Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 22(1): 138–151.
DOI: 10.3934/mbe.2025007
Received: 16 August 2024
Revised: 18 October 2024
Accepted: 18 December 2024
Published: 3 January 2025

Research article

A fully automated U-net based ROIs localization and bone age assessment method

Yuzhong Zhao[1], Yihao Wang[1], Haolei Yuan[2], Weiwei Xie[2], Qiaoqiao Ding[1,*] and Xiaoqun Zhang[1,*]

[1] Institute of Natural Sciences, School of Mathematical Sciences, MOE-LSC & Shanghai National Center for Applied Mathematics (SJTU Center), Shanghai Jiao Tong University, Shanghai 200030, China

[2] GenSci, Changchun 130012, China

* **Correspondence:** Email: dingqiaoqiao@sjtu.edu.cn, xqzhang@sjtu.edu.cn.

Abstract: Bone age assessment (BAA) is a widely used clinical practice for the biological development of adolescents. The Tanner Whitehouse (TW) method is a traditionally mainstream method that manually extracts multiple regions of interest (ROIs) related to skeletal maturity to infer bone age. In this paper, we propose a deep learning-based method for fully automatic ROIs localization and BAA. The method consists of two parts: a U-net-based backbone, selected for its strong performance in semantic segmentation, which enables precise and efficient localization without the need for complex pre- or post-processing. This method achieves a localization precision of 99.1% on the public RSNA dataset. Second, an InceptionResNetV2 network is utilized for feature extraction from both the ROIs and the whole image, as it effectively captures both local and global features, making it well-suited for bone age prediction. The BAA neural network combines the advantages of both ROIs-based methods (TW3 method) and global feature-based methods (GP method), providing high interpretability and accuracy. Numerical experiments demonstrate that the method achieves a mean absolute error (MAE) of 0.38 years for males and 0.45 years for females on the public RSNA dataset, and 0.41 years for males and 0.44 years for females on an in-house dataset, validating the accuracy of both localization and prediction.

Keywords: bone age assessment; U-net; ROIs localization; interpretability

1. Introduction

Bone age is a measure of skeletal maturity and a primary indicator of biological development of adolescents. Bone age assessment (BAA) has been widely applied in clinical practice. For example,

the difference between bone age and chronological age can be used to estimate adult height or to evaluate the efficiency of growth. In the procedure of BAA, usually X-ray images of left hands are taken and manually evaluated by clinicians according to some standards.

In the literature, two primary assessment methods are commonly referenced: the GP method established by Greulich et al. [1] and the Tanner Whitehouse (TW) [2] method. The so-called GP method is based on comparing bone features to a series of atlases representing different ages. The closest matching templates are selected and the corresponding bone age is inferred for the target image. This method is obviously subjective and highly depends on the experience of clinicians. With the development of imaging, the representative atlas used in the GP method has become outdated and cannot be easily adapted to different imaging settings and populations. The second method, known as the TW method [2], consists of selecting multiple key bones as regions of interest (ROIs) of hand images and evaluating the maturity level of each region. The bone age is obtained using a scoring formula. The standard of the TW method is generally more objective and interpretable. The TW method was further developed to the TW3 method in [3], a standard in many BAA literature. However, the TW method requires a large amount of workload for clinicians to extract the ROIs and score the maturity levels. The results still heavily depend on the experience and subjective judgements of radiologists.

To reduce the clinical workload and standardize the evaluation process, many computer-assisted methods have been proposed. BoneXpert [4] was the first widely accepted commercial BAA software extracting 13 ROIs proposed in the TW3-RUS method [3]. The system driven by conventional image processing techniques to evaluate the shapes of each bone in the hand images was reported that the system has a high rejection rate when the image quality is low or the bone shapes are far from the standard population [5]. In recent years, deep learning methods are making great success in various tasks of computer vision, medical image analysis, and bio-medicine. Deep learning based BAA methods have been considered in recent years [5–11].

Deep learning-based BAA methods were developed along two lines: whole image-based black-box methods and TW3-RUS ROI-based methods with interpretability. Whole image based deep learning method consists of two phrases: preprocessing and classification. Techniques such as denoising, rotation, and contrast enhancement in preprocessing are applied for the normalization of the input X-ray images. The normalized images are used to train a deep neural network (DNN), such as VGG16 [12] or the inception neural network [13]. For example, in [8], several popular DNN models pretrained on the ImageNet and a customized convolutional neural network (CNN) employed for BAA. In [7], a pretrained model from ImageNet and a large amount of image preprocessing steps were applied to improve the accuracy of the network. A recent report on DNNs for BAA on the asian population in Taiwan can be found in [14]. In general, this type can usually achieve a high accuracy, due to the fact that the whole procedure is a black-box model, the interpretability is missing, which is not acceptable for radiologists and patients in practice. Due to its high interpretability, TW-based DNN methods investigated in recent literature. To mimick the diagnosis procedure of radiologists, this line of work is usually composed of two steps. The ROIs are extracted by an object detection network with a dataset with ROIs coordinates. The ROIs are then directly input into a network for BAA, or used for the regression of skeletal maturity levels , if the ROIs maturity levels labels are available. The process preserves the interpretability of TW methods and the labor of ROIs extraction. Currently, object detection methods such as Faster R-CNN [15] are used for ROI extractions. As this

process can not directly output satisfactory results for X-ray images due to the high similarity of bones, some sophisticated pre-post processing methods are necessary for accurate ROIs extractions [6]. Besides, labeling the skeletal maturity imposes additional burdens for radiologists, for which the labels are not accessible in many datasets.

