

Review

# Artificial Intelligence (AI)-Based Systems for Automatic Skeletal Maturity Assessment through Bone and Teeth Analysis: A Revolution in the Radiological Workflow?

Elena Caloro <sup>1</sup>, Maurizio Cè <sup>1</sup>, Daniele Gibelli <sup>2</sup>, Andrea Palamenghi <sup>2</sup>, Carlo Martinenghi <sup>3</sup>, Giancarlo Oliva <sup>4</sup> and Michaela Cellina <sup>4,\*</sup>

<sup>1</sup> Postgraduation School in Radiodiagnostics, Università degli Studi di Milano, Via Festa del Perdono, 7, 20122 Milan, Italy

<sup>2</sup> Dipartimento di Scienze Biomediche per la Salute, Via Luigi Mangiagalli 31, 20133 Milan, Italy

<sup>3</sup> Radiology Department, San Raffaele Hospital, Via Olgettina 60, 20132 Milan, Italy

<sup>4</sup> Radiology Department, Fatebenefratelli Hospital, ASST Fatebenefratelli Sacco, Piazza Principessa Clotilde 3, 20121 Milan, Italy

\* Correspondence: michaela.cellina@asst-fbf-sacco.it

**Abstract:** Bone age is an indicator of bone maturity and is useful for the treatment of different pediatric conditions as well as for legal issues. Bone age can be assessed by the analysis of different skeletal segments and teeth and through several methods; however, traditional bone age assessment is a complicated and time-consuming process, prone to inter- and intra-observer variability. There is a high demand for fully automated systems, but creating an accurate and reliable solution has proven difficult. Deep learning technology, machine learning, and Convolutional Neural Networks-based systems, which are rapidly evolving, have shown promising results in automated bone age assessment. We provide the background of bone age estimation, its usefulness and traditional methods of assessment, and review the currently artificial-intelligence-based solutions for bone age assessment and the future perspectives of these applications.



**Citation:** Caloro, E.; Cè, M.; Gibelli, D.; Palamenghi, A.; Martinenghi, C.; Oliva, G.; Cellina, M. Artificial Intelligence (AI)-Based Systems for Automatic Skeletal Maturity Assessment through Bone and Teeth Analysis: A Revolution in the Radiological Workflow? *Appl. Sci.* **2023**, *13*, 3860. <https://doi.org/10.3390/app13063860>

Academic Editors: Peng Zhou and Chao Yang

Received: 19 January 2023

Revised: 12 March 2023

Accepted: 13 March 2023

Published: 17 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** bone age assessment; artificial intelligence; machine learning; computer-aided detection; pediatric radiology

## 1. Introduction

Bone age is a sign of an individual's skeletal and biological maturity, distinct from chronological age, which is determined by a person's date of birth [1].

Bone age estimation has different clinical roles: first, it is critical for assessing pediatric growth and maturity, particularly in a variety of syndromic and endocrine disorders associated with short stature, commonly associated with delayed bone age; second, it helps to establish the progression and treatment response of different pediatric endocrine and metabolic diseases; third, it can also be applied to predict adult height [2–5]. Moreover, the recently raised prevalence of precocious puberty [6] has increased the interest in children's development and the administration of growth hormone therapy, enhancing the value of bone age evaluation [7].

Estimating bone age is a challenge with significant social implications, as in South Asia, where 65% of all births are not registered by the age of 5 years [1]. It also has legal implications, as many countries are concerned about the growing number of children and adolescents who lack valid proof of their chronological age, particularly in cases of illegal immigration, delinquency, and juvenile labor [8]. Bone age can be traditionally estimated by different imaging techniques: the most applied are atlas-based such as the Greulich-Pyle (GP) [9] and Gilsanz-Ratibin (GR) [10]. The atlas-based methods are simple and fast, but several studies have suggested possible limitations, such as the high inter- and intra-observer variability and the dependency on the clinician's experience [11–13].

Moreover, ethnic variability is often undervalued by operators, as atlases may be efficient in age estimation only in specific ethnic groups [14,15].

The Tanner and Whitehouse method is a scoring method that derives bone age from the sum of the scores obtained by analyzing various regions of interest in specific bones of the left hand [16]. Unfortunately, this technique also has some limitations: it is a quite complex approach, requiring a significant amount of time and providing a sometimes-ambiguous classification because one bone shape can have two different pre-defined labels of the same feature [17,18].

Other methods assess the ossification stage of the medial clavicular epiphysis [19], or of the cervical vertebrae [20], whereas the Sauvegrain method evaluates the ossification of four anatomical elbow areas: the trochlea, lateral condyle, proximal radial epiphysis, and olecranon apophysis [21,22].

Each of these methods suffers from significant limitations, mainly related to the high inter-observer variability, operator dependence, and time required for correctly performing the evaluation.

The assessment of bone age is a representative example of how the concept of object detection and classification can be used, therefore, it is the perfect field for the application of artificial intelligence (AI) algorithms to develop automated systems of age classification.

Due to the limitations of traditional methods, the demand for an automated assessment method has always existed, and, to the best of our knowledge, in 1989, the first automated tool—HANDX—was created [23]. HANDX is a histogram model based on preprocessing, segmentation, and measurement that may be considered a first attempt for the automation of the task without the support of deep learning (DL) and machine learning (ML) tools.

Following initial attempts, some researchers have taken on the challenge of developing reliable tools for determining bone age quickly and precisely, thereby expediting the workflow. Unfortunately, despite numerous proposals, only a few systems have been commercialized. In this narrative review, we aim to describe the traditional methods used for bone age determination, including teeth analysis, and the different AI tools developed for this task to highlight the benefits and limitations of every approach to update radiologists on this important topic. Our hope is that the future purposes will overcome the limitations to be truly helpful in everyday clinical practice.

### *Introduction to AI Basic Terminology*

AI is a catch-all term for the capacity of a machine to mimic human abilities, such learning, reasoning, and problem-solving [24]. While human critical and cognitive abilities still appear irreplaceable, machines exhibit an additional unmatched advantage: the capacity to analyze enormous amounts of data in a very short amount of time [25].

The radiomics paradigm, also known as quantitative imaging, is supported by AI-based image analysis tools and entails viewing biomedical images as sets of numerical data rather than as simple images [26,27].

Machine learning (ML) is a subfield of AI, aiming at the development and tuning of learning algorithms that exhibit human-like learning properties for data analysis but are superior in performance, robustness, and velocity.

Different ML approaches use a large set of models to predict a quantity of interest (or higher-level parameters) that satisfies some predefined requirements, starting from a sample dataset [25,28,29]. By feeding these algorithms with enough data, they can extract knowledge from the data. After the training phase, the ML model should be validated and tested for its reliability in a different dataset: this is, in simple terms, the so-called radiomic workflow [30].

ML includes four main approaches: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning; each approach involves different levels of data pre-treatment, algorithmic strategies applied to map the relationship between data, and the problems that can be solved. The first three methods are the most widely applied in radiology [31].

The simplest type of ML, supervised learning, is suitable for very general classification tasks. To be trained, this type of ML requires a data set in which each input (independent variable) is labeled according to the corresponding output [32,33]. These algorithms gradually tune their performance while increasing exposure to samples during the training phase, thus learning to infer the relationship between data according to a mapping hypothesis (a model). During testing, the model associates one of the pre-defined class labels with fresh, unlabeled targets. When the actual labels are compared to the ones that were assigned, the algorithm's performance can be evaluated.

Supervised learning is the basis of many AI applications to predict the risk of cancer, survival, nodule detection, and classification [34].

Unsupervised learning algorithms require the machine to discover a hidden structure within data, or at least a subset of data, where no labels are provided. Unsupervised learning can be applied to a variety of data mining (extracting information from data) challenges, including clustering tasks (which aim to divide the dataset into groups based on particular feature characteristics) and association tasks (which aim to identify association rules within the dataset) [28].

Between supervised and unsupervised learning is semi-supervised learning. This method primarily uses unlabeled data, with a small amount of labeled data also included. As a result, this type of ML addresses the issue of low data availability by using a large amount of readily available but untagged data (such as undiagnosed images) to train precise classifiers [35].

Potential relationships between or within data can be inferred using a wide range of models, each with its applicable design for data representation, analysis, and loss function minimization (a function that describes the approximation toward the desired performance) [36]. In very general terms, models may be statistically based, or artificial-neural-networks-based; both models are required to be computationally feasible (that could be translated into a programming code for being automated by a machine) [37]. Each model can be viewed as a different tactic when thinking about the various ML approaches (such as supervised learning) as different strategies to find significant patterns in data.

In supervised learning, for example, these models represent various hypotheses for mapping the connection between input and output; the difference between the prediction and the observed data essentially determines the fitness of the hypothesis. Logistic regression, decision trees, and support vector machines are a few examples of statistical models used in supervised learning. While various strategies may apply to the same task, some are more effective than others in particular situations.

Artificial neural networks (ANN) became a popular model for dealing with data that could be considered a learning paradigm to address biomedical imaging tasks. As opposed to purely statistical models (such as logistic regression), artificial neural networks are inspired by biological neural networks of the human brain. ANNs learning properties are based on the computational properties of the neural ensemble, whose global function is the result of the coordinated activity of many smaller units, each of which performs an elementary computational operation, substantially summing the inputs—which are weighted differently—and transferring the information through the neuron when a certain threshold value is reached. Convolutional Neural Networks (CNNs) are a type of particularly complex ANN that has paved the way for DL in biomedical imaging [38–41]. DL could be defined as the high-throughput extraction of information from a huge amount of data using artificial neural networks, in which multiple layers of processing are used to extract progressively higher-level higher features. DL networks have been developed in recent years as a result of the availability of larger datasets and faster computation speeds. In a DL network, each layer represents a higher level of abstraction, with the number of layers determining the model depth.

Nowadays, CNNs are the most widely utilized artificial neural networks for issues involving radiological image processing and recognition and are inspired by the mammalian

visual cortex [17]. In CNN, the systems consist of various sub-networks hierarchically embedded into one another.

DL models can be used to find complex patterns in large data sets that go beyond the features that a radiologist could extract, such as tumor segmentation and feature extraction. Still, they do so by taking the entire image as an input. DL algorithms are well suited for imaging recognition tasks such as pattern detection and precise lesion segmentation [42].

## 2. Left Hand and Wrist Bone Age Assessment

### 2.1. Traditional Approaches

The ossification pattern in the hand and wrist bones is quite predictable and age-specific until the end of adolescence, when bone elongation is complete. Thus, the most widely used methods for estimating bone age are based on a comparison of the patient's level of maturation of hand and wrist bones to normal age standard levels. The radiograph of the left-hand wrist is the most used imaging technique for bone age assessment because it is fast and safe, widely available, and can obtain information from many bones included those in a small anatomical region.

Despite the various attempts to avoid radiation using ultrasound (US) or magnetic resonance imaging (MRI), the antero-posterior hand radiograph is still the most popular method because the effective dose of radiation received is between 0.0001–0.1 mSv [1,43]. Moreover, the quick acquisition of a radiograph is well suited to poor patient compliance, while US and MRI require more time and collaboration.

The GP method is based on the comparison of the appearance of a standard left-hand wrist X-ray with the nearest matching reference images provided in the atlas "The Radiographic Atlas of Skeletal Development of the Hand and Wrist" by Dr. William Walter Greulich and Dr. Sarah Idell Pyle (1959) [9].

This atlas contains reference standards of the left hand–left wrist radiography of subjects up to 18 years for females and 19 years for males.

The GP method is simple and relatively fast, but it is based on radiographs of North American Caucasians of good socioeconomic status, therefore, it does not apply to different populations. This method demonstrated good reliability for Australian and Middle Eastern ethnicities, but it was imprecise when applied to Asian, especially males, and to African females [44–47]

Furthermore, recent studies have shown that bone age is advanced compared to chronological age, even in Caucasian populations, most likely because children nowadays mature faster than those reported in the GP Atlas, which was published in 1959 [48].

