Traditional and New Methods of Bone Age Assessment-An Overview

© Monika Prokop-Piotrkowska1, Kamila Marszałek-Dziuba1, Elżbieta Moszczyńska1, Mieczysław Szalecki2, Elżbieta Jurkiewicz3

1Children’s Memorial Health Institute, Department of Endocrinology and Diabetology, Warsaw, Poland
2Jan Kochanowski University, Collegium Medicum, Kielce, Poland
3Children’s Memorial Health Institute, Department of Diagnostic Imaging, Warsaw, Poland

Abstract
Bone age is one of biological indicators of maturity used in clinical practice and it is a very important parameter of a child’s assessment, especially in paediatric endocrinology. The most widely used method of bone age assessment is by performing a hand and wrist radiograph and its analysis with Greulich-Pyle or Tanner-Whitehouse atlases, although it has been about 60 years since they were published. Due to the progress in the area of Computer-Aided Diagnosis and application of artificial intelligence in medicine, lately, numerous programs for automatic bone age assessment have been created. Most of them have been verified in clinical studies in comparison to traditional methods, showing good precision while eliminating inter- and intra-rater variability and significantly reducing the time of assessment. Additionally, there are available methods for assessment of bone age which avoid X-ray exposure, using modalities such as ultrasound or magnetic resonance imaging.

Keywords: Maturation, children, radiographs, deep learning, neural networks

Introduction

Maturation Indicators
The processes of growth and maturation in children are usually correlated, but they cannot be treated as one process as they may not be linear and may proceed at different paces. Due to numerous disturbances, such as growth hormone (GH) deficiency, deficiency of thyroid hormones or delayed puberty, but also sometimes in healthy children, the chronological age (CA) doesn’t match the biological age. This is because they are regulated by various factors, which include genes and nutrition, but also include many hormones, including GH, insulin-like growth factor-1, sex hormones and adrenal steroids such as cortisol, dehydroepiandrosterone, and testosterone (1,2). In paediatric endocrinology, it is especially important to assess the child’s growth and puberty in relation to biological age, rather than CA. Thus, clinicians have been looking for a good marker of maturation rate in children for decades (3).

Age at menarche is a solid biological indicator of maturity, but it is a one-off event and relates to only half of the population (3). Dentists, mainly orthodontists, use dental age judged using the Demirjian or Willems scale in daily practice but this practice has not been established as a reliable tool for other clinicians (3,4,5). Sexual characteristics, such as that made by assessment of position on the T anner scale, are useful only in the adolescent period and are very subjective. The only biological indicator of maturity, which is available from birth to adulthood, is bone age (BA) (3).

Bone Age
In paediatric endocrinology, BA is an important tool used in the clinical assessment of patients, mainly those suffering from growth and puberty disorders. Many parameters correlate better with BA than with CA including height velocity, menarche, muscle mass and bone mineral mass (6). Delayed BA is typical for GH deficiency, constitutional delay of growth, hypothyroidism, malnutrition and chronic...
illness (6,7). On the other hand, BA is advanced in many conditions that include precocious puberty and congenital adrenal hyperplasia, when there is a prolonged elevation of sex steroid levels (6,7,8). BA may be also marginally advanced in cases of overweight children, children with tall stature or premature adrenarche (1,6,8). In genetic overgrowth syndromes, for example Sotos syndrome, Beckwith-Wiedemann syndrome and Marshall-Smith syndrome, BA is usually significantly advanced (6). In all cases it is important to remember that advancement or delay of BA in relation to CA is a slow process, thus BA may not be altered in the case of examinations performed shortly after the first manifestations of a disorder and should be assessed in a temporal manner (7).

What is more, BA is used in forensic and legal medicine to estimate CA, for example in asylum seekers or unaccompanied minors without documents. In such cases an adequate assessment of age using precise methods is crucial. The consequences of incorrect assessment of a child as an adult may result in more restricted access to education, medical care or other forms of support provided for children (9).

This article considers different methods of BA assessment from the perspective of a paediatrician or paediatric endocrinologist (Table 1).

Traditional Methods

Although there have been attempts to assess BA by examinations of specific bones, such as the clavicle or iliac bone (Risser sign) (10,11,12,13,14,15), in paediatrics and paediatric endocrinology, the established way to obtain BA is by performing a radiograph of the hand and wrist of the non-dominant hand. Assessment of development of the bones can be performed in the traditional, manual way or using one of the automated methods. The manual method involves a comparison of obtained radiograph with radiographs in atlases. The manual methods can be divided into two groups depending on the type of atlas – holistic or analytic.

The first atlases were published shortly after the discovery of X-rays in 1895. In 1898, John Poland published the first one: “skiagraphic atlas showing the development of bones of the wrist and hand” (16). In his atlas, he depicted skiagraphs (positive reprints) of hand radiographs of 19 British children, aged between 1 and 17 years, with an attached description of each radiograph (16). However, the two most important publications in this field were issued in 1959 by Greulich and Pyle (17) and in 1962 by Tanner, Whitehouse and Healy (18).

Greulich-Pyle Atlas

‘The Radiographic Atlas of Skeletal Development of the Hand and Wrist’ by Greulich and Pyle (17) (GP) has been widely recognized and is used in many centers currently. This atlas was created based on radiographs of hands of paediatric patients referred to endocrinologists William Walter Greulich and Sarah Idell Pyle by paediatricians between the years 1951-1942. These patients were Caucasian children from a generally upper middle class background, living in Cleveland, Ohio, United States (19,20). This atlas consists of separate reference images for boys and girls aged 0-18 (boys) or 0-19 years (girls) in various intervals (3 months-1 year). Images are accompanied by an explanation of the gradual age-related changes in the bones at a given age and separate BAs calculated for each bone. Due to the natural variability of the BA of different bones in one individual, in some bones, it is often more or less advanced than the standard it is intended to represent. For example, a radiograph representing the age of 5 years 6 month (42 months) includes a 36-month first metacarpal and a 54-month lunate (17). BA is calculated by comparing the non-dominant wrist radiographs of the subject with the nearest matching reference radiographs provided in the atlas. Thus this method is termed a holistic method. Figure 1 presents GP atlas.
GP is the most popular method among clinicians and radiologists, as the assessment by GP is relatively quick and easy to learn. Although widely used, this method has significant drawbacks. BA assessment (BAA) using GP shows high inter- and intra-observer variability. In addition, given the reference population used in GP, this method may not be an appropriate, universal tool for use in various populations.

