Automated Bone Age Assessment Using Artificial Intelligence: The Future of Bone Age Assessment

Byoung-Dai Lee¹, Mu Sook Lee²

¹Division of Computer Science and Engineering, Kyonggi University, Suwon, Korea; ²Department of Radiology, Keimyung University Dongsan Hospital, Daegu, Korea

Bone age assessments are a complicated and lengthy process, which are prone to inter- and intra-observer variabilities. Despite the great demand for fully automated systems, developing an accurate and robust bone age assessment solution has remained challenging. The rapidly evolving deep learning technology has shown promising results in automated bone age assessment. In this review article, we will provide information regarding the history of automated bone age assessments, discuss the current status, and present a literature review, as well as the future directions of artificial intelligence-based bone age assessments.

Keywords: Bone age assessment; Left hand and wrist radiographs; Artificial intelligence; Convolutional neural network; Deep learning

INTRODUCTION

Although the definition of artificial intelligence (AI) varies, it commonly refers to the technology of creating machines demonstrating intelligence similar to that of humans [1].

AI is increasingly being used in medicine, considering the remarkable advances in the field. Deep learning is a subclass of AI that exploits artificial neural networks (ANNs). The most common ANN, which is used for image analysis and recognition-related problems, is the convolutional neural network (CNN). CNNs form a connectivity pattern between neurons similar to the organization of the animal visual cortex and use supervised learning (Fig. 1) [2]. CNNs show state-of-the-art performance for image classification and segmentation-related tasks. Recurrence neural networks (RNNs) are another type of ANN that is suitable for recognizing the regular patterns in data containing serial or temporal information. Thus, RNNs are widely used in fields that handle serial data such as music, text processing, lyric making, songwriting, language translation, and stock prediction (Fig. 2) [3].

AI has also been extensively employed in the field of medical imaging for applications such as detection and classification of an object or lesion, anatomical object localization, organ and substructure segmentation, lesion segmentation, and registration [4]. Numerous studies have exploited CNN-based deep learning for neuroanatomic structure segmentation, classification of neurodegenerative diseases, detection and classification of chest abnormalities, liver segmentation, classification of pathologic findings in the liver, detection and classification of breast cancer, and bone age assessment [5-11].

Bone age assessment is a representative case where the concept of object detection and classification can be applied. Numerous attempts have been made to develop an automated system in the past several years [12]. Bone age assessments have become a major target of the machine learning community, since the task is a typical object detection and classification problem of deep learning. In this approach, for a given input (e.g., the left hand radiograph including the distal radius and ulnar epiphysis) and a corresponding class (e.g., a class corresponding
to a bone age) is predicted [13]. Automated bone age assessments using CNN-based machine learning models have shown remarkable performance in recent years [11, 13-15].

The purpose of this review article was to discuss the current status of AI-based bone age assessments and its future direction. This study also aimed to provide an insightful introduction to the history of automated bone age assessments and present a literature review on AI-based bone age assessments.

**History of Automated Bone Age Assessments**

Bone age is a marker of bone maturity. It is determined based on the shape and maturity level of the primary and secondary ossification centers and the time of fusion between the two [16]. Atlas-based methods such as the Greulich-Pyle (GP) as well as the Gilsanz-Ratibin methods are used to determine bone age [17, 18]. In contrast, the Tanner-Whitehouse (TW) method, which is a scoring method, considers selected radiographic regions of interest (ROI) in specific bones of the left hand and categorizes the development level of each ROI into stages (from A to I). The total score is calculated by adding the scores assigned to each stage of development for each bone. Finally, the total score is converted into bone age [19]. Recently, the revised TW3 method was proposed, which assesses the maturity of the radius, ulna, and short bones (RUS) [19].

While studies have attempted to assess bone age using ultrasound and MRI, the validation process remains incomplete [20, 21]. Moreover, bone age assessments using left hand radiographs are less expensive than ultrasound and MRI. Although there are concerns regarding radiation exposure, the dose for bone age assessment with left hand radiographs is low (0.001–0.1 mSV) and is considered safe [22]. This dose is lower than the amount from a 20-minute exposure to natural radiation [23].

Bone age is a size-independent indicator of biological maturity. It is better associated with height velocity,
AI-Based Automated Bone Age Assessment: Is This the New Era of Bone Age Assessment?

Following the second AI boom in the 1980s and 1990s, the term ‘computer-aided detection’ (CAD) was coined [30]. Subsequently, after the third AI boom, CAD was divided into conventional and AI-based detection [31]. AI-based CAD is task-agnostic; it uses a deep learning algorithm to train itself with the given data [30].