This paper introduces a novel automatic BAA method that integrates the interpretability of the TW3 ROI-based approach with the global feature extraction of the GP method. We propose using a U-net neural network to simultaneously locate the ROIs in the TW3-RUS method [3], which achieves a high localization accuracy without the need for pre- or post-processing. U-net's ability to incorporate positional information allows precise identification of the phalanges, addressing limitations of conventional object detection methods. By combining both the local ROI-based features and the global features from whole images, our method achieves a high prediction accuracy without relying on the skeletal maturity levels. The method was validated on both a public and a private dataset, which demonstrates a improved prediction accuracy over the existing approaches that solely rely on global or local features while maintaining the model interpretability.

The main contributions of this paper can be summarized as follows:

- **Enhanced U-net Localization with Positional Awareness**: Unlike conventional object detection methods, the U-net network can accurately distinguish between the specified ROIs in the TW3-RUS method due to its ability to incorporate positional information. This enhancement ensures the precise identification of skeletal features, which improves the reliability of our system.
- **Combination of Local and Global Features without Skeletal Maturity Levels**: Our proposed method combines both the ROI-based local features and the global features extracted from whole images to achieve the high prediction accuracy without relying on predefined skeletal maturity levels. This hybrid approach leverages the interpretability of the TW3-based method with the global contextual awareness of the GP method, thus offering a balanced solution that enhances both the accuracy and the interpretability.
- **Extensive Validation on Public and Private Datasets**: The proposed method is validated on the public RSNA challenge dataset [16] and a private in-house dataset. Our method achieved a mean absolute error (MAE) of 0.38 years for males and 0.45 years for females on the public RSNA BAA dataset, and 0.41 years for males and 0.44 years for females on the private dataset. These results outperform the existing methods with only global or local features, which demonstrates the effectiveness and generalizability of our approach.

2. Methods

The fully automated ROIs localization and BAA prediction method is illustrated in Figure 1. In the following, we present the detail of the U-net based ROIs localization and feature extraction blocks for BAA.

2.1. ROIs localization

The idea of the TW3-RUS method [3] is to use a classification scoring algorithm based on the ossification levels of different bones. In the TW3-RUS, thirteens bones, namely the Radius, the Ulna and other short bones (Distal, Middel, Proximal Phalanx and Metacarpal) of the first, third and fifth finger of the left hand are chosen to assess the skeletal maturity level.