BAA by GP is very subjective and the standard error on a single determination in inter-observer studies ranges from 0.45 to 0.83 years (21,22,23,24,25). There is no standardization in how the bones are weighted. Depending on a rater, in clinical practice one may assign different weights to different bones, some raters may ignore the carpals and others may assign even half weight to the carpals during the assessment. Raters using the carpals reduce their importance at higher maturity but again not in a standardized manner (24).

It has been reported that currently boys and girls develop secondary sex characteristics earlier than decades ago in United States (26,27). Thus current use of the GP atlas, even in a similar population to the original source, may not be as precise as when it was created.

What is more, it has been proven that correlation of BA with CA, and consequently the applicability of GP, depends on ethnic origin (28,29). According to a recent meta-analysis it has been proven that in African females, in comparison to GP standards, BA is significantly advanced. Conversely, in Asian males, BA is significantly delayed between 6 and 9 years of age and significantly advanced at 17 years (28). This should be taken into consideration while assessing BA in these populations using the GP atlas.

There is an online version of GP uploaded by Brazilian Instituto Mineiro de Endocrinologia (28).

**Tanner-Whitehouse Atlas**

The second most popular tool for BA assessment is the Tanner-Whitehouse atlas (TW). The first version of TW was created in 1962 based on 2600 radiographs collected in the 1950s and 1960s of British children coming from average socio-economic class (18). It was later updated in 1983 to Tanner-Whitehouse 2 (TW2) and in 2001 the latest updated version was published - Tanner-Whitehouse 3 (TW3). These updates have attempted to adjust for the secular trends that influence the relationship between the total bone maturity score and BA (30). In several countries standardized TW methods have been created which change the relationship between the total maturity score and BA to make it suitable for different ethnic groups (31,32,33).

TW2 is an analytic or scoring method and it is based on the maturity levels of 20 regions of interest (ROI) in different bones of the hand and wrist. The level of development of each ROI is labeled as a given stage, which is then converted to a numerical score. A total maturity score is calculated by adding the scores of the ROIs and it is matched with the age of boys and girls separately.

The TW method is considered to be more objective than the holistic GP method and to also exhibit higher reproducibility than GP. Bull et al (21) reported that the intra-observer variation was greater using GP than TW (95% confidence interval, -2.46 to 2.18 vs -1.48 to 1.43, respectively). However, assessment using the TW method is more time-consuming. In a study performed by King et al (34) the average time required for TW assessment was calculated as 7.9 min. vs. 1.4 min. in the case of GP assessment. In this study the intra-observer variation between GP and TW assessment was also found to be insignificant (the average spread of results was 0.74 years for TW and 0.96 years for the GP). It should be noted that the sample size assessed by King et al (34) was considerably smaller than that assessed by Bull et al (21). A comparison of GP and TW methods is presented in Table 1 (Table 2).

**Other Atlases**

The FELS method was developed in 1988 using 13,823 serial radiographs of the left hand-wrist of boys and girls in

| Table 1. Bone age assessment methods | Manual | Automatic |
|-------------------------------------|--------|----------|
| Radiograph                          | - Greulich-Pyle Atlas (17) | - CASAS (55), |
|                                     | - Tanner-Whitehouse Atlas (30) | - BoneXpert (71,72) |
|                                     | - FELS Method (36) | - AI methods (97-109) |
|                                     | - Gilsanz and Ratib Atlas (37) | |
| MRI                                 | - Pediatric Hand MR Scanner (45,46) | - Method of Štern et al (51) |
|                                     | - Method of Tomei et al (48) | |
|                                     | - Method of Hojreh et al (49) | |
| USG                                 | - Femoral head cartilage thickness (44) | - BonAge (40) |
|                                     | - Risser’s stage (45) | |

MRI: magnetic resonance imaging, USG: ultrasonography
the Fels Longitudinal Study performed by William Cameron Chumlea, Alex F. Roche and David Thissen from two universities in Kansas and Ohio, US (35). It is based upon maturity indicators that represent radiographic features that occur during the maturation of every child (35). The set of maturity indicators is analysed with a computer program that provides the BA and the standard error for that assessment (35). However, the FELS method has not gained wide recognition.

In 2005 a digital atlas created by Vicente Gilsanz and Osman Ratib (GR) was published. It consists of artificially created, idealised images of hands and wrists, specific for age and sex. These images were produced by an analysis of the size, shape, morphology and density of ossification centres of 522 hand radiographs from healthy Caucasian children from Los Angeles, US (50% girls and 50% boys). Each image includes typical characteristics of development for each of the ossification centres (36). The images are of better quality and precision in comparison to GP. Another advantage is the regular spacing of the images at 6-monthly intervals from the ages of 2 to 6 years and yearly intervals from the age of 7 to 17 years (37). In one study the GR atlas was compared to GP and it was concluded that they were comparable in terms of precision. Yet again, however, the study was performed on a small number of examinations (38).

**Ultrasound Assessment**

Other imaging modalities, which have developed considerably over the years, now offer some advantages over the ubiquitous radiograph for assessment of BA. One of these is ultrasound (USG), the major advantage of which is that it does not expose the patient to any ionizing radiation, important when patients receive sequential assessment of BA. Some studies have been performed to establish different methods of BAA, including by performing USG (39).