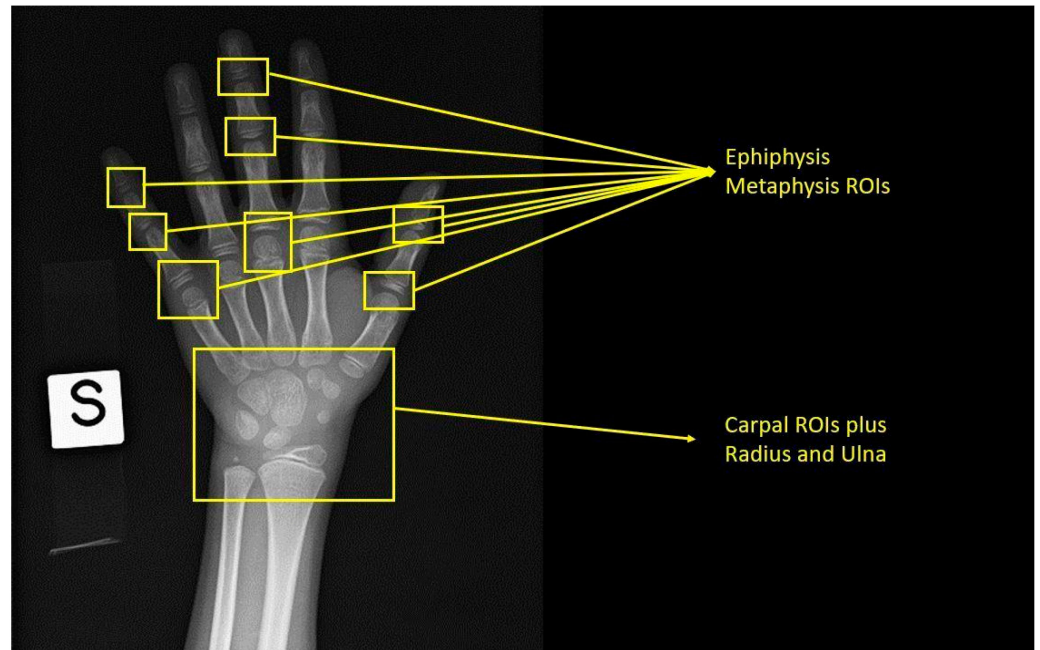
The GR Atlas [10] is a digital atlas, developed in 2005, including artificial and idealized standard hand–wrist radiographs images, specific for age and sex, created considering size, shape, morphology, and density as maturity characteristics of ossification centers in healthy children from 8 months to 18 years of age (at 6-month intervals between 2 and 6 years, and yearly intervals between 7 and 17 years). Compared to the older GP atlas, the new GR atlas provides images that are more precise and of a higher quality [49]. However, it also has more outliers [1] and it is affected by ethnic differences, as shown in a testing application on the Chinese population. According to this study, the GR atlas is not recommended in males aged 10–13 years, while the GP atlas does not fit properly for females aged 0–3 years [50].

The Tanner Whitehouse Method (1983) (TW2) [11] assesses the level of maturity of 20 selected regions of interest in specific bones of the hand and wrist in each age population. The development level for each region of interest is categorized into specific stages (from A to I) corresponding to a numerical score for every single bone. A global sex-specific maturity score is then obtained by summing all the individual scores and converting them into bone age [51].

In addition to the well-known ethical limitations, the TW2 method was affected by the evidence that new-generation children's bones were maturing more rapidly than in the past; therefore, a revised TW3 method was proposed in 2001 to assess the maturity of the

radius, ulna, and short bones (RUS), with a new version of centile charts [11]. Regardless of this evidence, TW2 remains the most known and used method [52].

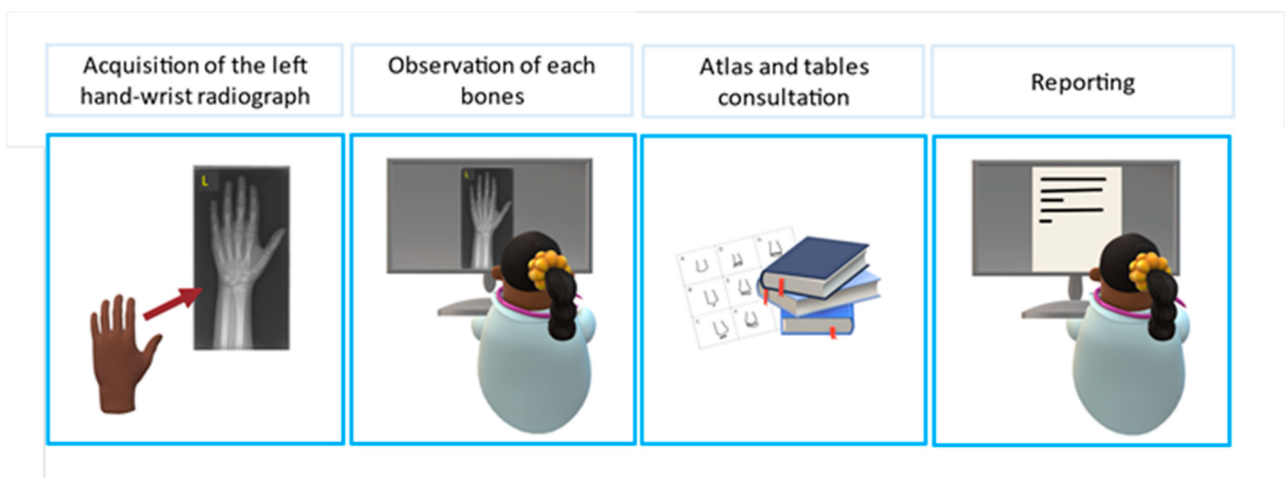
Despite the higher accuracy and reproducibility compared to the atlas-based techniques, the Tanner Whitehouse Method is more complex and time-consuming (Figure 1) and the classification is sometimes ambiguous as a particular bone shape can have two different predefined labels of the same feature [18].



**Figure 1.** Tanner Whitehouse Method (TW2) with a demonstration of the region of interest that should be placed for age calculation.

### 2.2. AI-Based Approaches

Traditional methods for determining bone age have undeniable limitations due to generational and ethnic differences. Furthermore, most radiologists regard the process as time-consuming and irritating, and the result is prone to inter- and intra-observer variability (Figure 2). Automated assessment may result in a quick, objective, and reproducible age calculation.



**Figure 2.** The steps of a traditional time-consuming age assessment procedure.

The first developed tool was a semi-automated system—HANDX—was based on the left hand and wrist X-ray and was introduced in 1989 [23]. In 1995, Gross et al. developed a neural network for the calculation of skeletal age based on measurements from hand X-rays [53].

To stimulate interest in this topic, the Radiological Society of North America (RSNA) launched the Pediatric Bone Age Machine Learning Challenge to create an algorithm or model with ML techniques to obtain an accurate assessment of skeletal age from a dataset of pediatric hand X-rays labeled by different experienced readers. They publicly recognized the team that created the algorithms with the best performances at 2017 RSNA annual meeting [54]. The datasets included 12,611 hand radiographs and a validation set of 1425 X-rays, and a separate test set of 200 images. Metha et al., for example, applied transfer learning to pre-train a neural network architecture obtaining a difference between the actual and estimated age of 5.9 months [55].

Pan et al. investigated potential model assembling benefits to improve the efficiency in automatic bone age estimation, including 48 submissions from the 2017 RSNA Pediatric Bone Age Machine Learning Challenge. The mean absolute deviation (MAD) and the mean pairwise model correlation were used to evaluate the performance of each potential model combination in ensembles of up to ten models. A single model expected generalization MAD was 4.55 months. The quartet that performed the best had a MAD of 3.79 months. Models in this group had an average pairwise correlation of 0.47. When the highest-ranking models based on individual scores were combined, the lowest MAD using eight models and a mean pairwise model correlation of 0.67 was 3.93 months [56]. The authors therefore supported ML contests to encourage the development of heterogeneous models whose predictions can be combined to achieve optimal performance.

Other authors tested the winning DL model of the 2017 RSNA Pediatric Bone Age Challenge on an internal validation set (1425 individuals) and an external test set (1202 individuals) of pediatric hand X-rays, using images reporting as ground truth for bone age. On the external test set, the bone age model performed well, with no change in MAD (6.8 months in the validation set versus 6.9 months in the external set). Model predictions would have resulted in clinically significant errors in 194 of 1202 images (16%) in the external test group. In both the age and Tanner stage subcategories of the external test set, as well as the internal validation set ( $p = 0.01$ ), the MAD was higher for girls than for boys. The authors concluded that, although clinically significant sex-, age-, and sexual-maturity-based biases in DL bone age were discovered, the model generalization to an external test group was good [57].

Kim et al. developed a GP-method-based DL technique to create a system for automated age assessment analyzing left-hand radiographs of 200 patients (3–17 years old) [58]. The reference was represented by the consensus of two experienced radiologists. The software showed a 69.5% concordance rate and significant correlations with the reference bone age ( $p < 0.001$ ), with a reduction of image reading times of 18.0% and 40.0% for reviewers one and two, respectively.

14,036 clinical hand radiographs from two pediatric hospitals were used to train and validate a DL neural network to assess bone age [59]. The mean of bone age estimates from the clinical report and additional three human reviewers were applied as the reference standard. Overall model performance was assessed by comparing the root mean square error (RMSE) and MAD between the model estimates and the reference standard bone ages. The mean difference between bone age estimates of the model and the reviewers was 0 years, with a mean RMSE and MAD of 0.63 and 0.50 years, respectively. Results from the model, the radiological report, and the readers' revisions were within the 95% limits of agreement. Therefore, the proposed model was able to provide bone age estimation with accuracy similar to that of an expert radiologist.

Mutasa et al. developed a customized neural network algorithm trained on 10,289 images of different skeletal age exams (8909 from their Picture Archiving and Communication System and 1383 from the public Digital Hand Atlas Database). Images were divided into

four cohorts: two for boys and girls < 10 years, and two for boys and girls 8–10 years old. The test set included left-hand radiographs taken for bone age assessment and trauma investigation and consisted of 20 X-rays of each 1-year-age cohort from 0 to 1 year to 14–15+ years, 50% of male subjects and 50% of female subjects. A 14-hidden-layer customized neural network was created. Data augmentation was applied to the network inputs [60] and a linear regression output was utilized. On the validation and test sets, MAE accuracies for young females were 0.654 and 0.561, respectively, and 0.662 and 0.497 for older females. For young males, the validation and test accuracies were 0.649 and 0.585, respectively, whereas for older males, they were 0.581 and 0.501, respectively. The customized neural network architecture reached an aggregate validation and test set MAE of 0.637 and 0.536, respectively.

Spampinato et al. tested different DL approaches in automated skeletal bone age assessment on a public dataset including different age ranges, genders, and ethnicities, and demonstrated an average discrepancy between the manual and automatic calculation of 0.8 years, considered as state-of-the-art performance [61].

Lee et al. developed a fully automated DL system to detect bones and tissues, construct a hand/wrist mask, standardize and preprocessed input X-rays, and calculate bone age using X-rays of the left hand and wrist of patients aged 5–18 years and more than 18 years, including 4278 images for females and 4047 for males. In the female and male cohorts, their fine-tuned CNN achieved an accuracy of 57.32% and 61.40%, respectively [62].

To process heterogenous features, Tong et al. developed a deep automated skeletal bone age assessment model based on CNNs and support vectors using the multiple kernels learning algorithm. This was not only aimed at the automated assessment of bone age from X-ray images of the hand and wrist, but also for the detection of heterogeneous information such as race and gender. The results demonstrated that the fused heterogeneous features offer a better description of the degree of bone maturation because they performed better and with higher bone age assessment accuracy [63].

Xu et al. first used a dataset of 2518 left-hand radiographs and applied a fine-grained classification to obtain the region of interest through automatic object detection, then executed the bone age assessment with a model based on the TW3 method: the accuracy of bone grading was 86.93%, with a MAE of bone age of 7.68 months on the clinical dataset [64].