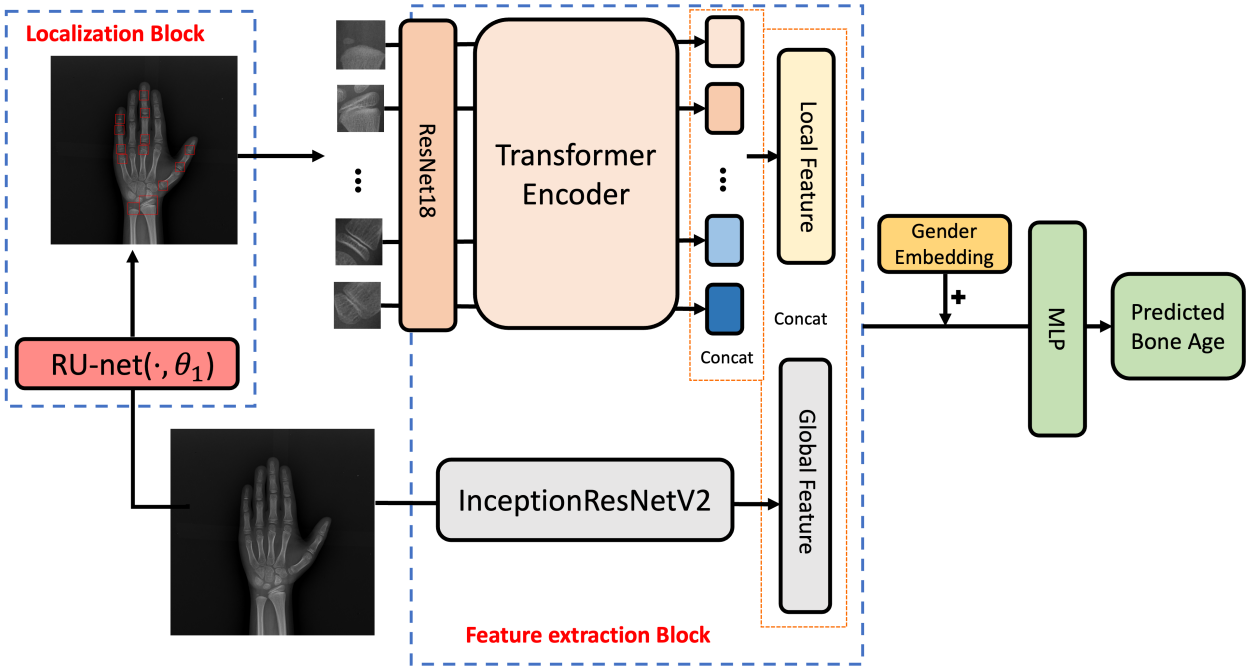

Figure 1. Our automatic bone age assessment model combines both ROIs and global features.

In previous studies, Faster R-CNN has been widely used for object detection in BAA tasks. However, as Son et al. [6] demonstrated, this approach suffers from limitations, including incorrect predictions where multiple ROIs are identified for the same anatomical structure. To address this, Son et al. introduced a two-step process: first locating the bones (bROI), and then localizing the specific ROIs to ensure the accuracy, which add complexity to the workflow. In our approach, we employ U-net [17] as the backbone network for ROIs extraction to overcome these issues. U-net was chosen due to its strong performance in semantic segmentation tasks, which makes it highly effective localizing small and complex anatomical structures. The network's encoder-decoder structure with skip connections preserves both low-level spatial details and high-level semantic information, which is crucial for the accurate ROI extraction. Specifically, our U-net consists of five down-sampling and five up-sampling layers, with each down-sampling layer followed by two 3x3 convolutions, ReLU activations, and 2x2 max-pooling operations. The up-sampling layers use transposed convolutions to upsample the feature maps, and skip connections between the corresponding layers ensure that the fine details are preserved during reconstruction. Given an input X-ray image $\boldsymbol{x}$ and thirteen labeled ROIs center coordinates $(x_i, y_i), i = 1, \cdots, 13$, we first generate fourteen square bounding boxes $\boldsymbol{M}_i, i = 1, \cdots, 14$ (with the thirteen channels corresponding to the thirteen ROIs and the last channel is the residual). With the training dataset pair $(\boldsymbol{x}, \{\boldsymbol{M}\}_{i=1}^{14})$, we apply RU-net [18], which is a variant of U-net that incorporates a total variation regularization at the last softmax layer, to train a network. We use the sum of mean squard error (MSE) and Dice loss [19] between the output mask and the ground truth as our loss function. As the output mask of RU-net is not always strictly square, we recalculate

the centroids of the former thirteen channels and regenerate 13 bounding boxes in order to obtain the ROIs. We note that the size of the bounding box of ROIs is set as 192×192 for the Radius and 96×96 for the other phalanges. The radius was resized to the same size of the other phalanges for the feature extraction. This process can be formulated as

$$[z_1, \cdots, z_{13}] = \mathbf{F}(\text{RUnet}(\boldsymbol{x}, \boldsymbol{\theta}_1)),$$

where $\boldsymbol{x} \in \mathbb{R}^{K \times K}$ is the original image, $\boldsymbol{\theta}_1$ is the parameter set of RUnet, $\mathbf{F}$ is the map of redrawing ROIs $z_1, \cdots, z_{13} \in \mathbb{R}^{k \times k}$ are the corresponding ROI images.

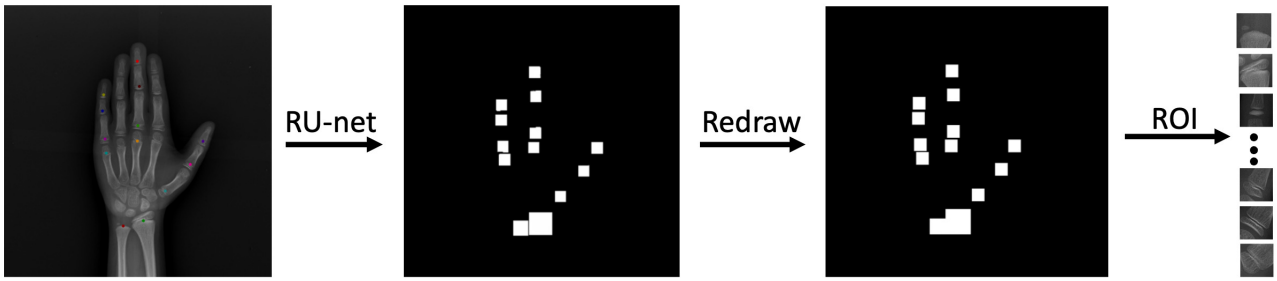

Figure 2. Redraw the output of the RU-net to get a normalized mask.

2.2. BAA prediction

For BAA prediction, the feature extraction block is composed of two parts. One is from the extracted ROIs images, and the other is synthesized from the whole images.