A result of one of these trials is BonAge® (Sunlight Medical Ltd, Tel Aviv, Israel) which consists of a device that performs an ultrasonographic examination and software that calculates the BA on the basis of this examination (19,40,41,42,43). BonAge® measures the ossifying cartilage structures of the wrist as an ultrasonic wave passes through the subject’s distal radius and ulnar epiphysis. According to the producer, BonAge® provides on-the-spot, easy-to-read, immediate results, without exposing children and adolescents to ionizing X-ray radiation, and moreover, it is objective and safe (40). The time of the examination is approximately five minutes although this can prove problematic in the smallest children (41).

Several studies have been performed to assess the precision of this instrument. Mentzel et al (41) and Shimura et al (42) concluded that the results of BonAge® examinations correlate closely with BA evaluated conventionally using the GP or TW2 method. However, in a more recent study performed by Khan et al (43) on a bigger number of patients it was shown that BonAge® tended to over read delayed BA and under read advanced BA and the authors concluded that ultrasonographic assessment should not yet be considered a valid replacement for radiographic BAA.

There has also been a report of ultrasonographic assessment of the thickness of anterior femoral head cartilage, which correlates strongly with the child’s CA and BA, standing height and body weight, according to the authors of the study (44). Ultrasonic examination of ossification of the iliac crest apophysis, (Risser’s sign), was also studied and it presented with high accuracy, specificity and sensitivity in comparison to hand X-ray examination and GP assessment (45).

Although the majority of the authors of these studies conclude that USG methods investigated are of good accuracy in comparison to hand X-ray, USG-based BAA is rarely used in daily practice. This may be because the examination needs to be performed by a trained specialist or there is a need for a specific device. In both cases, it takes more time to perform than an X-ray. Taking into consideration that most studies investigating the utility of USG in BAA were performed on relatively small groups of patients, the clinical utility of USG examination is as
yet unproven. Isolation of the forearm allows for minimal
radiation exposure and the radiation during hand X-ray is
very low (0.0005 mSv). However, in the future, USG may be
an advantageous method that may allow total elimination of
children’s exposure to ionizing radiation during BAA.

Magnetic Resonance Imaging Assessment

The first research in the field of BAA using magnetic
resonance imaging (MRI) was performed in 2007 to find a
tool suitable to establish the age of male football players
without unnecessary radiation exposure (29). Since in
some Asian and African countries registration at birth is
not compulsory, age determination is crucial to prevent
participation in the incorrect age group (29).

In 2012 Terada et al (46) reported a technique for BAA
based on MRI examination. BA was determined using an
open, compact, newly designed MR imager optimized for
evaluation of a child’s hand and wrist and it was scored by
two raters using the TW system adapted for the Japanese
population. Evaluation of this method was performed on a
group of 93 healthy Japanese children and a strong positive
correlation with BA and CA was demonstrated. What is
more, the intra-and inter-rater reproducibility rates were
significantly high (46). Another study from the same authors
was performed in 2014 to improve the performance of this
method (47). This was conducted on a group of 88 healthy
children with three raters assessing BA and it confirmed
the reliability and validity of this method (47). However, a
disadvantage of MRI is that it requires a relatively long time
to be performed (2 min and 44 sec), therefore it may not be
suitable for the youngest children, due to body movement.

Another study was performed by Tomei et al (48) and this
was published in 2014. They performed hand and wrist
MRIs on 179 healthy children aged 11-16 years old and
analyzed the correlation with CA. It was concluded that
BAA with MRI was feasible and showed good inter-observer
reproducibility (48).

In 2017 the results of another study were published
regarding the use of MRI in BAA. Hojreh et al (49) performed
hand MRI and X-ray examinations in 50 healthy volunteers
and 10 patients, all of whom were adolescents (aged 15 ± 2
years and 13.5 ± 2.6 years, respectively) and assessed both
examinations according to GP criteria. This study concluded
that the correlation between estimated patients’ ages on
radiographs assessed by GP and MRI was high with the
average estimated age difference between the MRIs and
radiographs being −0.05/−0.175 years. However larger,
multicenter studies are necessary to confirm the usefulness
of this method. There have also been attempts to automate
the BAA using MRI instead of radiography (50,51). The
comparison of RTG, USG and MRI methods is presented in
Table 3.

Automated Techniques

Due to the problems associated with BAA when using
traditional methods, such as inter- and intra-observer
variability and the fact that it is time-consuming, a need
emerged for new, objective tools that would provide
immediate results. As Computer-Aided Diagnosis (CAD) has
emerged and has started to be used in clinical practice, one
obvious procedure, which would be suitable for adaptation
to CAD was BAA, and BA was one of the first radiologic
examinations to be automated. This is not recent, however.
The first trials of CAD in BAA date back to 1989 when a
semi-automated system called HANDX was introduced by
Michael and Nelson (52). More recently, work on a system
which is based on assessment of phalangeal regions of
interest (PROI) was published by Pietka et al (53) in 1991.
In this method, the PROI were detected and the lengths of
the distal, middle, and proximal phalanx were measured

| Table 3. Comparison of radiographic, ultrasonographic and magnetic resonance imaging-based methods |
|---------------------------------------------------------------|
| **Method** | **Radiograph** | **Ultrasonography** | **Magnetic resonance imaging** |
| **Advantages** | - The most frequently used  
- Many recognised atlases  
- Easy to perform  
- Quick  
- Accessible  
- Doesn’t require a radiologist to perform, only to assess  
- Automated methods available | - No X-ray exposure | - No X-ray exposure  
- Accuracy validated in studies  
- There are attempts to automate BAA using MRI |
| **Disadvantages** | - X-ray exposure  
- Presence of radiologist required to perform  
- Time consuming  
- Only few studies on its accuracy | | - Not easily accessible  
- Relatively time consuming (quicker than USG) |

BAA: bone age assessment, USG: ultrasonography, MRI: magnetic resonance imaging
automatically. BA was estimated using the standard phalangeal length table, presented earlier by Garn et al (54).

