 2 years respectively.

Original test radiographs with augmentations for carpal region (M2), accuracies decrease to $47.47 \pm 6.04\%$, $77.99 \pm 4.44\%$ and $95.37 \pm 2.02\%$ within 0.5 years, 1 year and 2 years respectively. While original test radiographs with augmentations for the whole hand (M1), accuracies decrease to $38.79 \pm 2.20\%$, $73.06 \pm 1.39\%$ and $97.21 \pm 0.37\%$ within 0.5 years, 1 year and 2 years respectively.

Whereas original test radiographs without augmentations for extracted carpal bones (M3) achieve $51.52 \pm 10.50\%$, $82.24 \pm 8.90\%$ and $95.24 \pm 5.50\%$ accuracies within 0.5 years, 1 year and 2 years respectively. Consequently, original test radiographs without augmentations for carpal region (M2), accuracies decrease to 41.35 ± 6.04 , $66.71 \pm 4.44\%$ and $88.80 \pm 2.02\%$ within 0.5 years, 1 year and 2 years respectively. And original test radiographs without augmentations for the whole hand (M1), accuracies slightly increase to $35.87 \pm 2.20\%$, $64.46 \pm 1.39\%$ and $92.75 \pm 0.37\%$ within 0.5 years, 1 year and 2 years respectively.

Similarly with experiments on male radiographs, using extracted carpal bones obtains lower MAE, and M3 achieves a MAE of 4.14 months, which is much more accurate than using whole hand. Using augmentations gives us a much more lower MAE in comparison to not using augmentations.

4. Discussion

The experiments carried out in this work indicate that the use of extracted carpal bones (M3) in performing BAA by automatically regressing the bone age yields relatively higher accuracy for female and male cohorts respectively in comparison to the use of the entire carpal bone region (M2) - which consists of tissue and bone and the whole hand (M1). Previous work to separate bone and tissue found hand tissue muscle attenuation challenging to segment [5], however, in this paper, the carpal bones were effectively extracted and separated from the tissue hence increasing the BAA assessments. For children's BAA, carpal bones are a vital distinctive feature for skeletal maturity as shown on Figure 2, hence the use of the whole hand will introduce redundant features which are eliminated in this work. Furthermore, the use of the entire hand requires tremendous preprocessing due the variations in the original images as indicated by [12], however, in this work only the carpal ROI is vital hence lesser preprocessing is required. According to radiologists only the size and shape of carpal bones are the characteristics related with skeletal development, bony texture inside the carpal bone is not the factor that radiologist investigate in assessing the bone age [2].

The challenge that can be faced in carpal bone extraction is when some anomalous patients mature more rapidly and their carpal bones start ossification at a much earlier age. In this case, there will be no clear boundary between the metacarpal bones and carpal bones hence making the segmentation of the carpal region challenging. Another thing, when poor instruments are used in obtaining X-ray radiographs leading to very poor contrast images, it becomes difficult for the Canny edge detector to

M1 Using whole hand	M2 Using carpal region	M3 Using extracted carpals	Ground truth in months	Model's prediction			Accurate within 0.5 year			Accurate within 1 year			Accurate within 2 years		
				M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
			42.0	40.87	38.73	41.21	✓	✓	✓	✓	✓	✓	✓	✓	✓
			60.0	64.52	56.34	61.03	✓	✓	✓	✓	✓	✓	✓	✓	✓
			24.0	35.37	27.77	25.99	✗	✗	✗	✗	✗	✗	✗	✗	✗
			36.0	27.49	16.61	38.97	✗	✗	✗	✗	✗	✗	✗	✗	✗

Figure 7. Sample predictions for M1, M2 and M3.

find the boundary between tissue and bone in such images.

The evaluations illustrated on Figure 6 indicate that for limited data, a hold-out set estimator suffers a relatively higher variation in the performance evaluation for different samples of data; since the reduced training set poses a high risk of losing important patterns in the dataset resulting in underfitting. Whereas k-fold validation reduces this variance since it is less sensitive to the separating of the dataset, all observations in the dataset are used both for training and validation, and each observation is used for validation exactly once. The results shown on Figure 6 prove the effectiveness of 10-fold cross validation (original images with augmentations 90.15% for males and 67.23% for females) over a hold-out estimator method (original images with augmentations 86.19% for males and 68.04% for females); nearly all the experiments which utilized 10-fold cross validation outperformed the hold out estimator. The value for K was chosen as 10 which produced a lesser biased and higher variance performance because of the higher similarity and lower diversity between the training sets, although with a trade-off of slower computation. Crafting a more robust model that is less liable to overfitting was achieved by artificially creating training images through image augmentation techniques; this boosted the performance of our BAA regression network.

The accuracy increases relatively with data augmentation and when 10-fold cross validation is used. Further analysis of the results indicate that across all the experiments, our proposed work M3 outperforms the use of the whole hand and the entire carpal bone region. Samples of predictions obtained are shown on Figure 7. Our results show a trend of higher accuracy for male patients' bone age predictions, in contrast to the females' bone age predictions. This is attributed by the fact that girls mature faster than boys as evidenced by the dataset used in our experiments which showed that the girls' carpal bones reach ossification at a much earlier age than boys. Additionally, some radiographs overlapped into a number age groups hence requiring even more exploration to improve the bone age prediction.