Another proposed approach is based on the analysis of the carpal bones. Using projections in both the horizontal and vertical axes, Somkantha et al. [65] selected the carpal bone region and extracted the boundaries of the carpal bones. From the segmented carpal bones, they extracted five morphological features and applied them to regression using a support vector machine. Zhang et al. proposed a similar approach with extraction from carpal bones of hand-engineered features, then used this as an input for a fuzzy logic classifier [66]. As the carpal bones are typically fully developed by ages 5 to 7 and do not allow for meaningful discrimination after that age, this method could be applied only to younger children [10].

A recent study developed three DL models from a Chinese private dataset, an American public dataset, and a dataset combining the above two datasets, with test data labeled by ten senior pediatric radiologists, with a demonstrated mean absolute deviation of 0.42 years [67]. The authors noted that the result corresponded to good accuracy for the whole group but did not indicate an accurate estimation of individual bone age, as the kappa value was 0.714 and the agreement between the system and human clinical determinations was significantly different. They invited researchers to consider possible biases related to patients' sex and age, institutions, and radiologists.

The main characteristics of the above studies are listed in Table 1.

**Table 1.** Main characteristics of studies based on hand radiograph analysis for bone age estimation.

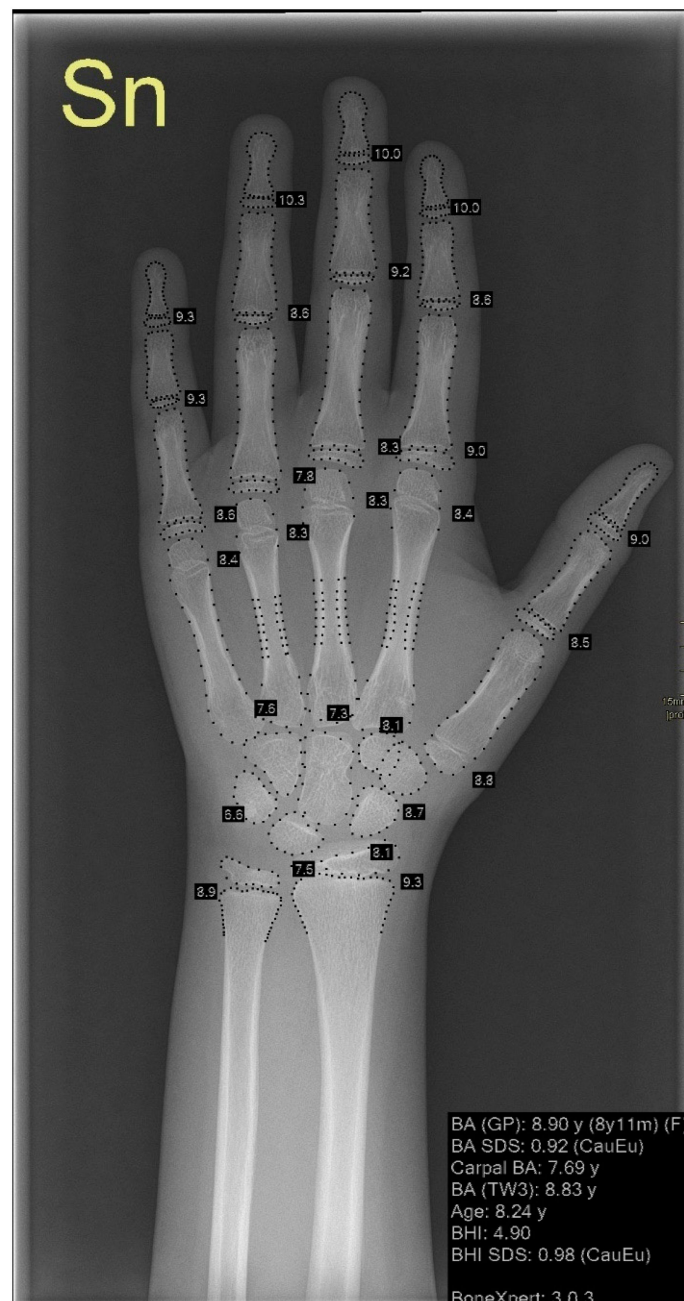
Authors	Number of Left-Hand Radiographs in the Dataset	Reference Standard Bone Age	AI Technique	RMSE (Years)	MAD (Years)	MAE (Years)
Halabi et al., 2018 [54]	Training set: 12,611 Validation set: 1425 Test set: 200	Radiology report provided by RSNA	Inception V3 for pixel information, additional dense layers, and multiple high-performing models		0.35	
Mehta et al., 2021 [55]	Training set: 12,611 Validation set: 1425 Test set: 200	Radiology report provided by RSNA	Inception V3 architecture applied on gamma-corrected images			0.492
Pan et al., 2019 [56]	200 cases test set divided into 1000 validation-test splits	Radiology report provided by RSNA	Combination of 8 models of RSNA Pediatric Bone Age ML Challenge		0.328	
Beheshtian et al., 2023 [57]	Internal validation set: 1425 External test set: 1202	Radiology report	Inception V3 + additional dense layers and multiple high-performing models		0.567 vs. 0.575	
Kim et al., 2017 [58]	Training set: 18,940 Test set: 200	2 experienced radiologists	VUNO Med-BoneAge (deep learning semiautomatic system, based on GP method)	0.6		
Larson et al., 2018 [59]	Training and validation set: 14,036 Test set: 200	Clinical report and 3 human reviewers	CNN model based on GP method	0.63	0.5	
Mutasa et al., 2018 [60]	Training set: 10,289 Test set: 300	Radiology report	14 hidden layer CNN based on GP method			0.536
Lee H. et al., 2017 [62]	Training set: 5828 Test set: 1249	Radiology report	ImageNet pre-trained, fine-tuned CNN	0.82–0.93		
Xu et al., 2022 [64]	Public dataset: 12,600 Clinical Training set: 2014 Clinical Test set: 504	Radiology report	CNN based on TW3 method			0.64; 0.54
Bai et al., 2022 [67]	Training set: 9607 + 11,226 Test set: 1246	10 senior radiologists	3 deep learning models		0.42	
Thodberg et al., 2009 [68]	Training set: 1559 Validation set: 122	Radiologists applying GP method	BoneXpert 2.1 (3 layers deep learning model based on GP and TW3 methods)	0.38		
Lee et al., 2021 [17]	Training set: 2684 Test set: 660	Radiology report (TW3)	HH-boneage. io (fully automated system based on TW3)	0.62		0.46
Zhao et al., 2022 [69]	Test set: 54	3 expert radiologists	Deep learning software by Deep Wise Artificial Intelligence Lab based on TW3 modified for Chinese people	TW3-RUS: 0.501 TW3-Carpal: 0.323	TW3-RUS: 0.379 TW3-Carpal: 0.229	

Only a few systems have been commercialized, one of which is BoneXpert, developed in 2008, which has been progressively upgraded and extended to estimate bone maturation until 19 years (Figure 3).

BoneXpert is currently used in Europe to automatically extract features such as the shape and density of hand–wrist bones, reducing reading times by 87% [70]: it automatically reconstructs the borders from X-ray images of the hand and wrist and calculates “intrinsic” bone ages for each of 13 bones (radius, ulna, and 11 short bones). Finally, BoneXpert converts intrinsic bone ages into TW or Greulich–Pyle GP bone ages. Images with abnormal bone morphology or extremely poor image quality are automatically rejected by the bone reconstruction method [68]. The tool has been demonstrated to be significantly more precise than the traditional application of the GP method and its performance has been improved with the new updates. For instance, version 2.1 achieved an overall root-mean-square deviation of 0.38 years versus 0.71 for the first version [68,70].



VUNO Med-BoneAge and HH-boneage.io are two other AI-based solutions that have been approved and commercialized in Korea by the Korea Food and Drug Administration. VUNO Med-BoneAge is a semi-automatic system that recommends the three most likely estimated bone ages to the physician, based on the GP method. HH-boneage.io, on the other hand, is a fully automatic system based on the TW3 method that predicts bone age with an MAE of 0.46 years and an RMSE of 0.62 years when compared to manual determination [17] (Table 2).



**Figure 3.** Example of an X-ray analyzed with BoneXpert. Standard radiograph of left hand and wrist. BoneXpert automatically recognizes bone segments and analyzes the hand–wrist bones. The analysis results are shown in the black box on the right, where bone age is calculated through both GP and TW3 methods.

Zhao et al. [69] investigated the positive effects of AI-based software on inter- and intra-observer variability in a recent prospective study, in which six board-certified residents

evaluated 56 left-hand wrist radiographs of pre-scholar children aged 3 to 6 years twice: once with and once without the assistance of artificial intelligence. Each resident evaluated the same images in the same way after 4 weeks. The results were compared to three experts' reference bone ages using the RMSE, MAD, and accuracy within 0.5 and 1 year. The study found that AI assistance improves inter-observer agreement and intra-observer reproducibility significantly, making it a valuable learning tool for residents.

**Table 2.** The table describes the most common traditional age-assessment methods and the most famous commercialized AI software applied to left-hand and wrist radiographs.

Traditional Method	Procedure	Commercially Available AI-Based Tool
Greulich and Pyle (GP)	Comparison with reference images contained in the Atlas	BoneXpert; VUNO Med-BoneAge
Gilsanz and Ratibin (GR)	Comparison with reference images contained in the Digital Atlas	
Tanner Whitehouse (TW)	Scoring the level of maturity of specific regions of interest based on the reference scale	BoneXpert; HH-boneage.io

As a fully automated tool has the potential to overcome variability among raters, it allows comparative studies of maturation across the world, beyond the differences among populations, because it could be applied to different populations [71]. For example, in a large comprehensive study in 2010, BoneXpert was applied to American children of four ethnicities—Caucasian, African American, Hispanic, and Asian—finding the largest deviation in Hispanic and Asian children older than 12 years who were about 1 year advanced relative to the GP standard. The automated method was validated for all ethnicities in males from 2.5 to 17 years and for females from 2 to 15 years [72]. In addition, the tool has been tested on Chinese children of middle and large Chinese cities without finding any significant differences compared to Chinese children in Los Angeles [71]. On the contrary, some differences were evident in children from Saudi Arabia, where the BoneXpert-derived GP method has been found suitable with no modification to Saudi Arabian females, while BoneXpert-derived TW3 can be applied to Saudi Arabian males [73].

One of the most important remaining challenges, therefore, is ethnic differences, with the need for more people of various ethnic and socioeconomic backgrounds to be included in future studies to increase the accuracy of AI systems to adapt to all ethnicities that compose our globalized society [74]. Tong et al. tested various AI-based models, such as kernel learning and convolutional neural network algorithms, to find the best solution for combining features from X-ray images, race, and gender, discovering that heterogeneous features improve bone age estimation performance and that multi-kernel learning algorithms best suit heterogeneous data [63].

Another limitation is that AI-based tools have shown limited efficacy in cases of abnormal bone morphology; therefore, despite technological advances, radiologists are still required for bone age evaluation. This is supported by a recent survey of European radiologists, in which only 18% said they used automatic solutions for bone age assessment alone, while 60% said they always checked radiographs to rule out underlying diseases [75]. To address this limitation, some automatic software provides information about bone mineral density and cortical thickness, which indicate bone health, particularly in chronically ill children. For example, a retrospective study published in 2016 found that bone health indices provided by BoneXpert largely correlate with dual-energy X-ray absorption (DXA) or peripheral quantitative computed tomography (pQCT) measurements, suggesting that they may be a valuable screening tool [76].

The main advantages and disadvantages of systems analyzing hand X-ray for bone age calculation are listed in Table 3.

**Table 3.** Advantages and disadvantages of traditional and AI-based approaches for bone age estimation from hand radiographs.