For the ROIs, we first employ the ImageNet pretrained ResNet18 to extract the features of each ROI. After a global average pooling, the third block feature map of each ROI is a vector of size $d_1 = 128$. This network will not be trained anymore. Then, these feature vectors of all 13 ROIs are input to a transformer encoder block. In this phrase, all ROI features will learn the similarity and correlation with each other. Finally, we will get 13 ROI output features. We can regard the transformer block as the imitating the estimating bone mature level phrase. Finally we concatenate all ROI output features together as the local feature with the $13 * d_1$ dimension.

For the global feature extraction, we employ the InceptionResNetV2 architecture due to its ability to efficiently capture both the local and global features through a combination of Inception modules and residual connections. The Inception modules allow the network to extract multi-scale features using parallel convolution layers with different kernel sizes (1×1, 3×3, and 5×5), while the residual connections improve the gradient flow and enable a deeper network training. This is critical where both the local details from ROIs and the global anatomical information are essential for accurate predictions. The global feature vector size is denoted as d_2. Moreover, the final stage is to concatenate features together to a $13 * d_1 + d_2$ dimensional vector, which incorporates both advantages of the GP-method and the TW3-method. This combination ensures that the model captures both the detailed anatomical information from specific ROIs and the broader context from the whole image, which is essential for accurate predictions. Beyond accuracy, this hybrid approach enhances the interpretability, which is a critical factor in clinical applications. By explicitly using ROIs that radiologists commonly rely upon

in manual assessments, the model mimics the clinical decision-making process. This allows clinicians to trace predictions back to specific features, making the results more transparent and understandable.

In order to distinguish the gender, we add the learnable gender embedding to the feature, which is similar to the position embedding in [20]. At last, we simply utilize the multi-layer perceptron (MLP) block with two fully-connected layers to output one scalar bone age. The overall network structure is summarized in Figure 1.

The overall architecture can be formulated as

$$y^{predict} = \mathrm{MLP}([\mathbf{L}(z_1, \cdots, z_{13}), \mathbf{G}(\boldsymbol{x})] + \boldsymbol{E}_{\mathrm{gender}}, \boldsymbol{\theta}_2)$$

where $y_{predict}$ is the predicted age, $\mathbf{L}$ is the local feature extraction block, $\mathbf{G}$ is the global feature extraction block, $\boldsymbol{E}_{\mathrm{gender}} \in \mathbb{R}^{13*d_1+d_2}$ is the gender embedding vector, and $\boldsymbol{\theta}_2$ is the parameter set of the whole regression network.

Finally, we use the mean squared error (MSE) as the loss function to train θ_2:

$$\min_{\boldsymbol{\theta}_2} \frac{1}{N} \sum_{i=1}^{N} \left(y_i^{predict} - y_i^{label} \right)^2,$$

where y_i^{label} is the i-th labeled bone age.

3. Experiments

3.1. Datasets

The proposed method was rigorously tested using two distinct datasets: the public RSNA challenge dataset [16] and a private dataset. This latter dataset is an assemblage of 10265 radiographic images meticulously gathered between March 2015 and October 2020. Notably, these images were predominantly procured from local primary and middle schools, representing a microcosm of the demographic variation within these establishments. A detailed demographic overview of both datasets is systematically presented in Table 1.

In examining the configuration of the RSNA dataset, the training set incorporated 6502 male specimens alongside 5427 female counterparts. Subsequently, the validation set is distinguished by the inclusion of 749 male and 631 female samples. Additionally, a distinct set of 200 samples, irrespective of gender categorization, has been specifically allocated for the testing phase. The RSNA dataset ensures consistent training and testing splits, which allows for fair and reproducible comparisons across studies using this dataset.

Building upon the method delineated in [5], we undertook a meticulous manual annotation process identifying 13 pivotal keypoints on a subset of 481 RSNA images. These images were judiciously selected to ensure a balanced distribution of bone age. 307 samples were designated for training, while the remaining 174 were set aside for testing. As for the proprietary dataset, it contains a total of 1500 samples endowed with annotated coordinates. 1200 samples were allocated for the training phase, and the residual 300 samples were exclusively utilized for testing.

The RSNA dataset, as outlined in [16], was meticulously annotated with respect to bone age by a cohort of six experienced radiologists. For our proprietary dataset, 80% of the total data was earmarked

for training. It translates to 3984 female samples and 4228 male samples. It is crucial to highlight that, the chronological age is utilized as the regression outputs for this private dataset. This decision was underpinned by the controlled nature of the cohort, which is representative of standard developmental patterns. Consequently, in the context of this dataset, the chronological ages can be academically interpreted as the inherent bone ages of the sampled population.

Table 1. Demographic of datasets.