Bone age images are an ideal dataset for training a deep learning solution, as there is a single image of the left hand and wrist and relatively standardized findings [14, 32]. In 2019, Dallora et al. [33] reviewed machine-learning-based automated bone age assessment solutions including the regression-based method, ANNs, CNNs, support vector machines, Bayesian networks, decision trees, and K-nearest neighbors [13, 34-40]. They summarized their systematic literature review as follows [33]: 1) most studies aimed to propose an automatic bone age assessment system; 2) research has focused on hand and wrist radiographs; 3) the ethnicity and socioeconomic aspects were not explored in detail; and 4) the average performance weighted by the sample size of the compared studies resulted in a MAE of 9.96 months. After this meta-analysis was published, many studies were conducted on various deep learning-based bone age assessment solutions, such as content-based image retrieval [41]. However, only a few of these AI-based bone age assessment solutions have been commercialized.

BoneXpert, an AI-based bone age assessment solution was introduced in 2008 [42]. BoneXpert is an AI system that uses feature extraction techniques and calculates bone age by analyzing the left hand radiograph based on 13 bones (radius, ulna, and 11 short bones in fingers 1, 3, and 5). Once the left hand radiograph is sent to the BoneXpert AI software server, the software applies an active appearance model (which has learned the regular shape and density distribution of each analyzed bone) to analyze the 13 bones. Lastly, the final bone age is determined by either the GP or the TW method. The left hand radiograph marked with the final bone age is transferred to the Picture of Archiving and Communication System (Fig. 3). BoneXpert is an established bone age assessment system and is widely used in Europe. It has been validated through comparisons of manual ratings in several studies. According to Booz et al. [43], the correlation between BoneXpert-derived and reference bone ages ($r = 0.99$) was significantly higher than that between the reader-calculated and reference bone ages ($r = 0.90$; $p < 0.001$). Moreover, BoneXpert requires considerably less time for image interpretation than manual rating using the GP method, thereby improving the time efficiency in routine clinical practice [43]. However, BoneXpert showed limited efficacy when fewer than eight bones were included, as well as in cases of poor image quality and abnormal bone morphology [43]. These limitations may be overcome through consistent version upgrades.
VUNO Med-BoneAge is the first AI-based bone age assessment solution approved by the Korea Food & Drug Administration and is commercialized in the country [15]. VUNO Med-BoneAge uses a deep learning solution. It was trained using 18,940 left hand radiographs analyzed by the GP method. On receiving the image to be analyzed, the system suggests three most likely estimated bone ages (i.e., first-, second-, and third-rank AI bone ages) based on probabilities along with similar images for comparison (Fig. 4). The accuracy of the first-rank bone age was 69.5%, which increased to 93% when the first-, second-, and third-rank bone ages were combined [15]. Strictly speaking, VUNO Med-BoneAge is a semiautomatic system, rather than being fully automatic, because a human has to choose one of the three bone age results and images suggested by the deep learning model. However, VUNO Med-BoneAge was reported to show a 29% reduction in the image interpretation time compared to manual rating [15].

Apart from VUNO Med-BoneAge, which uses the GP method, another solution using the TW3 method has been approved by the Korea Food & Drug Administration. The HH-boneage.io solution is a fully automatic system that localizes the epiphysis-metaphysis growth regions of the 13 bones, including the radius, ulna, and 11 short bones in fingers 1, 3, and 5, and estimates the corresponding bone age [11]. Once an image is uploaded, the system determines the ROIs of the 13 bones and calculates the maturity score for each ROI and the total score. The system then predicts bone age using a correlation matrix (Fig. 5). The system has a MAE and root mean squared error of 0.46 years and 0.62 years, respectively, when compared to the bone age

Fig. 3. Automated bone age assessment by BoneXpert. Once left hand and wrist radiographs are sent to the BoneXpert artificial intelligence software server, the software applies an active appearance model to analyze the 13 bones. Following this, the left hand and wrist radiographs marked with the final bone age are sent to the Picture of Archiving and Communication System (source: https://bonexpert.com).

Fig. 4. Automated bone age assessment by VUNO Med-BoneAge. On receiving the image to be analyzed, VUNO Med-BoneAge suggests three most likely estimated bone ages (i.e., first-, second-, and third-rank artificial intelligence bone ages) based on probabilities along with similar images for comparison (source: https://www.vuno.co).
Fig. 5. Automated bone age assessment by HH-boneage.io solution. The HH-boneage.io solution is a fully automatic system that localizes the epiphysis-metaphysis growth regions of 13 bones including the radius, ulna, and 11 short bones in fingers 1, 3, and 5, and estimates the corresponding bone age (source: http://www.boneage.io). ROI = regions of interest.
determined by an expert. The system demonstrated 97.6% accuracy within one year of ground truth [11].