	How It Works	Advantages	Disadvantages	
TRADITIONAL APPROACH	Greulich and Pyle (GP) method	Comparison between the patient radiography and reference images included in the Atlas	<ul style="list-style-type: none"> <li>– wide availability</li> <li>– long time experience</li> </ul>	<ul style="list-style-type: none"> <li>– time-consuming process</li> <li>– affected by ethnicity and generation differences</li> <li>– scarce inter-observer and intra-observer reproducibility</li> </ul>
	Gilsanz and Ratibin (GR) Atlas	Comparison between the patient radiography and reference images included in the Digital Atlas	<ul style="list-style-type: none"> <li>– wide availability</li> <li>– long time experience</li> <li>– high quality digital images</li> </ul>	<ul style="list-style-type: none"> <li>– time-consuming process</li> <li>– affected by ethnicity and generation differences</li> <li>– scarce inter-observer and intra-observer reproducibility</li> </ul>
	Tanner Whitehouse (TW3) Method	Age derived from a score calculated from the analysis of 20 ROIs	<ul style="list-style-type: none"> <li>– more precise than GP and GR methods</li> <li>– higher accuracy and reproducibility</li> </ul>	<ul style="list-style-type: none"> <li>– complexity</li> <li>– time-consuming process</li> <li>– affected by ethnicity and generation differences</li> </ul>
AI-BASED APPROACH	AI-based software provides an automatic result	<ul style="list-style-type: none"> <li>– high accuracy</li> <li>– high inter-rater and intra-rater reproducibility</li> <li>– fast process</li> <li>– more comparative results</li> </ul>	<ul style="list-style-type: none"> <li>– need for basic computer skills</li> <li>– affected by ethnicity differences (less than traditional approaches)</li> <li>– long time needed for validation and commercialization</li> <li>– potential legal issues</li> </ul>	

### 3. Dental Age Assessment

#### 3.1. Traditional Approach

The analysis of dental maturity is another age-assessment technique, as teeth mineralization is less affected by nutritional or endocrine abnormalities compared to the skeleton.

Teeth are a valuable source of age information, particularly in severe accidents where other body parts may not be usable due to burns or severe damage, but teeth are still present. Additionally, several studies have suggested that dental age evaluation is more closely related to physiological age than skeletal estimation [77,78].

Dental age estimation is substantially different between childhood and adulthood: in childhood, the atlas approach is one of the simplest traditional methods, which permits a comparison between the morphological stages of dental development of the subject with a reference standard age-matched orthopantomography. A popular atlas method has been proposed by Scour et al., including a table of reference radiographs of children aged from 4 months until 21 years [79].

Moorrees et al. also created an atlas system and provided various tables, each of which was tailored for a particular sex, while Anderson et al. changed the Moorrees et al. system, adding the third module and producing more detailed tables [1].

The method of Demirjian et al. (1973) [80,81] is based on a scoring system that analyzes the first seven teeth of the left lower quadrant and labels the development stage of each tooth from A to H. Lastly, a global sex-specific maturity score is obtained by summing up the single scores and then converting them into chronological age. This latter method showed a good correlation with the GR estimation [82] and with the chronological age of the Indian population [83].

Several studies had shown that Demirjian's method was affected by an overestimation bias, thus, Willems et al. revisited and tested it on the Belgian Caucasian population, creating new tables with scores expressed in years [84].

Teeth are a good age mark even in adulthood, as the permanent tooth is not a static element; on the contrary, the dentin and the dental pulp undergo pathological and physiological modifications, and every tooth can be analyzed to assess age. However, canines are

particularly useful because they are often still found in older people, in which the loss of various dental elements constitutes a common limitation [78].

The morphological techniques are based on the evaluation of regressive variations, such as the occlusal attrition, the coronal secondary dentine apposition, the apical resorption, the loss of periodontal attachment, and many others [77]. Radiological methods based on measurements of tooth, pulp, and root lengths and the calculation of many width ratios, performed on periapical radiographs, have been developed [85].

No matter the methods used and the accuracy of human experts, traditional dental age estimations show standard deviations of about 10–12 years because of the large spread of morphological variations that exist in nature [77,86] and the high intra- and inter-rater variation, both of which represent a considerable limitation [77].

### 3.2. AI-Based Approach

In a recent study, a CNN model was trained and tested to evaluate the pulp-to-tooth ratio of canines in 300 subjects aged 14 to 60 years using a radiological non-invasive method. They then compared CNN's predictive performance to that of linear regression models.

According to the statistical results, neural network models outperform regression models, with a RMSE of 4.40 years versus 10.26, respectively [78].

Despite of the potential revolutionary role of AI, the current literature is poor on proposed automatic solutions, although some attempts have been made in the last few years.

In 2021, DL procedures were applied to a large dataset of 10,257 orthopantomograms to assess dental age, based on the analysis of the left mandibular eight permanent teeth. An end-to-end convolutional neural network was developed that classified the dental age of the dataset considering three legal age thresholds: 14, 16, and 18 years old. Then the authors compared the results with the traditional methods, proving a more precise performance in the AI-based method. More specifically, the manual method showed an accuracy of 92.5%, 91.3%, and 91.8%, respectively, for each age threshold, while the CNN model reached 95.9%, 95.4%, and 92.3%, respectively [87].

As AI-based solutions have been usually tested on good-quality orthopantomograms image sets without any conditioning dental characteristics, Vila-Blanco et al. decided to use 2289 orthopantomograms of subjects aged 4.5–89.2 years without discarding bad quality images and those containing conditioning characteristics. They compared two CNNs that included or excluded the analysis of sex-specific features, concluding that the sex prediction can reduce the median error in age assessment by about 4 months [88].

Zabrowicz et al. published a recent study on the skeletal maturity assessment of children and teenagers on orthopantomography. They investigated the feasibility of developing a neural model to support the estimation of metric age and created a collection of 21 tooth and bone indicators. Three models were created for this study: one for men and women, one for women alone, and one for men alone. They included cases of men and women, with a quality for the test set of 0.99 and an error for the test set of 0.03. The test quality of the model determining the metric age of men was 0.99, while the test quality of the model comprising cases of women only had a quality for the test set of 0.96 and an error for the test set of 0.03 [89]. In the further development of this study, the authors built deep neural network models through the collection of 21 unique indicators, to determine whether deep neural network models were more accurate than earlier models. Depending on the learning set used, the generated models MAE ranged from 2.34 to 4.61 months, while their RMSE ranged from 5.58 to 7.49 months. R2 was a correlation value that varied between 0.92 and 0.96 [90].

Kim et al. demonstrated the use of artificial neural networks in 2021 [45] for information and image processing in dentistry. By applying a CNN to pantomographic images of first molars from X-rays, they looked at the estimation of age categories. Images of the right and left maxillary and mandibular first molars comprised the data collection. The research used 1586 pantomographic images in total. It was discovered that the traditional neural network focused on anatomical factors, such as the dental pulp chamber, alveolar bone

level, and interdental space. The networks created in this manner had very high efficiency, varying from 87.04% to 88.33% [91].

Deep CNN was employed in a 2021 article by Banjak et al. to estimate the age group [92]. There were 4035 orthopantomography images in the learning collection. The age of 89 archaeological skull remains was determined using the newly created neural network, which showed a 73% accuracy rate.

Milošević et al. [93] and Kahaki et al. [94] also used deep CNN to assess tooth age from dental radiographs. A learning set of 4035 orthopantomography and 76,416 dental radiographs of individuals between the ages of 19 and 90 was developed by Milošević's team. For panoramic images, the median error was 2.9 years, whereas for solitary tooth images, it was 4.6 years [93].

The efficacy of determining a patient's age using artificial intelligence was tested instead by Kahaki's team using 456 pantomographic images of children between the ages of 1 and 17. They developed 12 neural networks to depict the age ranges of males and females: 1–4, 5–7, 8–10, 11–13, 14–17, and 1–17. The networks with the best test quality of over 90% were for the 14–17 age range for both sexes. The test quality for the other age categories was greater than 80% [94].

Unfortunately, the limited number of attempts present in the literature suggests that the commercialization of reliable software could be far away.

## 4. Other Methods

### 4.1. Traditional Approaches

Several other traditional methods have been developed in the previous decades, exploiting various skeletal segments. We provide an overview of them below, even if they are less diffuse in the clinical practice.

In 1962, Sauvegrain et al. developed a comprehensive 27-point scoring system for determining bone age based on the analysis of four elbow ossification centers on the anteroposterior and latero-lateral projection of elbow radiography [95]. This system describes the different elbow ossification centers in terms of their appearance in chronological order, their fusion, and their relationship to age, as follows: capitellum (0–1 year; 10–15 years), radial head (2–6 years; 12–16 years), medial epicondyle (2–8 years; 13 years), trochlea (5–11 years; 10–18 years), olecranon (6–11 years; 13–16 years), lateral epicondyle (8–13 years; 12–16 years). The olecranon apophysis exhibits a distinct morphological development between 11 and 13 years of skeletal age in girls and between 13 and 15 years in boys. Due to the significant morphological changes in the elbow occurring every six months, this technique is considered a reliable way to determine skeletal age during puberty [22], but the double projection needed for the evaluation exposes individuals to a higher radiation dose than the single projection used, for example, for the GP [82].

As during adolescence, a secondary ossification center develops at the medial extremity of the clavicle and undergoes a complete fusion at the age of 22; the analysis of the medial clavicle epiphysis has also been proposed for bone age estimation between 18 and 22 years of age, on X-ray, CT, and US [96–98]. Schmeling et al., for example, proposed a classification system for the ossification of the medial clavicle epiphyseal cartilage on X-ray in 5 stages: 1: non-ossified epiphysis; 2: visible ossification center; 3: partial fusion; 4: complete fusion; and 5: the disappearance of the epiphyseal scar. Stage 3 was first identified in both sexes at the age of 16 years old: women entered stage 4 at age 20 and men at age 21, respectively, and the earliest detection of stage 5 in both sexes occurred at age 26 [99]. Other authors proposed a 4-stage scale with a fourth stage corresponding to a combination of the stages 4 and 5 from the Schmeling classification [99,100]. The main limitation is that a standard clavicle X-ray can be hindered by the overlapping of mediastinal structures, vertebrae, or ribs that may affect visualization of the medial epiphysis [1]. CT provides an accurate delineation of the clavicle medial end, but its widespread use is limited by radiation exposure.

Other authors proposed a bone age assessment based on the analysis of the femoral head, by assessing the depth of epiphyseal cartilage through US; as the ossification progresses, most of the cartilage is replaced by bone and hyaline articular cartilage [1].

Li et al. recently described a classification system for proximal humeral physis, based on the fusion of the external portion of the physis, divided into 5 stages: stage 1 represents an incompletely ossified lateral epiphysis; stage 2 shows an increased ossification of the lateral epiphysis; stage 3, in which the lateral half of the physis is open without evident fusion; stage 4, in which the lateral half of physis is partially fused; and stage 5, in which the fusion is complete [101].

The study of apophyseal ossification of the iliac crest has been proposed for forensic purposes and relies on the Risser sign, which is based on the degree of maturation of the iliac crest apophysis [102]. Its evaluation has been proposed with both CT [103] and US [104], but the ossification of the iliac crest apophysis is not uniform, thus resulting in discrepancies in bone age calculation and the need for an extensive validation.

Recent studies proposed different bone age assessments based on the study of the cervical vertebrae [105–107] to differentiate subjects who have not yet reached peak pubertal growth and those who have reached or passed it. In the study by Hsiang-Hua Lai, for example, the cervical vertebral maturation stages were determined through the latest Cervical Vertebral Maturation Stage Index and combined with the analysis of hand–wrist radiography [106]. Cervical vertebrae maturation analysis was proposed as a valid alternative to hand–wrist skeletal maturity calculation in the Taiwanese population.

Li et al. [108] created a bone age calculation based on the visualization of the calcaneal apophysis, divided into 5 stages: in stage 0, no ossification of the apophysis is visible; in stage 1, the apophysis covers less than 50% of the metaphysis; in stage 2 the apophysis interests more than 50% of the metaphysis; in stage 3, it has extended over the plantar surface and continues to extend over the dorsal surface without fusion; in stage 4, the fusion of the apophysis to metaphysis is visible but partial; and in stage 5, it is complete.

O'Connor et al. proposed a method to identify the relationship between the stage of the epiphyseal union at the knee joint and chronological age investigated in a modern Irish population [109]. The authors analyzed the anteroposterior and lateral knee X-rays of males and females aged 9–19 years, and classified fusion into four stages: 0, non-union; 1, beginning union; 2, active union; 3, recent union; and 4, complete union. The authors stated that in both males and females, the stage of epiphyseal union correlated with chronological age and that in this Irish population, the stages of union start earlier. Due to the highly selected population, the results are difficult to generalize.