Dataset \ Age	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
RSNA(M)	3	25	60	90	163	180	269	262	294	295	296	297	298	299	300	301	302	303	304
RSNA(F)	4	17	68	109	106	320	435	572	563	72	732	773	609	586	164	199	57	40	/
In-house (M)	/	/	/	/	/	1	355	566	742	875	919	538	422	407	392	69	2	/	/
In-house (F)	/	/	/	/	/	/	386	472	673	808	759	467	444	471	403	97	/	/	/

As the images are not in the same size, we first resized all the training images to 512×512 for the ROIs localization and global feature extraction. Since the contrast of the RSNA images is not uniform, we employed the histogram equalization algorithm CLAHE [21] for the RSNA images. The clip limit is set to 6 and the grid size is set to 8×8.

For the training of the first RU-net ROIs localization network, we performed ±45 degree rotations, random scaled up and down with ratios ranged 0.5 to 0.7, random horizontal flipping for data augmentation on RSNA dataset, in order to adapt to the changes in the size of palms and rotations. We found that this data augmentation can greatly improve the ROIs localization accuracy of this dataset. For our private dataset, the data augmentation process is not necessary since the localization accuracy is already very high with our method.

For the feature extraction network, we performed random scaling between the ratio of 0.8 to 1.1; the flipping, rotations, and translations for data augmentation to compensate different distances, positions and angles. Some examples of augmented images from RSNA can be found in Figure 3.

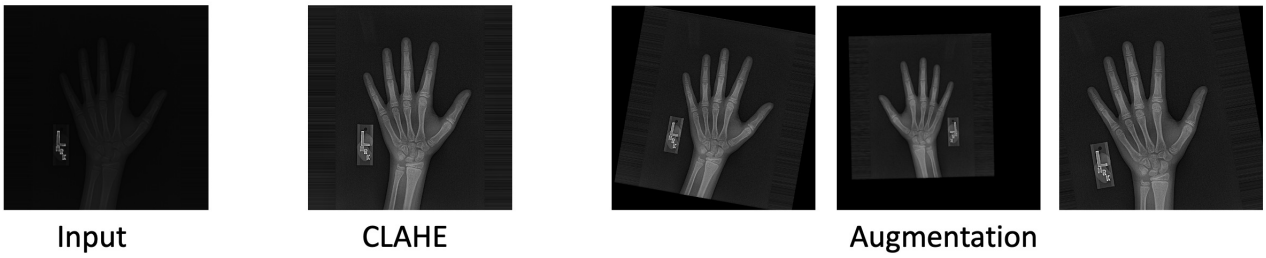

Figure 3. Data augmentation by rotation and scaling.

The weights of Inception-ResNet-v2 for the global feature were obtained by a pretrainning on ImageNet. This choice was based on the recognition that models pretrained on the large dataset ImageNet, often provide a robust starting point for further fine-tuning on specialized tasks. For the local feature extraction blocks and RU-net, the parameters of the CNN were initialized by the Kaiming initialization, which is a default setting in Pytorch.

We chose Adam as the optimizer with the hyperparameter $\beta_1 = 0.9$ and $\beta_2 = 0.999$ in both the localization and regression parts. The initial learning rates were set as 10^{-4}. RU-net was trained with 400 epochs and a batch size of 4 for the regression network, the total epochs number was set as 200 and the batch size was set to 16.

3.2. Results

We conducted a comprehensive validation of the localization accuracy of our proposed RU-net unit. Following the established localization criteria outlined in previous works such as Son et al. [6], Koitka et al. [5], and Everingham et al. [22], we utilized two well-established metrics, namely success rate and precision, to thoroughly evaluate the performance of our localization approach. Both of these metrics are widely recognized in the domain of localization. By utilizing such metrics, we aimed to offer a nuanced and detailed evaluation of our approach in comparison to the existing methodologies in the literature.

The success rate was determined based on stringent conditions, where all 13 ROIs were required to be accurately recognized, and the intersection over union (IoU) between each ground truth ROI and the corresponding predicted mask had to exceed 0.5. On the other hand, the precision rate was evaluated by considering the probability of accurate detection for each bone, accounting for potential discrepancies between the ground truth and predicted results.

In Table 2, we present a comprehensive comparison of the localization accuracy reported in the relevant literature, with a particular emphasis on the studies by Son et al. [6] and Koitka et al. [5]. These comparative evaluations span a range of diverse datasets, which encompass both the aforementioned studies and our methodology. Our U-net based approach demonstrates a commendably high accuracy in localizing the ROIs, as evidenced by the comparative results. This heightened accuracy is further underscored by the comparative metrics delineated in the table.

Table 2. ROIs localization accuracy comparison.

Method	Dataset	Success rate	precision
Son et al. [6]	In-house	98.4	–
Koitkaa et al. [5]	RSNA	–	99.0
Proposed	In-house	98.7	99.3
	RSNA	96.5	99.1

To assess the BAA error, we use the mean absolute error (MAE), which is a standard metric commonly employed in the literature. This metric quantifies the absolute discrepancies between the predicted age and the labelled age, which provides an effective measure of the accuracy and reliability of BAA prediction models. The metric is

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i^{predict} - y_i^{label}\right|.$$

In Table 3, we present the MAE results reported in relevant literature, showcasing findings from different datasets, including In-house, RSNA, or Digital Hand-Atlas [23], and model types, including whole image-based and/or ROI-based approaches. In our study, we report the mean MAE separately

for each gender, as our model was trained on gender-specific data. When comparing our work with other methods, we strive to align with the datasets used in prior studies as closely as possible. For the RSNA dataset, we strictly followed the predefined splits between the training, validation, and testing sets to ensure a fair comparison. However, for other datasets—especially private datasets—reproducing identical experimental setups may not always be feasible due to the limited access to the same data. In such cases, we base our comparisons on the reported metrics and methodologies in the original works.