The MediAI-BA solution is another AI-based bone age assessment solution, approved by the Korea Food & Drug Administration, which uses the TW3 method [44]. Unlike the traditional TW3 method, this system analyzes seven epiphysis-metaphysis growth regions in the radius, ulna, metacarpal of the 1st finger, and metacarpal, proximal, middle, and distal phalanges of the 3rd finger (Fig. 6). The MediAI-BA solution has a MAE of 0.59 years [44].

**Current Issues in AI-Based Automated Bone Age Assessment**

AI-based automated bone age assessments are validated against manual bone age assessments by experts. Left hand radiographs are manually analyzed using the GP or TW method for use as reference data. However, there is no method to confirm that the manually determined age represents the true bone age [24]. The question of whether AI-based assessment results should be compared with those obtained from other assessment methods (aside from the ones using the GP or TW method) and with other types of medical images, such as ultrasound and MRI (besides left hand and wrist radiographs), persists during the development of an AI-based automated bone age assessment solution. The application of AI-based bone age assessment solutions to populations of different ethnicities is another significant issue. Bone age assessment is influenced by ethnicity, since it has been proven that different populations show different rates of skeletal maturation [45]. Therefore, it is reasonable to doubt the direct application of AI-based bone age assessment solutions, particularly those based on the GP method, in populations of different ethnicities other than the groups considered during development. Recognizing this issue, studies are being conducted on their applicability in diverse ethnic groups. For instance, according to a recent study that validated BoneXpert in American children of four ethnicities (African American, Asian, Caucasian, and Hispanic), the AI-based bone age assessment solution could analyze images of all ethnicities and therefore eliminate the problem associated with rater variability [46]. In addition, although AI may significantly increase the efficiency of bone age assessment by pediatric radiologists, congenital or acquired abnormal bone morphologies cannot be detected on left hand radiographs. Thus, the necessity for manual assessments persists in patients who undergo surgeries such as epiphysiodesis or show sequelae of growth plate injury. In addition, most research and commercially usable solutions for AI-based bone age assessment have focused on left hand radiographs. MRI offers some advantages compared to hand radiographs, such as no requirement for ionizing radiation, more accurate analysis of the growth
plate, and reduced subjective influence of the examiners [47-49]. However, there are limited research papers on automated bone age assessment using MRI [36, 39, 49, 50]. Although these studies reported good performance of MRI in the estimation of bone age, a comparative study between radiograph- and MRI-based age estimation solutions is needed. Lastly, high heterogeneity in research and solutions is also an issue. Owing to the high heterogeneity in terms of age ranges, dataset sizes, and performance metrics, the comparison between studies is challenging [33]. Accumulation of data regarding the clinical efficacy through multi-center and multi-national clinical trials must precede use of such AI-based solutions.

However, the development of and research on automated bone age assessment using AI will continue, which will result in the commercialization of more solutions. The existing commercialized solutions will also undergo a series of upgrades. Therefore, the accuracy of AI-based automated bone age assessment solutions will improve and the current limitations will be addressed.

CONCLUSION

The era of AI may be considered an immense revolution with an irreversible tendency. The research objective of enhancing the performance of AI-based systems has now shifted to exploring ways to efficiently exploit this intelligence to further optimize human performance. Currently, AI in medical imaging is expected to act as an assistant that can reduce the burden of doctors rather than competing against them. AI-based automated bone age assessments can reduce the burden of radiologists who handle a large number of images to determine bone age. It can also significantly reduce the subjectivity, and inter- and intra-observer variabilities associated with traditional bone age assessment methods.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Mu Sook Lee. Data curation: all authors. Methodology: all authors. Project administration: all authors. Resources: all authors. Software: Byoung-Dai Lee. Supervision: Mu Sook Lee. Validation: all authors. Visualization: all authors. Writing—original draft: Byoung-Dai Lee. Writing—review & editing: Mu Sook Lee.