Another study tried to estimate bone age through the ossification stage of the distal femoral epiphysis on a 3T MRI in the age group 10 to 30 years, with a classification into five stages of epiphyseal fusion [110].

A possible approach to grade skeletal maturity is based on the study of the apophyseal ossification of the iliac crest, according to Risser's sign, a 5-score classification, as follows: 1, 25% iliac apophysis ossification, localized in the anterior superior iliac spine; 2, 50% iliac apophysis ossification with extension across the iliac wing; 3, 75% iliac apophysis ossification; 4, 100% ossification, with no fusion to the iliac crest; and 5, iliac apophysis fusion to the iliac crest and representing the cessation of growth [111]. This approach has also been investigated for forensic purposes using quantitative descriptors in an Australian population, retrospectively collecting X-rays and CTs of individuals aged 7–25, demonstrating complete fusion between 17.3–19.2 and 17.1–20.1 years in males and females, respectively [103].

The main characteristics of the above studies are collected in Table 4.

**Table 4.** The different methods for bone age estimation based on the analysis of different skeletal segments.

Authors	Anatomical Region	Imaging Technique	Process Description	Age Range of Subjects	Disadvantages
Sauvegrain et al. [95]	Elbow	Radiography (AP and LL projections)	Evaluation of 4 elbow ossification centers, basing on a 27-point scoring system	0–18	The double projection exposes to a higher radiation dose
Schmeling et al. [99]	Medial clavicle epiphysis	Radiography	Evaluation of ossification degree of the cartilage, basing on a 5-stage classification system	18–22	Conventional clavicle X-ray can be hindered by overlapping images related to the mediastinal structures, vertebrae, or ribs
Li et al. [101]	Proximal humeral physis	Radiography	Evaluation of the humeral head epiphysis and fusion of the external portion of the physis, based on a 5-stage scale	10–15	Historical collection of radiographs (dated 1926–1942)
Lottering et al. [103]	Iliac crest	CT	Evaluation of the apophyseal ossification of the iliac crest, according to Risser’s sign, a 5-score classification	7–25	The ossification of iliac crest apophysis is not uniform, thus it can create some discrepancies
Schmidt et al. [104]	Iliac crest	US	Apophyseal ossification of the iliac crest, according to Risser’s sign, a 5-score classification	11–20	The ossification of iliac crest apophysis is not uniform, thus it can create some discrepancies
Soegiharto et al. [105]	Cervical vertebrae	Radiography (lateral cephalometric)	Evaluation of C2, C3 and C4, basing on a 6-stage maturation scale	8–17	The method originally was developed more than 5 decades ago, without a fair description of the classification system, until a few years ago.
Li et al. [108]	Calcaneal apophysis	Radiography (lateral foot projection)	Evaluation of the calcaneal apophysis, based on a 5-stage scale	7–16	Ethnical differences
O’Connor et al. [109]	Knee	Radiography (AP and LL knee projections)	Evaluation of the stage of the epiphyseal union at the knee joint, basing on a 5-stage scale	9–19	The study was applied to a highly selected population (Irish), thus the results are difficult to generalize
Krämer et al. [110]	Distal femur	3T MRI	Evaluation of the ossification stage of the distal femoral epiphysis, based on a 5-stage scale	10–30	Unbalanced age distribution of subjects, particularly in the lower age groups; only one sectional plane and only one MRI weighting were considered.

#### 4.2. AI-Based Approaches

In 2018, the applicability of the Sauvegrain method through DL was investigated by Bin Baik et al. [112]; as elbow X-rays are less commonly applied for age assessment, they first applied the data augmentation to expand the number of images needed to for DL algorithms and trained a CNN on 576 images, then compared the automatic age estimation results with those performed by experts. The MAE of the automatic version of the Sauvegrain method was 2.8 months and the Mean Absolute Percentage Error was 0.018, demonstrating a performance similar to that of experienced radiologists [112].

Automatic methods for bone age assessment in the knee are mainly based on MRI, through the assessment of growth plate ossification; some authors developed a DL system trained on 185 coronal and 404 sagittal studies of Caucasian males aged 13 to 21, involving image pre-processing, bone segmentation, and age estimation. The combined use of CNN and three ML-based algorithms allowed an accuracy of 90.9%, a sensitivity of 88.6%, and a specificity of 94.2% in classification and a MAE of 0.49 years in age regression [113].

In their proposal, Dallora et al. combined two CNN models: the first one selected the most useful images from the MRI study, which were then analyzed by a second module

forage estimation trained on knee MRI scans of 402 volunteers between the ages of 14 and 21. The bone age of male subjects chronologically aged between 14 and 20.5 years could be determined with a MAE of 0.793 years, and the bone age of female subjects aged between 14 and 19.5 years could be determined with a MAE of 0.988 years. Using a cut-off age of 18 years, it was possible to classify subjects with an accuracy of 98.1% for males and 95.0% for females [114].

Other authors proposed an automatic age bone calculation based on MRI images of hands, clavicles, and teeth: a deep CNN was trained on a dataset of 322 subjects in the age range between 13 and 25 years, with the demonstration of mean absolute prediction error in regressing chronological age of  $1.01 \pm 0.74$  years [115].

Bone age estimation can be also performed through the cervical vertebral maturation analysis: Liao et al. developed a DL system for automatic calculation and propose a CNN called iCVM [116]; whereas Kim et al. developed eight ML models including 13 anatomic landmarks of the cervical vertebrae as input data, the width–height ratio, width–perpendicular height ratio, concavity ratio, patients' chronological age, and sex. The MAE, round MAE, and RMSE values were, respectively, 0.90, 0.87, and 1.20 [117].

Peng et al. tested the performance of three DL models (VGG19, Inception-V3, and Inception-ResNet-V2) to pelvic radiographs by collecting 962 pelvic X-rays from adolescents and young adults (481 males, 481 females) between 11 and 21 years and divided them into training and validation (80%) and test sets (20%).

The performances were compared by calculating the RMSE, MAE, and Bland–Altman plots between the automatic age estimation and the chronological ages: compared to the chronological age, the RMSE and MAE of VGG19 were 1.29 and 1.02 years, respectively; the RMSE and MAE of the Inception-V3 model were 1.17 and 0.82 years, respectively; while the RMSE and MAE of the Inception-ResNet-V2 model were 1.11 and 0.84 years, respectively. Lastly, the lowest mean value of differences between age estimates and the chronological ages by Bland–Altman plots was reached by the Inception-ResNet-V2 model [118].

Pelvic bones play an important role in forensic bone age assessment; however, due to the overlapping of pelvic organs in radiography, which reduces accuracy, there are insufficient applications of machine learning to pelvic X-ray images. In 2022, Peng et al. attempted to overcome this limitation by applying the previously mentioned CNN models to 2164 images after U-Net detected and segmented the regions of interest. The segmented images were then combined with the original images to enhance the key areas and, as a result, reduce the pelvic organ profiles. Hence, the three CNN models were applied both to enhanced and not enhanced images, showing better performances after regional augmentation. For example, the RMSE of the Inception-V3, Inception-ResNet-V2, and VGG19 networks applied on enhanced images were 0.93 years, 1.12 years, and 1.14 years, respectively, while the RMSE became 1.22 years, 1.25 years, and 1.63 years, respectively, on not enhanced images [119].

This result suggests that a good segmentation tool could be an excellent starting point to create new automatic tools, perhaps analyzing anatomical regions that have never been investigated so far.

Some segmentation tools have already been tested on different bones and various radiological techniques, ranging from lateral X-ray images of the patella to 3D MRI images of the hip joint [120,121].

Nowadays, studies are focusing on the most used traditional procedures, including hand–wrist radiographs evaluation, but we cannot exclude that in the future one or more software with a multi-disciplinary approach will exist to best suit each case.

The main results of the above studies are collected in Table 5.



**Table 5.** The available AI applications for bone age estimation based on the analysis of skeletal segments different from hand–wrist.

Authors	Object of Analysis	Dataset	AGE	CNN	MAE	Accuracy (%)	RMSE (Years)
Bin Baik et al. [112]	Elbow radiographs	576	adolescents	U-Net, RPN+, F-RCNN, VGC16	2.8 months		
Der Mauer et al. [113]	Knee MRI	589	13–21	N4ITK, MAdM, U-Net, AgeNet2D	0.67 ± 0.49 y	90.9	
Dallora et al. [114]	Knee MRI	402	14–21	GoogleNet, ResNet-50, Inception-v3, VGG, AlexNet, DenseNet, U-Net	0.793–0.988 y	95–98.1	
Štern et al. [115]	MRI of hands, clavicle, and teeth	322	13–25	Inception V3	1.01 ± 0.74 y		
Kim et al. [117]	Cervical vertebrae in lateral cephalograms	499	6–18	BayesianRidge, Ridge, LinearRegression, HuberRegressor, SGDRegressor, RandomForestRegressors, TheilSenRegressor, AdaBoostRegressor and LinearSV	0.9 y		1.2
Peng et al. [118]	Pelvic radiographs	962	11–21	Inception-V3, Inception-ResNet-V2, and VGG19	0.82–1.02 y		1.11–1.29
Peng et al. [119]	Pelvic radiographs	2164	11–21	Inception-V3, Inception-ResNet-V2, and VGG19, U-Net	0.93–1.14 y		1.22–1.63

## 5. Challenges and Perspectives

The perspectives opened by the AI approaches are huge and represent the main promise and challenge for radiology in the next years. However, when dealing with the creation and diffuse adoption of AI-based software in clinical practice, some difficulties should be considered.

Developing AI-based tools takes a multidisciplinary team effort and knowledge, large amounts of high-quality data, and a rigorous workflow.

Access to a large amount of data is essential for the development of reliable AI approaches [122]: large datasets for training, validating, and testing these models are essential. This problem has been partially addressed in recent years by the rising availability of open-source image repositories and data sharing. Large dataset requirements may also encourage cross-institutional collaboration in research.

Another significant problem relates to the so-called “black box” phenomenon and the difficulty in understanding results from AI-based models [24,123]. Executing complex tasks requires the use of complex models that operate at a level of abstraction that makes it challenging to fully comprehend how they interact with the characteristics that radiologists typically classify. For instance, “decision tree” type models have high interpretability because the algorithm on which they are based can be broken down and understood when used to operate a classification task. However, they may perform inaccurately when used to operate more complex tasks. On the other hand, AI models such as Support Vector Machines or CNN which are considered particularly suitable for imaging tasks, show great accuracy at the expense of less interpretability.

Model validation is a critical step in the creation of ML models. The data used to train the model in the real world represents a sample of the population that may or may not be fully representative. When used outside of the training set, a model’s performance might degrade if the sample is not sufficiently representative.

This is where validation, which is necessary and not optional, comes into play.

Among the many models, random sampling, leave-one-out cross-validation, k-fold cross-validation, and bootstrapping are some of the most popular. Choosing the best model is influenced by several variables.

Focusing on AI-based bone age-assessment methods, the most important problem is the application to people of various ethnic and socioeconomic backgrounds, as it is well-known that different populations show different rates of skeletal maturation [124]. It is therefore justifiable to be skeptical about the direct applicability of AI-based bone age-assessment methods, particularly those based on the GP approach, in populations of different ethnicities aside from the ones considered during the AI system development.

Studies on the application of these methods in various ethnic groups are being carried out to solve this problem, demonstrating that the AI-based bone age assessment could correctly analyze images of all ethnicities [72].

Due to the ongoing development and research, more AI solutions based on larger datasets including different ethnicities will be commercialized in the future [125]. A certain number of improvements and upgrades will also be made to the currently available software. As a result, AI-based methods for automated bone age assessment will become more accurate and reliable across populations, and current flaws will be addressed [17].

As AI algorithms become more common in daily reporting activities, radiologists will need to gain a deeper understanding of them and be aware of the purposes for which the models were developed. More importantly, understanding these tools and their limitations in clinical practice will be critical.