Our approach, which combines ROIs and whole image information, demonstrates the lowest MAE compared to other similar experimental settings. Notably, our findings are consistent with the results reported in [5], where a similar MAE of 0.38 was achieved using ensemble filtering techniques for the testing results, as stated by the authors.

Our method's superior performance can be attributed to the synergistic utilization of ROIs and whole image information, which enables comprehensive and accurate predictions. By leveraging gender-specific data during the model training process, we are able to capture and account for the potential anatomical differences between genders, which enhances the prediction accuracy. It is worth mentioning that our study's findings align with the current literature, which underscores the importance of incorporating both local and global features in medical image analysis tasks.

Table 3. MAE comparison.

Method	DataSet	Model type	MAE (M/F)
Lee et al. [7]	In-house	Whole image	0.82/0.93 (RMSE)
Son et al. [6]	In-house	ROIs	0.46
Chen et al. [24]	In-house	Whole image	0.46
Iglovikov et al. [10]	RSNA	Whole image	0.51
Lee et al. [25]	RSNA	Whole image	0.53
Koitkaa et al. [5]	RSNA	ROIs	**0.38**
Spampinato et al. [8]	Digital Hand-Atlas [23]	Whole image	0.79
Simukayi et al. [26]	Digital Hand-Atlas	Whole image	0.54
Toan et al. [27]	Digital Hand-Atlas	ROIs	0.59
Zhou et al. [28]	Digital Hand-Atlas	ROIs	0.72
Tong et al. [29]	Digital Hand-Atlas	Whole image	0.55
Proposed	In-house	Whole image	0.45
	In house	Whole image + ROIs	0.41
	RSNA	Whole image	0.44
	RSNA	Whole image + ROIs	**0.38**

3.3. Discussion: Model performance across age groups

The evaluation of the model's performance across different age groups for the RSNA dataset, as illustrated in Figure 4, reveals variations in the MAE on age and gender. These differences highlight the potential shortcomings of the model in certain age brackets.