**CASAS**

However, the first system to be used by different authors in studies was CASAS - a computerized image analysis system for estimating TW2 BA (55). This semi-automated system was introduced by Tanner and Gibbons in 1994 and it used the 13 bones of TW RUS system (radius, ulna and short bones) for BAA. These bones had to be located manually on the screen by a rater (correct positioning was assured by computer templates of each bone stage) and then automatic scoring was performed. Tanner and Gibbons (55) concluded that CASAS was more reliable and valid than manual TW RUS rating (56). Although other researchers have also reported that CASAS was useful and reliable (57,58), this system has not been widely adopted. The major drawback was that it took more time to estimate BA with CASAS than a manual TW assessment. In addition, difficulties with BAA in cases of abnormally shaped bones restricted the use of CASAS in some pathological conditions.

More recently there have been numerous approaches to BAA automation (58-71) and the most important ones are described below.

**BoneXpert**

This automated tool for BAA was created in 2008 by the Visiana company, based in Holte, Denmark (72,73,74). This computer program analyses BA automatically, in several steps. The first step is the definition of borders and intensity of the radiologic image of 13 points of interest of the same 13 bones used in the TW RUS system, that is the radius, ulna and 11 short bones. During this first step the system also defines if the picture is complete and of appropriate technical quality. In the next step, BA is assessed for each of the 13 bones separately. The last step is the transformation of the summary BA according to GP and TW criteria (72,73).

Figure 2 presents BAA by BoneXpert. BAA is available for ages 2.5-19 years for boys and 2-18 years for girls (version 2.4.5.1) (75). The data set used for the creation of this program consisted of 1678 hand radiographs of healthy Danish children and children from Belgium diagnosed with a range of disorders, such as Turner syndrome (73).

To date several papers have been published that verify the reliability and precision of BAA using BoneXpert in comparison to GP in different populations (Table 4). In European populations, studies have been conducted among healthy children from the Netherlands (405 patients), German children with short stature (1,097 patients), precocious or early puberty (116 patients), congenital adrenal hyperplasia (100 patients) and with various other endocrinological disturbances (514 patients) (75,76,77,78,79). Moreover, there was a study conducted with 1100 healthy American children from four different ethnic groups (Caucasian, African American, Asian and Hispanic) (22) and another on 515 eutrophic, overweight and obese children from Brazil (80). Research into the validity of BoneXpert has also been performed in Asian populations, including a study on 597 healthy children from Shanghai, China (81), in a large population of 6026 healthy children from five different cities in China (82) and among Japanese children, using 185 radiographs from 22 healthy children and 284 radiographs from 22 patients diagnosed with GH deficiency (83).

What is more, studies have confirmed the validity of BAA via BoneXpert in groups of children suffering from different disorders, including juvenile idiopathic arthritis (84), in severely disabled children (85) and, as previously noted, children with short stature (76), precocious puberty (77) and congenital adrenal hyperplasia (78). All these studies conclude that BoneXpert is a suitable tool to perform BAA, it is faster than traditional methods and eliminates rater variability. However, it should be noted that one of the authors of most of these studies is a person connected to the commercial activity of Visiana company, the producer of BoneXpert.
BoneXpert has several critical limitations. BA is not identified directly, the prediction depends on the relationship between CA, which is an input to the system, and BA (62). The system is brittle and will reject radiographs when there is excessive noise, in one study it rejected 4.5% of individual bones (81). Finally, until recently BoneXpert did not take the carpal bones into consideration, although in younger children they contain discriminative features. This has been changed in the latest version - BoneXpert 3.0 released in September 2019 - which now does include carpal bones in the analysis.

An additional feature that BoneXpert offers is measurement of a parameter called the Bone Health Index (BHI) (86), which is a unique parameter. BHI is a measurement of bone mass counted as a function of cortical thickness of three central metacarpals and their width and length. The program also automatically calculates standard deviation (SD) values for BHI, based on cohort data of Caucasian children (86). There are several research studies on the comparison of BHI values and traditional methods of bone mass measurement. In one study BHI was compared to dual-energy X-ray-absorption (DXA) and peripheral quantitative computed tomography (pQCT) in a cohort of paediatric patients from paediatric endocrine or paediatric oncology outpatient clinics and it was concluded that BHI values showed a strong positive correlation with DXA readings and total bone mineral density, as assessed via pQCT, also positively correlated with the BHI (87,88). In another study on a group of patients with juvenile idiopathic arthritis, BHI measured by BoneXpert was correlated to measurements of bone mineral density by DXA, however, the correlation of Z-scores of bone mineral density measured by the two methods was weaker (89). The authors of these studies noted that a significant advantage of using BHI, in comparison to DXA or pQCT, was that radiation exposure was lower and in low-risk peripheral areas. Also, BHI has already been used in research studies of BA in patients with juvenile idiopathic arthritis (89). There is an extension to BoneXpert, known as digital X-ray radiogrammetry (DXR). DXR measures the cortical bone thickness in the shafts of the metacarpals and has been shown to be effective in the assessment of hand bone loss caused by rheumatoid arthritis (90).

Another advantage of BoneXpert is a prediction of the final height of a child (91,92), which is a vital element of clinical assessment of a child with short stature. Methods in current routine use take into consideration BAA using traditional methods – GP or TW. The variability of these assessments is the main reason for the variability of predicted final height. When BAA derived from BoneXpert is used, it is possible to predict final height in an objective, precise way. This program takes into consideration sex, CA, height and BA of a child in order to predict their final height. One can also add the height of parents and height at menarche to obtain even more reliable outcome. It is also compulsory to classify the child into one of nine population groups, five within the Caucasian ethnicity, Asian Chinese, Asian American, Hispanic and African American. The result of these calculations is accompanied by an SD value and the true height values will be within the indicated range with 68% probability (93). This method’s accuracy has been validated in a clinical study (91).