Education will play a key role in the widespread adoption of these solutions [54].

## 6. Conclusions

The era of AI might be viewed as an ongoing revolution with an unstoppable trend.

The focus of research has switched from improving the performance of AI-based systems to finding effective ways to use AI to help optimize human performance.

We present an overview of AI applications for bone age calculation, based on the analysis of different bone segments and teeth, trusting that the future diffuse adoption of automated bone age evaluations powered by AI can ease the workload for radiologists who must process a large number of pictures to identify bone age. Additionally, it can greatly lessen the subjectivity, inter- and intra-observer variability, and other above-discussed problems related to conventional age-determining methods.

**Author Contributions:** Conceptualization: M.C. (Michaela Cellina); methodology: M.C. (Maurizio Cè) and M.C. (Michaela Cellina); literature research: E.C., G.O. and A.P.; data curation: C.M. and E.C.; writing—original draft preparation: M.C. (Maurizio Cè), E.C., A.P. and M.C. (Michaela Cellina); review and editing: E.C., M.C. (Michaela Cellina), D.G. and C.M.; supervision: M.C. (Michaela Cellina). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mughal, A.M.; Hassan, N.; Ahmed, A. Bone age assessment methods: A critical review. *Pak. J. Med. Sci.* **2014**, *30*, 211–215. [[CrossRef](#)]
2. Satoh, M. Bone age: Assessment methods and clinical applications. *Clin. Pediatr. Endocrinol.* **2015**, *24*, 143–152. [[CrossRef](#)]
3. Creo, A.L.; Schwenk, W.F. Bone Age: A Handy Tool for Pediatric Providers. *Pediatrics* **2017**, *140*, e20171486. [[CrossRef](#)]
4. Martin, D.D.; Wit, J.M.; Hochberg, Z.; Säwendahl, L.; van Rijn, R.R.; Fricke, O.; Cameron, N.; Caliebe, J.; Hertel, T.; Kiepe, D.; et al. The Use of Bone Age in Clinical Practice—Part 1. *Horm. Res. Paediatr.* **2011**, *76*, 1–9. [[CrossRef](#)] [[PubMed](#)]

5. Ostojic, S.M. Prediction of adult height by Tanner-Whitehouse method in young Caucasian male athletes. *QJM Int. J. Med.* **2012**, *106*, 341–345. [CrossRef] [PubMed]
6. Kim, Y.J.; Kwon, A.; Jung, M.K.; Kim, K.E.; Suh, J.; Chae, H.W.; Kim, D.H.; Ha, S.; Seo, G.H.; Kim, H.-S. Incidence and Prevalence of Central Precocious Puberty in Korea: An Epidemiologic Study Based on a National Database. *J. Pediatr.* **2019**, *208*, 221–228. [CrossRef] [PubMed]
7. Kim, J.R.; Lee, Y.S.; Yu, J. Assessment of Bone Age in Prepubertal Healthy Korean Children: Comparison among the Korean Standard Bone Age Chart, Greulich-Pyle Method, and Tanner-Whitehouse Method. *Korean J. Radiol.* **2015**, *16*, 201–205. [CrossRef]
8. Karami, M.; Rabiei, M.; Riahinezhad, M. Evaluation of the pelvic apophysis with multi-detector computed tomography for legal age estimation in living individuals. *J. Res. Med. Sci.* **2015**, *20*, 209. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4468222/> (accessed on 17 January 2023).
9. Greulich, W.W.; Pyle, S.I. *Radiographic Atlas of Skeletal Development of the Hand and Wrist Professor of Anatomy*; Stanford University School of Medicine: Stanford, CA, USA, 1959.
10. Gilsanz, V.; Ratib, O. *Hand Bone Age: A Digital Atlas of Skeletal Maturity*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 1–96. [CrossRef]
11. Serinelli, S.; Panetta, V.; Pasqualetti, P.; Marchetti, D. Accuracy of three age determination X-ray methods on the left hand-wrist: A systematic review and meta-analysis. *Leg. Med.* **2011**, *13*, 120–133. [CrossRef]
12. Kim, S.; Oh, Y.; Shin, J.; Rhie, Y.; Lee, K. Comparison of the Greulich-Pyle and Tanner Whitehouse (TW3) Methods in Bone age Assessment. *J. Korean Soc. Pediatr. Endocrinol.* **2008**, *13*, 50–55.
13. Berst, M.J.; Dolan, L.; Bogdanowicz, M.M.; Stevens, M.A.; Chow, S.; Brandser, E.A. Effect of Knowledge of Chronologic Age on the Variability of Pediatric Bone Age Determined Using the Greulich and Pyle Standards. *Am. J. Roentgenol.* **2001**, *176*, 507–510. [CrossRef]
14. Cunha, E.; Baccino, E.; Martrille, L.; Ramsthaler, F.; Prieto, J.; Schuliar, Y.; Lynnerup, N.; Cattaneo, C. The problem of aging human remains and living individuals: A review. *Forensic Sci. Int.* **2009**, *193*, 1–13. [CrossRef]
15. Bull, R.K.; Edwards, P.D.; Kemp, P.M.; Fry, S.; Hughes, I.A. Bone age assessment: A large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods. *Arch. Dis. Child.* **1999**, *81*, 172–173. [CrossRef] [PubMed]
16. Reynolds, B.C.; Beattie, T.J.; Ramage, I.J.; Lucas, P.; Law, C.; Baird, J. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*, 3rd ed.; W.B. Saunders: London, UK, 2001.
17. Lee, B.-D.; Lee, M.S. Automated Bone Age Assessment Using Artificial Intelligence: The Future of Bone Age Assessment. *Korean J. Radiol.* **2021**, *22*, 792–800. [CrossRef] [PubMed]
18. Aja-Fernández, S.; Garcia, R.D.L.; Martín-Fernández, M.; Alberola-Lopez, C. A computational TW3 classifier for skeletal maturity assessment. A Computing with Words approach. *J. Biomed. Informatics* **2004**, *37*, 99–107. [CrossRef] [PubMed]
19. Kellinghaus, M.; Schulz, R.; Vieth, V.; Schmidt, S.; Schmeling, A. Forensic age estimation in living subjects based on the ossification status of the medial clavicular epiphysis as revealed by thin-slice multidetector computed tomography. *Int. J. Leg. Med.* **2009**, *124*, 149–154. [CrossRef]
20. Soegiharto, B.M.; Moles, D.; Cunningham, S.J. Discriminatory ability of the skeletal maturation index and the cervical vertebrae maturation index in detecting peak pubertal growth in Indonesian and white subjects with receiver operating characteristics analysis. *Am. J. Orthod. Dentofac. Orthop.* **2008**, *134*, 227–237. [CrossRef]
21. DiMéglio, A.; Charles, Y.P.; Daires, J.-P.; De Rosa, V.; Kaboré, B. Accuracy of the Sauvegrain Method in Determining Skeletal Age During Puberty. *J. Bone Jt. Surg.* **2005**, *87*, 1689–1696. [CrossRef]
22. Canavese, F.; Charles, Y.P.; DiMéglio, A. Skeletal age assessment from elbow radiographs. Review of the literature. *Chir. Organi Mov.* **2008**, *92*, 1–6. [CrossRef]
23. Michael, D.; Nelson, A. HANDX: A model-based system for automatic segmentation of bones from digital hand radiographs. *IEEE Trans. Med. Imaging* **1989**, *8*, 64–69. [CrossRef]
24. Coppola, F.; Faggioni, L.; Regge, D.; Giovagnoni, A.; Golfieri, R.; Bibbolino, C.; Miele, V.; Neri, E.; Grassi, R. Artificial intelligence: Radiologists' expectations and opinions gleaned from a nationwide online survey. *La Radiol. Med.* **2020**, *126*, 63–71. [CrossRef] [PubMed]
25. Nakaura, T.; Higaki, T.; Awai, K.; Ikeda, O.; Yamashita, Y. A primer for understanding radiology articles about machine learning and deep learning. *Diagn. Interv. Imaging* **2020**, *101*, 765–770. [CrossRef] [PubMed]
26. Cellina, M.; Pirovano, M.; Ciocca, M.; Gibelli, D.; Floridi, C.; Oliva, G. Radiomic analysis of the optic nerve at the first episode of acute optic neuritis: An indicator of optic nerve pathology and a predictor of visual recovery? *La Radiol. Med.* **2021**, *126*, 698–706. [CrossRef]
27. Scapicchio, C.; Gabelloni, M.; Barucci, A.; Cioni, D.; Saba, L.; Neri, E. A deep look into radiomics. *La Radiol. Med.* **2021**, *126*, 1296–1311. [CrossRef] [PubMed]
28. Jung, A. *Machine Learning*; Springer Nature: Singapore, 2022. [CrossRef]
29. Granata, V.; Fusco, R.; De Muzio, F.; Cutolo, C.; Setola, S.V.; Dell'Aversana, F.; Grassi, F.; Belli, A.; Silvestro, L.; Ottaiano, A.; et al. Radiomics and machine learning analysis based on magnetic resonance imaging in the assessment of liver mucinous colorectal metastases. *La Radiol. Med.* **2022**, *127*, 763–772. [CrossRef] [PubMed]
30. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [CrossRef]