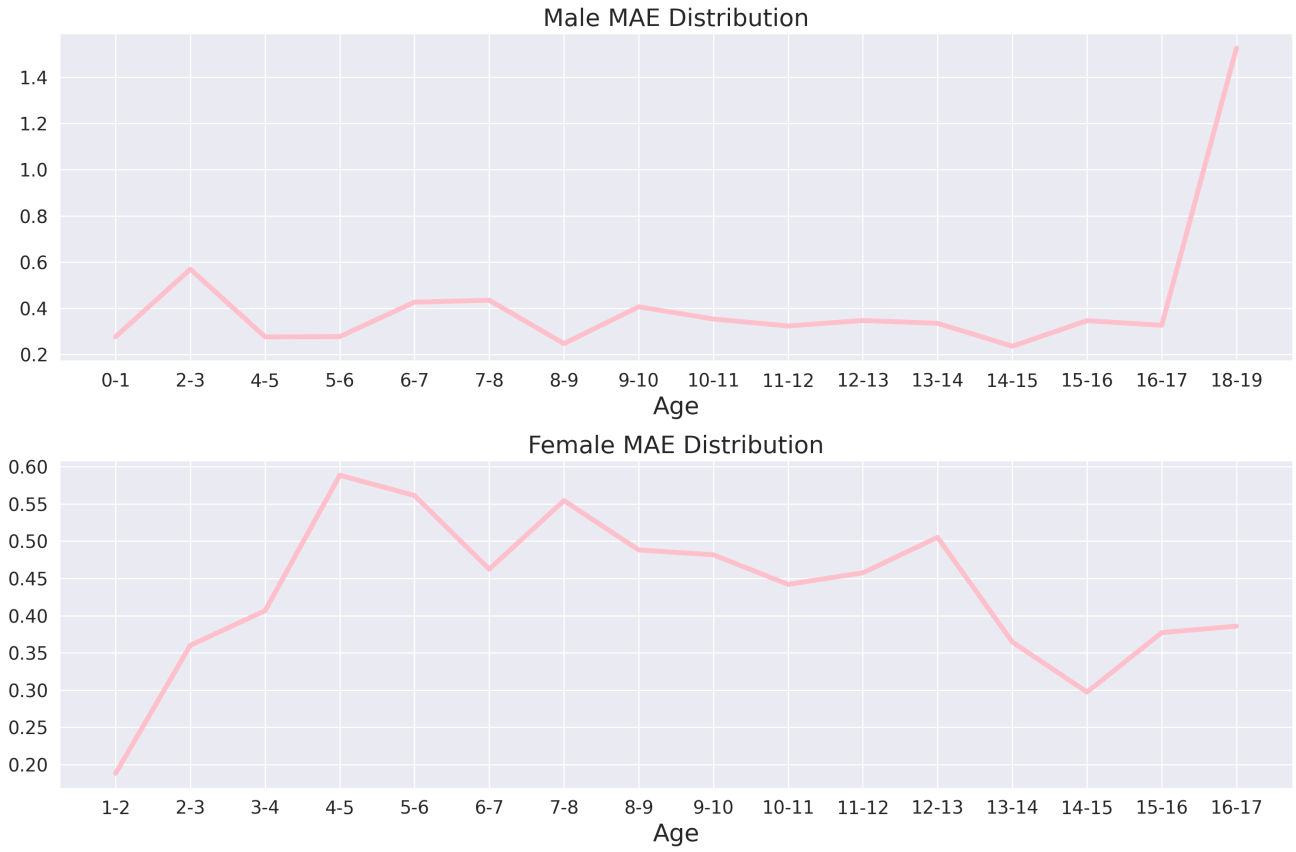

Figure 4. Mean Absolute Error (MAE) distribution across different age groups for male and female.

The MAE distribution for males shows a fairly stable performance for most age groups, with errors generally remaining below 0.4 for the majority of age groups. However, a notable spike occurs in the 18-19 age group, where the error exceeds 1.4, which indicates a significant drop in the predictive accuracy for this particular cohort. This suggests that the model struggles with older male individuals, mainly due to the limited training data for this age group. Another small spike is observed in the 2-3 age range, which indicates the model may also have difficulties with predicting the outcomes accurately for very young children.

In contrast, the MAE distribution for females shows a different trend, with the highest errors occurring in the younger age groups, particularly from 4-5 years, where the MAE peaks around 0.6. The error decreases in subsequent age groups, which indicates the improved accuracy as the age increases, while there are smaller peaks in the 8-9 and 12-13 age groups. This suggests that the model's performance is less stable in predicting the outcomes for younger females. These fluctuations could indicate that certain age-specific features for females are not well-represented in the model, which leads to inconsistent predictions.

The observed performance variations across age groups may be attributed to the following factors:

- **Data Distribution Imbalance**: Poor generalization in certain age groups may result from

underrepresentation in the training data. Insufficient samples for very young children or older individuals limit the model's ability to learn relevant patterns leading to higher errors.

- **Insufficient Capture of Age-Specific Features**: The increased MAE for males aged 18-19 and females aged 4-5 suggests that the model may not fully capture age-specific features. Adolescents exhibit a greater variability in health and behavior, while younger children may require specialized feature sets not present in the model.

These challenges underscore the importance of addressing data imbalance and incorporating relevant age-specific features through regularization techniques or balanced sampling strategies to improve model performance across all age groups.

4. Conclusions

In this paper, we proposed a fully automated method that combines the strengths of the GP and TW3 methods. By employing a U-net-based ROIs localization approach, the system reduces the clinicians' workload while maintaining the high interpretability and accuracy. Our method was validated on both public and in-house datasets, showing superior performance in terms of the localization precision and the MAE error compared to existing methods.

The method can streamline clinical workflows by automating BAA, reducing the assessment variability, and enabling faster and reliable diagnoses. This may facilitate earlier and more accurate detection of growth disorders. Although the method is robust across different datasets, future work should investigate the effects of image resolution and demographic variability to further enhance the generalizability in clinical practice.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by NSFC (Nos. 12090024 and 12201402) and Startup Fund for Young Faculty at SJTU (SFYF at SJTU). We thank the Student Innovation Center at Shanghai Jiao Tong University for providing us the computing services.

Conflict of interest

Xiaoqun Zhang is an editorial board member for Mathematical Biosciences and Engineering and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