### Table 4. Studies assessing the validity of BoneXpert vs. the Greulich-Pyle method

| Study | Author | Year | Size | Origin | Health status | Validity claimed |
|-------|--------|------|------|--------|---------------|-----------------|
|       | Van Rijn et al (75) | 2009 | 405  | Netherlands | Healthy | Yes |
|       | Martin et al (76) | 2008 | 1097 | Germany | Short stature | Yes |
|       | Martin et al (77) | 2011 | 116  | Germany | Precocious or early puberty | Yes |
|       | Martin et al (78) | 2013 | 100  | Germany | Congenital adrenal hyperplasia | Yes |
|       | Booz et al (79) | 2020 | 514  | Germany | Various endocrinological disturbances | Yes |
|       | Thodberg and Sävendahl (22) | 2010 | 1100 | American (4 ethnic groups) | Healthy | Yes |
|       | Artioli et al (80) | 2019 | 515  | Brasil | Healthy, overweight and obese | Yes |
|       | Zhang et al (81) | 2016 | 397  | Shanghai | Healthy | Yes |
|       | Zhang et al (82) | 2013 | 6026 | China | Healthy | Yes |
|       | Martin et al (83) | 2010 | 44  | Japan | Healthy, deficiency of growth hormone | Yes |
|       | Anink et al (84) | 2014 | 69  | Netherlands | Juvenile idiopathic arthritis | Yes |
|       | Mergler et al (85) | 2016 | 95  | Netherlands | Severely disabled | Yes |
Artificial Intelligence and Machine Learning

New possibilities of automating BAA emerged with the use of artificial intelligence (AI) and machine learning, especially the specific type of machine learning known as deep learning. The most popular use a convolutional neural network (CNN), which has already found application in areas such as detection of patterns of interstitial lung disease on CT imaging (94) or segmenting the vascular network of the human eyes on fundus photographs (95). In recent years there has been tremendous progress in this field and there have been numerous publications reporting the automation of BAA using CNN (96-108).

In 2017 Radiological Society of North America (RSNA) conducted a challenge to assess BA from paediatric hand radiographs (RSNA Pediatric Bone Age Machine Learning Challenge 2017), as part of efforts to spur the creation of AI tools for radiology (109,110). The goal of the RSNA 2017 Machine Learning Challenge was to develop an algorithm which can most accurately determine BA using a validation set of paediatric hand radiographs. The results were evaluated by determining the mean difference and the mean absolute difference (MAD) between the performance of each system and the mean of all reviewers’ estimates. The company 16 Bit were placed first in the competition with a MAD of 4.265 months and concordance correlation coefficient of 0.991 (111). The training data set available for competitors contained 12612 images from two American hospitals with a minimum age of 1 month, maximum age of 19 years and mean (SD) age of 10 years and 7 months (3 years 6 months) (111). Their Paediatric Bone Age Calculator is freely available on the website 16Bit.ai, although it is provided with the rider that the application is strictly for demonstration purposes and should not be used for clinical decision making (111). However, this tool has already been validated by a group of Canadian researchers, who compared its results to BAA using the GP atlas in a group of 213 male and 213 female patients and found that the differences between BA assessed by these two methods was not statistically significant (median difference was 0.53 years) and concluded that the tool created by 16 Bit is suitable for clinical use (112).

Another attempt to automate BAA using CNN was described in 2016 by Spampinato et al (113). They compared performance of several approaches, ranging from existing, off-the-shelf CNN, through existing pre-trained CNN (with general imagery) and fine-tuned programs to custom, trained from scratch only on BA radiographs (113). All of these CNNs were tested on the same, public data set, the Digital Hand Atlas Database System, provided in 2007 by Gertych et al (114). This atlas includes 1391 digitized, left-hand radiographs from evenly distributed, normally developed children of Caucasian, Asian, African-American and Hispanic origin, both male and female, with an age range from 1 to 18 years. Spampinato et al (113) conclude that the best performance was observed with BoNet, which was an original, new CNN trained from scratch specifically to assess hand radiographs (114).

Another study in this area deserving attention, as it is especially thorough and methods used have been precisely described, concerns a system called the Fully Automated Deep Learning System for BAA, which was created in 2017 by a group of researchers from Massachusetts General Hospital, Harvard Medical School. They used a pre-trained, fine-tuned CNN to create a new tool for BAA, using a large number of hand radiographs that included 4278 for females and 4047 for males but excluded children aged 0-4 years (115). This system calculates BA and provides a result as a number with representative picture and presents four more pictures of BA +1, +2, -1, -2 years. Thus the radiologist can verify the result and compare it with the closest ones. It achieved an accuracy of 57.32% and 61.4% for the female and male cohorts on held-out test images. Female test radiographs were assigned a BAA within 1 year 90.59% of the time and within 2 years 98.11% of the time. Male test radiographs were assigned 94.18% within 1 year and 99.00% within 2 years. It should be noted that this system does not reject malformed images (115). These authors also compared the BAA performance of a cohort of paediatric radiologists with and without the assistance of their tool for automatic BAA (116). They concluded that AI improves the radiologist’s performance for BAA by increasing accuracy and decreasing variability and root mean squared error. The best results were achieved when radiological assessment was assisted by AI and this was better than using AI alone, a radiologist alone, or a pooled cohort of experts (116).

A comparison of chosen AI methods and BoneXpert is presented in Table 5. Due to the small number of radiographs in training and validating data sets, all the systems based on CNNs used data augmentation (increasing the number of radiographs by rotating the pictures, adding noise, etc.). In some studies authors tested more than one type of CNN. In these studies the CNN with the best performance is presented in the table.