31. Chen, X.; Wang, X.; Zhang, K.; Fung, K.-M.; Thai, T.C.; Moore, K.; Mannel, R.S.; Liu, H.; Zheng, B.; Qiu, Y. Recent advances and clinical applications of deep learning in medical image analysis. *Med. Image Anal.* **2022**, *79*, 102444. [CrossRef]
32. Colombo, E.; Fick, T.; Esposito, G.; Germans, M.; Regli, L.; van Doormaal, T. Segmentation techniques of brain arteriovenous malformations for 3D visualization: A systematic review. *La Radiol. Med.* **2022**, *127*, 1333–1341. [CrossRef]
33. Matsoukas, S.; Scaggiante, J.; Schuldt, B.R.; Smith, C.J.; Chennareddy, S.; Kalagara, R.; Majidi, S.; Bederson, J.B.; Fifi, J.T.; Mocco, J.; et al. Accuracy of artificial intelligence for the detection of intracranial hemorrhage and chronic cerebral microbleeds: A systematic review and pooled analysis. *La Radiol. Med.* **2022**, *127*, 1106–1123. [CrossRef]
34. Chiu, H.-Y.; Chao, H.-S.; Chen, Y.-M. Application of Artificial Intelligence in Lung Cancer. *Cancers* **2022**, *14*, 1370. [CrossRef]
35. Moore, M.M.; Slonimsky, E.; Long, A.D.; Sze, R.W.; Iyer, R.S. Machine learning concepts, concerns and opportunities for a pediatric radiologist. *Pediatr. Radiol.* **2019**, *49*, 509–516. [CrossRef] [PubMed]
36. Tan, X.J.; Cheor, W.L.; Lim, L.L.; Ab Rahman, K.S.; Bakrin, I.H. Artificial Intelligence (AI) in Breast Imaging: A Scientometric Umbrella Review. *Diagnostics* **2022**, *12*, 3111. [CrossRef] [PubMed]
37. Ullah, N.; Khan, J.A.; Almakdi, S.; Khan, M.S.; Alshehri, M.; Alboaneen, D.; Raza, A. A Novel CovidDetNet Deep Learning Model for Effective COVID-19 Infection Detection Using Chest Radiograph Images. *Appl. Sci.* **2022**, *12*, 6269. [CrossRef]
38. Nakamura, Y.; Higaki, T.; Honda, Y.; Tatsugami, F.; Tani, C.; Fukumoto, W.; Narita, K.; Kondo, S.; Akagi, M.; Awai, K. Advanced CT techniques for assessing hepatocellular carcinoma. *La Radiol. Med.* **2021**, *126*, 925–935. [CrossRef] [PubMed]
39. Han, S.; Lee, J.; Lee, S. Activation Fine-Tuning of Convolutional Neural Networks for Improved Input Attribution Based on Class Activation Maps. *Appl. Sci.* **2022**, *12*, 12961. [CrossRef]
40. Alshehri, A.; AlSaeed, D. Breast Cancer Detection in Thermography Using Convolutional Neural Networks (CNNs) with Deep Attention Mechanisms. *Appl. Sci.* **2022**, *12*, 12922. [CrossRef]
41. Han, D.; Chen, Y.; Li, X.; Li, W.; Zhang, X.; He, T.; Yu, Y.; Dou, Y.; Duan, H.; Yu, N. Development and validation of a 3D-convolutional neural network model based on chest CT for differentiating active pulmonary tuberculosis from community-acquired pneumonia. *La Radiol. Med.* **2022**, *128*, 68–80. [CrossRef]
42. Parekh, V.S.; Jacobs, M.A. Deep learning and radiomics in precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **2019**, *4*, 59–72. [CrossRef]
43. Mettler, F.A., Jr.; Huda, W.; Yoshizumi, T.T.; Mahesh, M. Effective Doses in Radiology and Diagnostic Nuclear Medicine: A Catalog. *Radiology* **2008**, *248*, 254–263. [CrossRef]
44. Alshamrani, K.; Messina, F.; Offiah, A.C. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. *Eur. Radiol.* **2019**, *29*, 2910–2923. [CrossRef]
45. Lin, N.-H.; Ranjitkar, S.; Macdonald, R.; Hughes, T.; Taylor, J.A.; Townsend, G. New growth references for assessment of stature and skeletal maturation in Australians. *Aust. Orthod. J.* **2006**, *22*, 1–10. [PubMed]
46. Soudack, M.; Ben-Shlush, A.; Jacobson, J.; Raviv-Zilka, L.; Eshed, I.; Hamiel, O. Bone age in the 21st century: Is Greulich and Pyle’s atlas accurate for Israeli children? *Pediatr. Radiol.* **2012**, *42*, 343–348. [CrossRef] [PubMed]
47. Büken, B.; Şafak, A.A.; Yazıcı, B.; Büken, E.; Mayda, A.S. Is the assessment of bone age by the Greulich–Pyle method reliable at forensic age estimation for Turkish children? *Forensic Sci. Int.* **2007**, *173*, 146–153. [CrossRef]
48. Calfee, R.P.; Sutter, M.; Steffen, J.A.; Goldfarb, C.A. Skeletal and chronological ages in American adolescents: Current findings in skeletal maturation. *J. Child. Orthop.* **2010**, *4*, 467–470. [CrossRef]
49. Kaplowitz, P.; Srinivasan, S.; He, J.; McCarter, R.; Hayeri, M.R.; Sze, R. Comparison of bone age readings by pediatric endocrinologists and pediatric radiologists using two bone age atlases. *Pediatr. Radiol.* **2010**, *41*, 690–693. [CrossRef]
50. Lin, F.-Q.; Zhang, J.; Zhu, Z.; Wu, Y.-M. Comparative study of Gilsanz-Ratib digital atlas and Greulich-Pyle atlas for bone age estimation in a Chinese sample. *Ann. Hum. Biol.* **2014**, *42*, 523–527. [CrossRef] [PubMed]
51. Kahleyss, S.; Hoepffner, W.; Keller, E.; Willgerodt, H. The determination of bone age by the Greulich-Pyle and Tanner-Whitehouse methods as a basis for the growth prognosis of tall-stature girls. *Pediatr. Relat. Top.* **1990**, *29*, 137–140. Available online: <https://europepmc.org/article/MED/2352741> (accessed on 17 January 2023).
52. Ahmed, M.L.; Warner, J.T. TW2 and TW3 bone ages: Time to change? *Arch. Dis. Child.* **2007**, *92*, 371–372. [CrossRef]
53. Gross, G.W.; Boone, J.M.; Bishop, D.M. Pediatric skeletal age: Determination with neural networks. *Radiology* **1995**, *195*, 689–695. [CrossRef]
54. Halabi, S.S.; Prevedello, L.; Kalpathy-Cramer, J.; Mamonov, A.B.; Bilbily, A.; Cicero, M.; Pan, I.; Pereira, L.A.; Sousa, R.; Abdala, N.; et al. The RSNA Pediatric Bone Age Machine Learning Challenge. *Radiology* **2019**, *290*, 498–503. [CrossRef]
55. Mehta, C.; Ayeesha, B.; Sotakanal, A.; Nirmala, S.R.; Desai, S.D.; Suryanarayana, K.V.; Ganguly, A.D.; Shetty, V. Deep Learning Framework for Automatic Bone Age Assessment. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Virtual, 1–5 November 2021; pp. 3093–3096.
56. Pan, I.; Thodberg, H.H.; Halabi, S.S.; Kalpathy-Cramer, J.; Larson, D.B. Improving Automated Pediatric Bone Age Estimation Using Ensembles of Models from the 2017 RSNA Machine Learning Challenge. *Radiol. Artif. Intell.* **2019**, *1*, e190053. [CrossRef]
57. Beheshtian, E.; Putman, K.; Santomartino, S.M.; Parekh, V.S.; Yi, P.H. Generalizability and Bias in a Deep Learning Pediatric Bone Age Prediction Model Using Hand Radiographs. *Radiology* **2023**, *306*, 2. [CrossRef] [PubMed]
58. Kim, J.R.; Shim, W.H.; Yoon, H.M.; Hong, S.H.; Lee, J.S.; Cho, Y.A.; Kim, S. Computerized Bone Age Estimation Using Deep Learning Based Program: Evaluation of the Accuracy and Efficiency. *Am. J. Roentgenol.* **2017**, *209*, 1374–1380. [CrossRef] [PubMed]

59. Larson, D.B.; Chen, M.C.; Lungren, M.P.; Halabi, S.S.; Stence, N.V.; Langlotz, C.P. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. *Radiology* **2018**, *287*, 313–322. [CrossRef] [PubMed]
60. Mutasa, S.; Chang, P.D.; Ruzal-Shapiro, C.; Ayyala, R. MABAL: A Novel Deep-Learning Architecture for Machine-Assisted Bone Age Labeling. *J. Digit. Imaging* **2018**, *31*, 513–519. [CrossRef]
61. 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. [CrossRef] [PubMed]
62. Lee, H.; Tajmir, S.; Lee, J.; Zissen, M.; Yesiwas, B.A.; Alkasab, T.K.; Choy, G.; Do, S. Fully Automated Deep Learning System for Bone Age Assessment. *J. Digit. Imaging* **2017**, *30*, 427–441. [CrossRef]
63. 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. [CrossRef]
64. Xu, X.; Xu, H.; Li, Z. Automated Bone Age Assessment: A New Three-Stage Assessment Method from Coarse to Fine. *Healthcare* **2022**, *10*, 2170. [CrossRef]
65. Somkantha, K.; Theera-Umpun, N.; Auephanwiriyakul, S. Bone Age Assessment in Young Children Using Automatic Carpal Bone Feature Extraction and Support Vector Regression. *J. Digit. Imaging* **2011**, *24*, 1044–1058. [CrossRef]
66. Zhang, A.; Gertych, A.; Liu, B.J.; Huang, H.K. Bone age assessment for young children from newborn to 7-year-old using carpal bones. *Med. Imaging 2007 PACS Imaging Inform.* **2007**, *6516*, 651618. [CrossRef]
67. Bai, M.; Gao, L.; Ji, M.; Ge, J.; Huang, L.; Qiao, H.; Xiao, J.; Chen, X.; Yang, B.; Sun, Y.; et al. The uncovered biases and errors in clinical determination of bone age by using deep learning models. *Eur. Radiol.* **2022**, 1–13. [CrossRef] [PubMed]
68. Thodberg, H.H.; Kreiborg, S.; Juul, A.; Pedersen, K.D. The BoneXpert Method for Automated Determination of Skeletal Maturity. *IEEE Trans. Med. Imaging* **2008**, *28*, 52–66. [CrossRef]
69. Zhao, K.; Ma, S.; Sun, Z.; Liu, X.; Zhu, Y.; Xu, Y.; Wang, X. Effect of AI-assisted software on inter- and intra-observer variability for the X-ray bone age assessment of preschool children. *BMC Pediatr.* **2022**, *22*, 644. [CrossRef] [PubMed]
70. Booz, C.; Yel, I.; Wichmann, J.L.; Boettger, S.; Al Kamali, A.; Albrecht, M.H.; Martin, S.S.; Lenga, L.; Huizinga, N.; D'Angelo, T.; et al. 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**, *4*, 6–8. [CrossRef]
71. Zhang, S.-Y.; Liu, G.; Ma, C.-G.; Han, Y.-S.; Shen, X.-Z.; Xu, R.-L.; Thodberg, H.H. Automated Determination of Bone Age in a Modern Chinese Population. *ISRN Radiol.* **2013**, *2013*, 874570. [CrossRef]
72. Thodberg, H.H.; Säwendahl, L. Validation and Reference Values of Automated Bone Age Determination for Four Ethnicities. *Acad. Radiol.* **2010**, *17*, 1425–1432. [CrossRef]
73. Alshamrani, K.; Hewitt, A.; Offiah, A. Applicability of two bone age assessment methods to children from Saudi Arabia. *Clin. Radiol.* **2019**, *75*, 156.e1–156.e9. [CrossRef]
74. Klünder-Klünder, M.; Espinosa-Espindola, M.; Lopez-Gonzalez, D.; Loyo, M.S.-C.; Suárez, P.D.; Miranda-Lora, A.L. Skeletal Maturation in the Current Pediatric Mexican Population. *Endocr. Pract.* **2020**, *26*, 1053–1061. [CrossRef]
75. Thodberg, H.H.; Thodberg, B.; Ahlqvist, J.; Offiah, A.C. Autonomous artificial intelligence in pediatric radiology: The use and perception of BoneXpert for bone age assessment. *Pediatr. Radiol.* **2022**, *52*, 1338–1346. [CrossRef]
76. Schündeln, M.M.; Marschke, L.; Bauer, J.J.; Hauffa, P.K.; Schweiger, B.; Führer-Sakel, D.; Lahner, H.; Poeppel, T.D.; Kiewert, C.; Hauffa, B.P.; et al. 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. [CrossRef]
77. Willems, G. A review of the most commonly used dental age estimation techniques. *J. Forensic Odonto-Stomatol.* **2001**, *19*, 9–17.
78. Farhadian, M.; Salemi, F.; Saati, S.; Nafisi, N. Dental age estimation using the pulp-to-tooth ratio in canines by neural networks. *Imaging Sci. Dent.* **2019**, *49*, 19–26. [CrossRef] [PubMed]
79. Schour, I.; Massler, M. Studies in Tooth Development: The Growth Pattern of Human Teeth. *J. Am. Dent. Assoc.* **1940**, *27*, 1778–1793. [CrossRef]
80. Demirjian, A.; Goldstein, H.; Tanner, J.M. A new system of dental age assessment. *Hum. Biol.* **1973**, *45*, 211–227. [PubMed]
81. Khadiri, N.; Halilah, T.; Khandakji, M.N.; Jost-Brinkmann, P.-G.; Bartzela, T. The adaptation of Demirjian's dental age estimation method on North German children. *Forensic Sci. Int.* **2019**, *303*, 109927. [CrossRef]
82. Magon, P.; Viswanathan, V.K. "Bone Age", *Revision Classes in Pediatrics*; Jaypee Brothers Medical Publishers (P) Ltd.: New Delhi, India, 2008; p. 54. [CrossRef]
83. Aggarwal, A.; Kulkarni, S.; Sheikh, S.; Aggarwal, O.; Mehta, S.; Gupta, D. Correlation between Radiographic evaluation of dental age and chronological age: A study on 6 to 16 years human population of Ambala using Demirjian method. *J. Oral Sign* **2012**, *4*, 63–67.
84. Willems, G.; Van Olmen, A.; Spiessens, B.; Carels, C. Dental Age Estimation in Belgian Children: Demirjian's Technique Revisited. *J. Forensic Sci.* **2001**, *46*, 15064. [CrossRef]
85. Kvaal, S.I.; Kolltveit, K.M.; Thomsen, I.O.; Solheim, T. Age estimation of adults from dental radiographs. *Forensic Sci. Int.* **1995**, *74*, 175–185. [CrossRef]
86. Kvaal, S.; Solheim, T. A non-destructive dental method for age estimation. *J. Forensic Odonto-Stomatol.* **1994**, *12*, 6–11. Available online: <https://pubmed.ncbi.nlm.nih.gov/9227083/> (accessed on 15 January 2023).