1. W. Greulich, S. I. Pyle, *Radiographis Atlas of the Skeletal Development of the Hand and Wrist*, Stanford University Press, 1959.
2. J. M. Tanner, R. H. Whitehouse, N. Cameron, W. A. Marshall, M. J. R. Healy, H. Gold-stein, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*, Academic press, London, 1976.
3. R. M. Malina, G. P. Beunen, Assessment of skeletal maturity and prediction of adult height (TW3 method), *Am. J. Hum. Biol.*, **14** (2002), 788–789. https://doi.org/10.1002/ajhb.10098
4. H. H. Thodberg, S. Kreiborg, A. Juul, K. D. Pedersen, The bonexpert method for automated determination of skeletal maturity, *IEEE Trans. Med. Imaging*, **28** (2008), 52–66. https://doi.org/10.1109/TMI.2008.926067
5. S. Koitka, M. S. Kim, M. Qu, A. Fischer, C. M. Friedrich, F. Nensa, Mimicking the radiologists' workflow: Estimating pediatric hand bone age with stacked deep neural networks, *Med. Image Anal.*, **64** (2020), 101743. https://doi.org/10.1016/j.media.2020.101743
6. S. J. Son, Y. Song, N. Kim, Y. Do, N. Kwak, M. S. Lee, et al., TW3-based fully automated bone age assessment system using deep neural networks, *IEEE Access*, **7** (2019), 33346–33358. https://doi.org/10.1109/ACCESS.2019.2903131
7. H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, et al., Fully automated deep learning system for bone age assessment, *J. Digital Imaging*, **30** (2017), 427–441. https://doi.org/10.1007/s10278-017-9955-8
8. C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, R. Leonardi, Deep learning for automated skeletal bone age assessment in X-ray images, *Med. Image Anal.*, **36** (2017), 41–51. https://doi.org/10.1016/j.media.2016.10.010
9. P. Gong, Z. Yin, Y. Wang, Y. Yu, Towards robust bone age assessment: Rethinking label noise and ambiguity, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2020), 621–630. https://doi.org/10.1007/978-3-030-59725-2_60
10. V. I. Iglovikov, A. Rakhlin, A. A. Kalinin, A. A. Shvets, Paediatric bone age assessment using deep convolutional neural networks, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2018), 300–308. https://doi.org/10.1007/978-3-030-00889-5_34
11. D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V. Stence, C. P. Langlotz, Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs, *Radiology*, **287** (2018), 313–322. https://doi.org/10.1148/radiol.2017170236
12. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint*, 2015, arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556
13. C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Thirty-first AAAI Conference on Artificial Intelligence*, **31** (2017). https://doi.org/10.1609/aaai.v31i1.11231

14. C. F. Cheng, E. T. C. Huang, J. T. Kuo, K. Y. K. Liao, F. J. Tsai, Report of clinical bone age assessment using deep learning for an asian population in Taiwan, *BioMedicine*, **11** (2021), 50–58. https://doi.org/10.37796/2211-8039.1256

15. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 1137–1139. https://doi.org/10.1109/TPAMI.2016.2577031

16. S. S. Halabi, L. M. Prevedello, J. Kalpathy-Cramer, A. B. Mamonov, A. Bilbily, M. Cicero, et al., The RSNA pediatric bone age machine learning challenge, *Radiology*, **290** (2019), 498–503. https://doi.org/10.1148/radiol.2018180736

17. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

18. F. Jia, J. Liu, X. C. Tai, A regularized convolutional neural network for semantic image segmentation, *Anal. Appl.*, **19** (2021), 147–165. https://doi.org/10.1142/S0219530519410148

19. F. Milletari, N. Navab, S. A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)*, (2016), 565–571. https://doi.org/10.1109/3DV.2016.79

20. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint*, 2021, arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929

21. S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, et al., Adaptive histogram equalization and its variations, *Comput. Vision Graphics Image Proc.*, **39** (1987), 355–368. https://doi.org/10.1016/S0734-189X(87)80186-X

22. M. Everingham, S. Eslami, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: A retrospective, *Int. J. Comput. Vision*, **111** (2015), 98–136. https://doi.org/10.1007/s11263-014-0733-5

23. A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, H. Huang, Bone age assessment of children using a digital hand atlas, *Comput. Med. Imaging Graphics*, **31** (2007), 322–331. https://doi.org/10.1016/j.compmedimag.2007.02.012

24. X. Chen, J. Li, Y. Zhang, Y. Lu, S. Liu, Automatic feature extraction in X-ray image based on deep learning approach for determination of bone age, *Future Gener. Comput. Syst.*, **110** (2020), 795–801. https://doi.org/10.1016/j.future.2019.10.032

25. J. H. Lee, K. G. Kim, Applying deep learning in medical images: The case of bone age estimation, *Healthcare Inf. Res.*, **24** (2018), 86–92. https://doi.org/10.4258/hir.2018.24.1.86

26. S. Mutasa, P. D. Chang, C. Ruzal-Shapiro, R. Ayyala, Mabal: A novel deep-learning architecture for machine-assisted bone age labeling, *J. Digital Imaging*, **31** (2018), 513–519. https://doi.org/10.1007/s10278-018-0053-3

27. T. D. Bui, J. J. Lee, J. Shin, Incorporated region detection and classification using deep convolutional networks for bone age assessment, *Artif. Intell. Med.*, **97** (2019), 1–8. https://doi.org/10.1016/j.artmed.2019.04.005

28. J. Zhou, Z. Li, W. Zhi, B. Liang, D. Moses, L. Dawes, Using convolutional neural networks and transfer learning for bone age classification, in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, (2017), 1–6. https://doi.org/10.1109/DICTA.2017.8227503

29. C. Tong, B. Liang, J. Li, Z. Zheng, A deep automated skeletal bone age assessment model with heterogeneous features learning, *J. Med. Syst.*, **42** (2018), 1–8. https://doi.org/10.1007/s10916-018-1091-6

AIMS Press