ORCID iDs

Byoung-Dai Lee https://orcid.org/0000-0002-4028-6168

Mu Sook Lee https://orcid.org/0000-0002-0382-5564

REFERENCES

1. Sarmah SS. Concept of artificial intelligence, its impact and emerging trends. Int Res J Eng Technol 2019;6:2164-2168
2. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun Acm 2017;60:84-90
3. Yüksel K, Skarbek W. Convolutional and recurrent neural networks for face image analysis. Found Comput Decis Sci 2019;44:331-347
4. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;24:60-88
5. Kushibar K, Valverde S, González-Villá S, Bernal J, Cabezas M, Oliver A, et al. Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. Med Image Anal 2018;48:177-186
6. Álvarez JD, Matias-Guiu JA, Cabrera-Martin MN, Risco-Martín JL, Ayala JL. An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders. BMC Bioinformatics 2019;20:491
7. Hwang EJ, Park CM. Clinical implementation of deep learning in thoracic radiology: potential applications and challenges. Korean J Radiol 2020;21:511-525
8. Wang K, Mamidipalli A, Retson T, Bahrami N, Hasenstab K, Blansit K, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. Radiol Artif Intell 2019;1:180022
9. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med 2020;3:23
10. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577:89-94
11. Son SJ, Song Y, Kim N, Do Y, Kwak N, Lee MS, et al. TW3-based fully automated bone age assessment system using deep neural networks. IEEE Access 2019;7:33346-33358
12. 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
13. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. J Digit Imaging 2017;30:427-441
14. Tajmir SH, Lee H, Shailam R, Gale HI, Nguyen JC, Westra SJ,
et al. Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. *Skeletal Radiol* 2019;48:275-283

15. Kim JR, Shim WH, Yoon HM, Hong SH, Lee JS, Cho YA, et al. Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency. *AJR Am J Roentgenol* 2017;209:1374-1380

16. Mansourvar M, Ismail MA, Herawan T, Raj RG, Kareem SA, Nasarudin FH. Automated bone age assessment: motivation, taxonomies, and challenges. *Comput Math Methods Med* 2013;2013:391626

17. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist, 1st ed. Stanford: Stanford University Press, 1959

18. Gilsanz V, Ratib O. A digital atlas of skeletal maturity. Berlin: Springer, 2011

19. Tanner JM, Healy MJ, Goldstein H, Cameron N. Assessment of skeletal maturity and prediction of adult height: TW3 method. London: W.B Saunders Company, 2001

20. Khan KM, Miller BS, Hoggard E, Somani A, Sarafoglou K. Application of ultrasound for bone age estimation in clinical practice. *J Pediatr* 2009;154:243-247

21. Hojreh A, Gamper J, Schmook MT, Weber M, Prayer D, Herold CJ, et al. Hand MRI and the Greulich-Pyle atlas in skeletal age estimation in adolescents. *Skeletal Radiol* 2018;47:963-971

22. Mettler FA Jr, Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology* 2008;248:254-263

23. Martin DD, Witt JM, Hochberg Z, Sävendahl L, van Rijn RR, Fricke O, et al. The use of bone age in clinical practice-part 1. *Horm Res Paediatr* 2011;76:1-9

24. van Rijn RR, Thodberg HH. Bone age assessment: automated techniques coming of age? *Acta Radiol* 2013;54:1024-1029

25. Roche AF, Rohmann CG, French NY, Dávila GH. Effect of training on replicability of assessments of skeletal maturity (Greulich-Pyle). *Am J Roentgenol Radium Ther Nucl Med* 1970;108:511-515

26. Michael DJ, Nelson AC. HANDX: a model-based system for automatic segmentation of bones from digital hand radiographs. *IEEE Trans Med Imaging* 1989;8:64-69

27. Pietka E, McNitt-Gray MF, Kuo ML, Huang HK. Computer-assisted phalangeal analysis in skeletal age assessment. *IEEE Trans Med Imaging* 1991;10:616-620

28. Tanner JM, Oshman D, Lindgren G, Gunbaum JA, Elsouki R, Labarthe D. Reliability and validity of computer-assisted estimates of Tanner-Whitehouse skeletal maturity (CASAS): comparison with the manual method. *Horm Res* 1994;42:288-294

29. Yoo JW, Lee JM, Kim WY. A bone age assessment method based on normalized shape model. *J Korea Multimed Soc* 2009;12:383-396

30. Oakden-Rayner L. The rebirth of CAD: how is modern AI different from the CAD we know? *Radiology: Artificial Intelligence* 2019;1:e180089
46. Thodberg HH, Sävendahl L. Validation and reference values of automated bone age determination for four ethnicities. *Acad Radiol* 2010;17:1425-1432

47. Margalit A, Cottrill E, Nhan D, Yu L, Tang X, Fritz J, et al. The spatial order of physeal maturation in the normal human knee using magnetic resonance imaging. *J Pediatr Orthop* 2019;39:e318-e322

48. Fan F, Zhang K, Cui JH, Hu N, Deng ZH. Forensic age estimation of living persons from the knee: comparison of MRI with radiographs. *Forensic Sci Int* 2016;268:145-150

49. Štern D, Payer C, Urschler M. Automated age estimation from MRI volumes of the hand. *Med Image Anal* 2019;58:101538

50. Dallora AL, Berglund JS, Brogren M, Kvist O, Diaz Ruiz S, Dübbel A, et al. Age assessment of youth and young adults using magnetic resonance imaging of the knee: a deep learning approach. *JMIR Med Inform* 2019;7:e16291