87. Guo, Y.-C.; Han, M.; Chi, Y.; Long, H.; Zhang, D.; Yang, J.; Yang, Y.; Chen, T.; Du, S. Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. *Int. J. Leg. Med.* **2021**, *135*, 1589–1597. [CrossRef]
88. Vila-Blanco, N.; Carreira, M.J.; Varas-Quintana, P.; Balsa-Castro, C.; Tomas, I. Deep Neural Networks for Chronological Age Estimation From OPG Images. *IEEE Trans. Med. Imaging* **2020**, *39*, 2374–2384. [CrossRef]
89. Zaborowicz, K.; Biedziak, B.; Olszewska, A.; Zaborowicz, M. Tooth and Bone Parameters in the Assessment of the Chronological Age of Children and Adolescents Using Neural Modelling Methods. *Sensors* **2021**, *21*, 6008. [CrossRef]
90. Zaborowicz, M.; Zaborowicz, K.; Biedziak, B.; Garbowski, T. Deep Learning Neural Modelling as a Precise Method in the Assessment of the Chronological Age of Children and Adolescents Using Tooth and Bone Parameters. *Sensors* **2022**, *22*, 637. [CrossRef]
91. Kim, S.; Lee, Y.-H.; Noh, Y.-K.; Park, F.C.; Auh, Q.-S. Age-group determination of living individuals using first molar images based on artificial intelligence. *Sci. Rep.* **2021**, *11*, 1–11. [CrossRef] [PubMed]
92. Banjšak, L.; Milošević, D.; Subašić, M. Implementation of Artificial Intelligence in Chronological Age Estimation from Orthopantomographic X-ray Images of Archaeological Skull Remains. *Bull. Int. Assoc. Paleodont.* **2020**, *14*, 122–129. Available online: [www.paleodontology.com](http://www.paleodontology.com) (accessed on 24 February 2023).
93. Milošević, D.; Vodanović, M.; Galić, I.; Subašić, M. Automated estimation of chronological age from panoramic dental X-ray images using deep learning. *Expert Syst. Appl.* **2021**, *189*, 116038. [CrossRef]
94. Kahaki, S.M.; Nordin, M.J.; Ahmad, N.; Arzoky, M.; Ismail, W. Deep convolutional neural network designed for age assessment based on orthopantomography data. *Neural Comput. Appl.* **2020**, *32*, 21–22. [CrossRef]
95. Sauvegrain, J. Etude de la maturation osseuse du coude. *Ann. Radiol.* **1962**, *5*, 542–550. Available online: <https://cir.nii.ac.jp/crid/1572543025397534592> (accessed on 17 January 2023).
96. Hermetet, C.; Saint-Martin, P.; Gambier, A.; Ribier, L.; Sautenet, B.; Rérolle, C. Forensic age estimation using computed tomography of the medial clavicular epiphysis: A systematic review. *Int. J. Leg. Med.* **2018**, *132*, 1415–1425. [CrossRef] [PubMed]
97. Benito, M.; Muñoz, A.; Beltrán, I.; Labajo, E.; Perea, B.; Sánchez, J.A. Assessment of adulthood in the living Spanish population based on ossification of the medial clavicle epiphysis using ultrasound methods. *Forensic Sci. Int.* **2018**, *284*, 161–166. [CrossRef]
98. Shedje, R.; Kanchan, T.; Warriar, V.; Dixit, S.G.; Krishan, K. Forensic age estimation using conventional radiography of the medial clavicular epiphysis: A systematic review. *Med. Sci. Law* **2021**, *61*, 138–146. [CrossRef] [PubMed]
99. Schmeling, A.; Schulz, R.; Reisinger, W.; Mühler, M.; Wernecke, K.-D.; Geserick, G. Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography. *Int. J. Leg. Med.* **2004**, *118*, 5–8. [CrossRef] [PubMed]
100. Kreitner, K.-F.; Schweden, F.J.; Riepert, T.; Nafe, B.; Thelen, M. Bone age determination based on the study of the medial extremity of the clavicle. *Eur. Radiol.* **1998**, *8*, 1116–1122. [CrossRef] [PubMed]
101. Li, D.T.; Cui, J.J.; DeVries, S.; Nicholson, A.D.; Li, E.; Petit, L.; Kahan, J.B.; Sanders, J.O.; Liu, R.W.; Cooperman, D.R.; et al. Humeral Head Ossification Predicts Peak Height Velocity Timing and Percentage of Growth Remaining in Children. *J. Pediatr. Orthop.* **2018**, *38*, e546–e550. [CrossRef] [PubMed]
102. Bitan, F.D.; Veliskakis, K.P.; Campbell, B.C. Differences in the Risser Grading Systems in the United States and France. *Clin. Orthop. Relat. Res.* **2005**, *436*, 190–195. [CrossRef]
103. Lottering, N.; Alston-Knox, C.L.; MacGregor, D.M.; Izatt, M.T.; Grant, C.A.; Adam, C.J.; Gregory, L.S. Apophyseal Ossification of the Iliac Crest in Forensic Age Estimation: Computed Tomography Standards for Modern Australian Subadults. *J. Forensic Sci.* **2016**, *62*, 292–307. [CrossRef]
104. Schmidt, S.; Schmeling, A.; Zwiesigk, P.; Pfeiffer, H.; Schulz, R. Sonographic evaluation of apophyseal ossification of the iliac crest in forensic age diagnostics in living individuals. *Int. J. Leg. Med.* **2011**, *125*, 271–276. [CrossRef]
105. Rhee, C.H.; Shin, S.M.; Choi, Y.S.; Yamaguchi, T.; Maki, K.; Kim, Y.I.; Kim, S.S.; Park, S.B.; Son, W.S. Application of statistical shape analysis for the estimation of bone and forensic age using the shapes of the 2nd, 3rd, and 4th cervical vertebrae in a young Japanese population. *Forensic Sci. Int.* **2015**, *257*, 513.e1–513.e9. [CrossRef]
106. Lai, E.H.-H.; Liu, J.-P.; Chang, J.Z.-C.; Tsai, S.-J.; Yao, C.-C.J.; Chen, M.-H.; Chen, Y.-J.; Lin, C.-P. Radiographic Assessment of Skeletal Maturation Stages for Orthodontic Patients: Hand-Wrist Bones or Cervical Vertebrae? *J. Formos. Med. Assoc.* **2008**, *107*, 316–325. [CrossRef]
107. Schlégl, T.; O’Sullivan, I.; Varga, P.; Than, P.; Vermes, C. Determination and correlation of lower limb anatomical parameters and bone age during skeletal growth (based on 1005 cases). *J. Orthop. Res.* **2016**, *35*, 1431–1441. [CrossRef]
108. Li, S.Q.; Nicholson, A.D.; Cooperman, D.R.; Liu, R.W. Applicability of the Calcaneal Apophysis Ossification Staging System to the Modern Pediatric Population. *J. Pediatr. Orthop.* **2019**, *39*, 46–50. [CrossRef]
109. O’Connor, J.E.; Bogue, C.; Spence, L.D.; Last, J. A method to establish the relationship between chronological age and stage of union from radiographic assessment of epiphyseal fusion at the knee: An Irish population study. *J. Anat.* **2008**, *212*, 198–209. [CrossRef] [PubMed]
110. Krämer, J.A.; Schmidt, S.; Jürgens, K.-U.; Lentschig, M.; Schmeling, A.; Vieth, V. Forensic age estimation in living individuals using 3.0T MRI of the distal femur. *Int. J. Leg. Med.* **2014**, *128*, 509–514. [CrossRef]
111. Risser, J.C. The Classic: The Iliac Apophysis: An Invaluable Sign in the Management of Scoliosis. *Clin. Orthop. Relat. Res.* **2009**, *468*, 646–653. [CrossRef]

112. Baik, S.B.; Cha, K.G. A Study on Deep Learning Based Sauvegrain Method for Measurement of Puberty Bone Age. *arXiv* **2018**, arXiv:1809.06965.
113. Der Mauer, M.A.; Well, E.J.-V.; Herrmann, J.; Groth, M.; Morlock, M.M.; Maas, R.; Säring, D. Automated age estimation of young individuals based on 3D knee MRI using deep learning. *Int. J. Leg. Med.* **2020**, *135*, 649–663. [[CrossRef](#)]
114. Dallora, A.L.; Berglund, J.S.; Brogren, M.; Kvist, O.; Ruiz, S.D.; Dübbel, A.; Anderberg, P. Age Assessment of Youth and Young Adults Using Magnetic Resonance Imaging of the Knee: A Deep Learning Approach. *JMIR Public Health Surveill.* **2019**, *7*, e16291. [[CrossRef](#)] [[PubMed](#)]
115. Stern, D.; Payer, C.; Giuliani, N.; Urschler, M. Automatic Age Estimation and Majority Age Classification From Multi-Factorial MRI Data. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1392–1403. [[CrossRef](#)] [[PubMed](#)]
116. Liao, N.; Dai, J.; Tang, Y.; Zhong, Q.; Mo, S. iCVM: An Interpretable Deep Learning Model for CVM Assessment Under Label Uncertainty. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4325–4334. [[CrossRef](#)]
117. Kim, D.; Kim, J.; Kim, T.; Kim, T.; Kim, Y.; Song, I.; Ahn, B.; Choo, J.; Lee, D. Prediction of hand-wrist maturation stages based on cervical vertebrae images using artificial intelligence. *Orthod. Craniofacial Res.* **2021**, *24*, 68–75. [[CrossRef](#)] [[PubMed](#)]
118. Peng, L.; Wan, L.; Wang, M.W.; Li, Z.; Wang, P.; Liu, T.A.; Wang, Y.H.; Zhao, H. Comparison of Three CNN Models Applied in Bone Age Assessment of Pelvic Radiographs of Adolescents. *J. Forensic Med.* **2020**, *36*, 622–630. [[CrossRef](#)]
119. Peng, L.-Q.; Guo, Y.-C.; Wan, L.; Liu, T.-A.; Wang, P.; Zhao, H.; Wang, Y.-H. Forensic bone age estimation of adolescent pelvis X-rays based on two-stage convolutional neural network. *Int. J. Leg. Med.* **2022**, *136*, 797–810. [[CrossRef](#)]
120. Chen, H.C.; Wu, C.H.; Lin, C.J.; Liu, Y.H.; Sun, Y.N. Automated segmentation for patella from lateral knee X-ray images. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 3553–3556. [[CrossRef](#)]
121. Xia, Y.; Fripp, J.; Chandra, S.S.; Schwarz, R.; Engstrom, C.; Crozier, S. Automated bone segmentation from large field of view 3D MR images of the hip joint. *Phys. Med. Biol.* **2013**, *58*, 7375–7390. [[CrossRef](#)]
122. Vicini, S.; Bortolotto, C.; Rengo, M.; Ballerini, D.; Bellini, D.; Carbone, I.; Preda, L.; Laghi, A.; Coppola, F.; Faggioni, L. A narrative review on current imaging applications of artificial intelligence and radiomics in oncology: Focus on the three most common cancers. *La Radiol. Med.* **2022**, *127*, 819–836. [[CrossRef](#)]
123. Langlotz, C.P. Will Artificial Intelligence Replace Radiologists? *Radiol. Artif. Intell.* **2019**, *1*, e190058. [[CrossRef](#)] [[PubMed](#)]
124. Ontell, F.K.; Ivanovic, M.; Ablin, D.S.; Barlow, T.W. Bone age in children of diverse ethnicity. *Am. J. Roentgenol.* **1996**, *167*, 1395–1398. [[CrossRef](#)] [[PubMed](#)]
125. Sardanelli, F.; Colarieti, A. Open issues for education in radiological research: Data integrity, study reproducibility, peer-review, levels of evidence, and cross-fertilization with data scientists. *La Radiol. Med.* **2022**, *128*, 133–135. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.