Conclusion

For clinicians, especially paediatric endocrinologists, it is very important to assess BA as precisely as possible to be able to make the right diagnosis and monitor closely the development of a child, the progress of a disease or effects
of treatment. The traditional methods used to date have very significant drawbacks. These drawbacks include being highly time consuming, having a high inter- and intra-rater variability, making comparison of chronologically sequential examinations of one patient difficult and the need to possess a physical copy of the atlas. The new automated BAA techniques provide instant results, eliminate inter- and intra-rater variability and all only need access to the software. Much research in this field is currently underway and the results are very promising. Most of the programs described herein have been validated in clinical studies, in comparison to traditional BAA and they show very good precision while possessing the benefits of automated BAA systems. There are already some widely available options for clinical use, including BoneXpert and the Paediatric Bone Age Calculator from 16Bit.ai. It is to be expected that these automated tools will continue to gain acceptability and widespread usage, making the traditional atlas-based BAA a thing of the past.

Ethics

Peer-review: Externally peer-reviewed.

Authorship Contributions

Concept: Elżbieta Jurkiewicz, Mieczysław Szalecki, Elżbieta Moszczyńska, Monika Prokop-Piotrkowska, Design: Elżbieta Jurkiewicz, Mieczysław Szalecki, Elżbieta Moszczyńska, Monika Prokop-Piotrkowska, Literature Search: Monika Prokop-Piotrkowska, Kamila Marszałek-Dziuba, Writing: Monika Prokop-Piotrkowska, Kamila Marszałek-Dziuba.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Reinehr T, de Sousa G, Wabitsch M. Relationships of IGF-I and androgens to skeletal maturation in obese children and adolescents. J Pediatr Endocrinol Metab 2006;19:1133-1140.
2. Phillip M, Moran O, Lazar L. Growth without growth hormone. J Pediatr Endocrinol Metab 2002;15(Suppl 5):1267-1272.
3. Cox LA. The biology of bone maturation and ageing. Acta Paediatr Suppl 1997;423:107-108.
4. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. Hum Biol 1973;45:211-227.
5. Sehrawat JS, Singh M. Willems method of dental age estimation in children: A systematic review and meta-analysis. J Forensic Leg Med 2017;52:122-129. Epub 2017 Aug 25
6. Martin DD, Wit JM, Hochberg Z, Sävendahl L, van Rijn RR, Fricke O, Cameron N, Caliebe J, Hertel T, Kiepe D, Albertsson-Wikland K, Thodberg HH, Binder G, Ranke MB. The use of bone age in clinical practice - part 1. Horm Res Paediatr 2011;76:1-9. Epub 2011 Jun 21
7. Spadoni GL, Cianfarani S. Bone age assessment in the workup of children with endocrine disorders. Horm Res Paediatr 2010;73:2-5.
8. Kim SE, Jang JW, Ahn MB, Kim SH, Cho WK, Cho KS, Park SH, Jung MH, Suh BK. The association between skeletal maturation and adrenal androgen levels in obese children and adolescents. Ann Pediatr Endocrinol Metab 2017;22:108-114. Epub 2017 Jun 28
9. Menjívar C, Perreira KM. Undocumented and unaccompanied: children of migration in the European Union and the United States. J Ethn Migr Stud 2019;45:197-217. Epub 2017 Dec 21
10. Kaur G, Khandelwal N, Jasuja OP. Computed tomographic studies on ossification status of medial epiphysis of clavicle: Effect of slice thickness and dose distribution. J Indian Acad Forensic Med 2010;32:298-302.
49. Hojreh A, Gamper J, Schmook MT, Weber M, Prayer D, Herold CJ, Noebauer-Huhmann JM. Hand MRI and the Greulich-Pyle atlas in skeletal age estimation in adolescents. Skeletal Radiol 2018;47:963-971. Epub 2018 Jan 25

50. Ebner T, Stern D, Donner R, Bischof H, Urschler M. Towards automatic bone age estimation from MRI: localization of 3D anatomical landmarks. Med Image Comput Assist Interv 2014;17:421-428.

51. Stern D, Payer C, Urschler M. Automated age estimation from MRI volumes of the hand. Med Image Anal 2019;58:101538. Epub 2019 Jul 31

52. 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.

53. 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.

54. Garn SM, Hertzog KP, Poznanski AK, Nagy JM. Metacarpophalangeal length in the evaluation of skeletal malformation. Radiology 1972;105:375-8.

55. Tanner JM, Gibbons RD. A computerized image analysis system for estimating Tanner-Whitehouse 2 bone age. Horm Res 1994;42:282-287.

56. Tanner JM, Oshman D, Lindgren G, Grunbaum 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.

57. Frisch H, Riedl S, Waldhör T. Computer-aided estimation of skeletal age and comparison with bone age evaluations by the method of Greulich-Pyle and Tanner-Whitehouse. Pediatr Radiol 1996;26:226-231.

58. Van Teunenbroek A, De Waal W, Roks A, Chinof P, Fokker M, Mulder P. De Muinck Keizer-Schrama S, Drop S. Computer-aided skeletal age scores in healthy children, girls with Turner syndrome, and in children with constitutionally tall stature. Pediatr Res 1996;39:360-367.

59. Hill K, Pynsent PB. A fully automated bone-ageing system. Acta Paediatr Suppl 1994;406:81-83.

60. Rucci M, Coppini G, Nicolletti I, Chelli D, Valli G. Automatic analysis of digital radiographs for the assessment of skeletal age: a subsymbolic approach. Comput Biomed Res 1995;28:239-256.

61. Somkantha K, Theera-Umporn N, Aejuanwiriyakul S. Bone age assessment in young children using automatic carpal bone feature extraction and support vector regression. J Digital Imaging 2011;24:1044-1058.

62. Seok J, Hyun B, Kasa-Vubu J, Girard A. Automated Classification System for Bone Age X-ray Images. IEEE Int Conf Syst Man Cybern 2012;208-213.