5. Conclusion

In this paper, we gave a deep-learning method to automatically segment and extract carpal bones as a region of interest for BAA. This paper made a comparison of the use of the whole hand, the entire carpal bone region which comprises of bones and tissue, and the extracted carpal bones with tissue eliminated, while utilizing a data driven approach with a regression CNN. As evidenced by the results obtained, the technique proposed in this paper has great capability to yield much higher accuracy results. Given a larger training dataset will significantly suppress the BAA inter-rater variation.

Acknowledgments

The work is supported by Zhejiang Provincial Natural Science Foundation of China under grants No. LY18F020034, LY18F020032, and National Natural Science Foundation of China under grants No.61801428, 61502424, and partially supported by the Ministry of Education of China under grant of No. 2017PT18 and the Zhejiang University Education Foundation under grant of No. K18-511120-004 and No. K17-511120-017.

Conflict of interest

The authors declare no conflict of interest.

References

1. R. M. Malina and G. P. Beunen, Assessment of skeletal maturity and prediction of adult height (TW3 method), *Am. J. Human Biol.*, **14**(2002), 788–789.
2. D. R. Kirks and N. T. Griscom, Practical pediatric imaging: Diagnostic radiology of infants and children, *Lippincott-Raven*, (1998).
3. W. W. Greulich and S. I. Pyle, Radiographic atlas of skeletal development of the hand and wrist, *Stanford University Press*, (1959).
4. C. Athanasios, B. Maria, K. George, et al., Bone age estimation and prediction of final height in patients with β -thalassaemia major: a comparison between the two most common methods, *Pediat. Radiol.*, **37**(2007), 1241–1246.
5. T. S. Levitt, M. W. Hedgcock, J. W. Dye, et al., Bayesian inference for model-based segmentation of computed radiographs of the hand, *Artif. Intell. Med.*, **5**(1993), 365–387.
6. J. L. Zhang, P. Liu, F. Zhang, et al., CloudNet: Groundbased cloud classification with deep convolutional neural network, *Geophys. Res. Lett.*, (2018), 8665–8672.
7. D. W. Zhang, D. Y. Meng and J. W. Han, Co-saliency detection via a self-paced multiple-instance learning framework, *TPAMI*, **39**(2017), 865–878.
8. C. Bai, L. Huang, X. Pan, et al., Optimization of deep convolutional neural network for large scale image retrieval, *Neurocomputing*, **303**(2018), 60–67.
9. X. He, H. Zhang, M. Landis, et al., Unsupervised boundary delineation of spinal neural foramina using a multi-feature and adaptive spectral segmentation, *Med. Image Anal.*, **36**(2017), 22–40.

10. C. Xu, L. Xu, Z. Gao, et al., Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture, *Med. Image Anal.*, **50**(2018), 82–94.
11. T. Xiao, L. Liu, K. Li, et al., Comparison of transferred deep neural networks in ultrasonic breast masses discrimination, *BioMed. Res. Int.*, **36**(2018), 1–9.
12. L. Hyunkwang, T. Shahein, S. Giordano, et al., Fully automated deep learning system for bone age assessment, *J. Digit. Imaging*, **30**(2017), 427–441.
13. C. Spampinato, S. Palazzo, D. Giordano, et al., Deep learning for automated skeletal bone age assessment in X-Ray images, *Med. Image Anal.*, **36**(2017), 41–51.
14. X. Ren, T. Li, X. Yang, et al., Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph, *IEEE J. Biomed. Health*, (2018).
15. S. Mutasa, P. D. Chang, C. Ruzal-Shapiro, et al., MABAL: a Novel Deep-Learning Architecture for Machine-Assisted Bone Age Labeling, *J. Digit. Imaging*, **31**(2018), 513–519.
16. E. Pietka, L. Kaabi, M. L. Kuo, et al., Feature extraction in carpal-bone analysis, *IEEE Trans. Med. Imaging*, **12**(1993), 44–49.
17. J. Seok, B. Hyun, J. Kasa-Vubu, et al., Automated classification system for bone age x-ray images, *IEEE SMC*, (2012), 208–213.
18. P. Liskowski and K. Krawiec, Segmenting retinal blood vessels with deep neural networks, *IEEE Trans. Med. Imaging*, **35**(2016), 2369–2380.
19. K. Somkantha, N. Theera-Umpon and S. Auephanwiriyakul, Bone age assessment in young children using automatic carpal bone feature extraction and support vector regression, *J. Digit. Imaging*, **24**(2011), 1044–1058.
20. C. W. Hsieh, A fuzzy-based growth model with principle component analysis selection for carpal bone-age assessment, *Med. Biol. Eng. Comput.*, **48**(2010), 579–588.
21. A. Zhang, A. Gertych and B. J. Liu, Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones, *Comput. Med. Imag. Grap.*, **31**(2007), 299–310.
22. D. Giordano, C. Spampinato and G. Scarciofalo, An automatic system for skeletal bone age measurement by robust processing of carpal and epiphysial/metaphysial Bones, *IEEE T. Instrum. Meas.*, **59**(2010), 2539–2553.
23. P. Y. Hao, Y. J. Chen, S. Chokuwa, et al., Skeletal bone age assessment based on deep convolutional neural networks, *PCM*, (2018), 408–417.
24. S. Yadav and S. Shukla, Analysis of k-Fold cross-validation over hold-out validation on colossal datasets for quality classification, *IACC*, (2016), 78–83.