63. Cao F, Huang HK, Pietka E, Gilsanz V. Digital hand atlas and web-based bone age assessment: system design and implementation. Comput Med Imaging Graph 2000;24:297-307.

64. Gross GW, Boone JM, Bishop DM. Pediatric skeletal age: determination with neural networks. Radiology 1995;195:689-695.

65. Sato K, Ashizawa K, Anzo M, Otsuki F, Kaneko S, Tanaka T, Tsukagoshi K, Nimura A, Matsuoka H. Assessment with neural networks. Radiology 1995;195:689-695.

66. Mahmoodi S, Sharif BS, Chester EG, Owen JP, Lee R. Skeletal growth estimation using radiographic image processing and analysis. IEEE Trans Inf Technol Biomed 2000;4:292-297.

67. Zhang A, Gertych A, J LB. Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones. Comput Med Imaging Graph 2007;31:299-310.

68. Hsieh CW, Jong TL, Tiu CM. Bone age estimation based on phalanx information with fuzzy constrain of carpals. Med Biol Eng Comput 2007;45:283-295. Epub 2007 Jan 23

69. Liu J, Qi J, Liu Z, Ning Q, Luo X. Automatic bone age assessment based on intelligent algorithms and comparison with TW3 method. Comput Med Imaging Graph 2008;32:678-684.

70. Trisant-Vega A, Arribas JJ. A radius and ulna TW3 bone age assessment system. IEEE Trans Biomed Eng 2008;55:1463-1476.

71. Hsieh CW, Liu TC, Jong TL, Tiu CM. A fuzzy-based growth model with principle component analysis selection for carpal bone-age assessment. Med Biol Eng Comput 2010;48:579-588. Epub 2010 Apr 20

72. Thodberg HH, Kreiborg S, Juel A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging 2009;28:52-66.

73. Thodberg HH. An Automated Method for Determination of Bone Age. J Clin Endocrinol Metab 2009;94:2239-2244.

74. Thodberg HH, van Rijn RR, Jenni OG, Martin DD. Automated determination of bone age from hand X-rays at the end of puberty and its applicability for age estimation. Int J Legal Med 2017;131:771-780. Epub 2016 Oct 18

75. van Rijn RR, Lequin MH, Thodberg HH. Automatic determination of Greulich and Pyle bone age in healthy Dutch children. Pediatr Radiol 2009;39:591-597. Epub 2009 Jan 6

76. Martin DD, Deusch D, Schweizer R, Binder G, Thodberg HH, Ranke MB. Clinical application of automated Greulich-Pyle bone age determination in children with short stature. Pediatr Radiol 2009;39:598-607. Epub 2009 Mar 31

77. Martin DD, Meister K, Schweizer R, Ranke MB, Thodberg HH, Binder G. Validation of automatic bone age rating in children with precocious and early puberty. J Pediatr Endocrinol Metab 2011;24:1009-1014.

78. Martin DD, Heil K, Heckmann C, Zierf A, Schaefer J, Ranke MB, Binder G. Validation of automatic bone age determination in children with congenital adrenal hyperplasia. Pediatr Radiol 2013;43:1615-1621. Epub 2013 Oct 5

79. Booz C, Yel I, Wichmann JL, Boetgter S, Al Kamali A, Albrecht MH, Martin SS, Lenga L, Huizinga NA, D'Angelo T, Cavallaro M, Vogl TJ, Bodele B. Artificial intelligence in bone age assessment: accuracy and efficiency of a novel fully automated algorithm compared to the Greulich-Pyle method. Eur Radiol Exp 2020;28:4-6.

80. Artioli TO, Alvares MA, Carvalho Macedo VS, Silva TS, Avrichir R, Kochi C, Longui CA. Bone age determination in eutrophic, overweight and obese Brazilian children and adolescents: a comparison between computerized BoneXpert and Greulich-Pyle methods. Pediatr Radiol 2019;49:1185-1191. Epub 2019 May 31

81. Zhang J, Lin F, Ding X. Maturation disparity between hand-wrist bones in a chinese sample of normal children: An analysis based on automatic bonexpert and manual greulich and pyle atlas assessment. Korean J Radiol 2016;17:435-442. Epub 2016 Apr 14

82. Zhang SY, Liu G, Ma CG, Han YS, Shen XZ, Xp RL, Thodberg HH. Automated determination of bone age in a modern chinese population. ISRN Radiol 2013;2013:874570.

83. Martin DD, Sato K, Sato M, Thodberg HH, Tanaka T. Validation of a new method for automated determination of bone age in Japanese children. Horm Res Paediatr 2010;73:398-404. Epub 2010 Apr 14

84. Anikir J, Nusman CM, van Suijlekom-Smit LW, van Rijn RR, Maas M, van Rossum MA. Automated determination of bone age and bone mineral density in patients with juvenile idiopathic arthritis: a feasibility study. Arthritis Res Ther 2014;16:424.

85. Mergler S, de Man SA, Boot AM, Heus KGCB B, Huijbers WAR, van Rijn RR, Penning C, Evenhuis HM. Automated radiogrammetry is a
feasible method for measuring bone quality and bone maturation in severely disabled children. Pediatr Radiol 2016;46:1017-1022. Epub 2016 Mar 30.

86. Thodberg HH, van Rijn RR, Tanaka T, Martin DD, Kreiborg S. A paediatric bone index derived by automated radiogrammetry. Osteoporos Int 2010;21:1391-1400. Epub 2009 Nov 24.

87. Schündeln MM, Marschke L, Bauer JJ, Hauffa PK, Schweiger B, Führer-Sakel D, Lahner H, Poeppl TD, Kiewert C, Hauffa BP, Grasemann C. A Piece of the Puzzle: The Bone Health Index of the BoneXpert Software Reflects Cortical Bone Mineral Density in Pediatric and Adolescent Patients. PLoS One 2016;11:e0151936.

88. Nusman CM, Anink J, Van Rossum MAJ, Van Rijn RR, Maas M, Van Suijlekom-Smit IWA. Bone health assessment of patients with juvenile idiopathic arthritis: A comparison between DXA and BoneXpert. Pediatr Radiol 2014;44(Suppl 1):S521-S522.

89. Twilt M, Pradsgaard D, Spannow AH, Horlyck A, Heuck C, Herlin T. Joint cartilage thickness and automated determination of bone age and joint health in juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2017;15:63.

90. Pfeil A, Thodberg HH, Renz DM, Reinhardt L, Oelzner P, Wolf G, Böttcher J. Metacarpal bone loss in patients with rheumatoid arthritis estimated by a new Digital X-ray Radiogrammetry method - Initial results. BMC Musculoskelet Disord 2017;18:1-10.

91. Thodberg HH, Jenni OG, Caflisch J, Ranke MB, Martin DD. Prediction of adult height based on automated determination of bone age. J Clin Endocrinol Metab 2009;94:4868-4874. Epub 2009 Nov 19.

92. Satoh M. Bone age: assessment methods and clinical applications. Clin Pediatr Endocrinol 2015;24:143-152. Epub 2015 Oct 24.

93. Adult Height Predictor. Accessed on: 29.07.2021. Available from: https://wwwpesanetwork.com/adult-height-predictor

94. Anthimopoulos M, Christodoulidis S, Ebner L, Chrissie A, Mougiakakou S. Lung Pattern Classification for Intestinal Lung Diseases Using a Deep Convolutional Neural Network. IEEE Trans Med Imaging 2016;35:1207-1216. Epub 2016 Feb 29.

95. Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. IEEE Trans Med Imaging 2016;35:2369-2380. Epub 2016 Mar 24.

96. Lee JH, Kim KG. Applying deep learning in medical images: the case of bone age estimation. Healthe Inform Res 2018;24:86-92. Epub 2018 Jan 31.

97. Hao PY, Chokuwa S, Xie XH, Wu FL, Wu J, Bai C. Skeletal bone age assessments for young children based on regression convolutional neural networks. Math Biosci Eng 2019;16:6454-6466.

98. Duc T, Lee J, Shin J. Incorporated region detection and classification using deep convolutional networks for bone age assessment. Artif Intell Med 2019;97:1-8.

99. Wang F, Gu X, Chen S, Liu Y, Shen Q, Pan H, Shi L, Jin Z. Artificial intelligence system can achieve comparable results to experts for bone age assessment of Chinese children with abnormal growth and development. PeerJ. 2020;8:e8854.

100. Wang F, Gu X, Chen S, Liu Y, Shen Q, Pan H, Shi L, Jin Z. Artificial intelligence system can achieve comparable results to experts for bone age assessment of Chinese children with abnormal growth and development. PeerJ. 2020;8:e8854.

101. Mutasa S, Chang PD, Ruzal-Shapiro C, Ayayla R. MABAL: a Novel deep-learning architecture for machine-assisted bone age labeling. J Digit Imaging 2018;31:513-519.

102. Liu Y, Zhang C, Cheng J, Chen X, Wang ZJ. A multi-scale data fusion framework for bone age assessment with convolutional neural networks. Comput Biol Med 2019;108:161-173.

103. Ren X, Li T, Yang X, Wang S, Ahmad S, Xiang L, Stone SR, Li L, Zhan Y, Shen D, Wang Q. Regression Convolutional Neural Network for Automated Pediatric Bone Age Assessment From Hand Radiograph. IEEE J Biomed Health Inform 2019;23:2030-2038. Epub 2018 Oct 19.

104. 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. Epub 2016 Oct 29.

105. Tong C, Liang B, Li J, Zheng Z. A Deep Automated Skeletal Bone Age Assessment Model with Heterogeneous Features Learning. J Med Syst 2018;42:249.

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

107. Van Steenkiste T, Ruysinck J, Janssens O, Vandewalle G.骨質密度の測定と骨生长期の推定。Ann Rheum Dis 2013;72:1374-1380.

108. Larson DB, Chen MC, Lungren MP, Stence NV, Langlois CP. Performance of a deep-learning neural network model in assessing skeletal maturity on. Radiology 2018;287:313-322.

109. Halabi SS, Prevedello LM, Kalpathy-cramer J, Mamonov AB. The RSNA pediatric bone age machine learning Challenge. Radiology 2018;290:498-503.

110. Siegel EL. What Can We Learn from the RSNA Pediatric Bone Age Machine Learning Challenge? Radiology 2018;290:504-505.

111. Predicting Skeletal Age. Available from: https://www.16bit.ai/bone-age

112. Larson DB, Chen MC, Lungren MP, Stence NV, Langlois CP. Performance of a deep-learning neural network model in assessing skeletal maturity on. Radiology 2018;287:313-322.

113. Gerges M, Eng H, Chihina H, Cooper A. Modernization of bone age assessment: comparing the accuracy and reliability of an artificial intelligence algorithm and shorthand bone age to Greulich and Pyle. Skeletal Radiol 2020;49:1449-1457.

114. 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.

115. Gertych A, Zhang A, Sayre J, Hspsieh-Kurkowska S, Huang HK. Bone Age Assessment of Children using a Digital Hand Atlas. Comput Med Imaging Graph 2007;31:322-331.

116. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Choy G, Do S. Fully Automated Deep Learning System for Bone Age Assessment. J Digit Imaging 2017;30:427-441.

117. Tajmir SH, Lee H, Shaham R, Gale HI, Nguyen JC, Westra SJ, Lim R, Yune S, Gee MS, Do S. Artificial Intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. Skeletal Radiol 2019;48:275-283. Epub 2018 Aug 1.