

Applicability and robustness of an artificial intelligence-based assessment for Greulich and Pyle bone age in a German cohort

Anwendbarkeit und Robustheit einer auf künstlicher Intelligenz basierenden Analyse des Knochenalters nach Greulich und Pyle in einer deutschen Kohorte

Authors

Johanna Pape¹ , Franz Wolfgang Hirsch¹, Oliver Johannes Deffaa², Matthew D. DiFranco³, Maciej Rosolowski⁴, Daniel Gräfe¹

Affiliations

- 1 Pediatric Radiology, University Hospital Leipzig, Germany
- 2 Pediatric Surgery, University Hospital Leipzig, Germany
- 3 ImageBiopsy Lab, ImageBiopsy Lab, Vienna, Austria
- 4 Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Leipzig, Germany

Keywords

children, radiography, artificial intelligence, bone age, Greulich&Pyle

received 9.6.2023

accepted 4.10.2023

published online 8.12.2023

Bibliography

Fortschr Röntgenstr 2024; 196: 600–606

DOI 10.1055/a-2203-2997

ISSN 1438-9029

© 2023. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence

Dr. Johanna Pape

Pediatric Radiology, University Hospital Leipzig, Liebigstraße 20a, 04103 Leipzig, Germany

Tel.: +49/3 41/9 72 05 08

johanna.pape@medizin.uni-leipzig.de



Supplementary material is available under <https://doi.org/10.1055/a-2203-2997>

ABSTRACT

Purpose The determination of bone age (BA) based on the hand and wrist, using the 70-year-old Greulich and Pyle (G&P) atlas, remains a widely employed practice in various institutions today. However, a more recent approach utilizing artificial intelligence (AI) enables automated BA estimation based on the G&P atlas. Nevertheless, AI-based methods encounter limitations when dealing with images that deviate from the standard hand and wrist projections. Generally, the extent to which BA, as determined by the G&P atlas, corre-

sponds to the chronological age (CA) of a contemporary German population remains a subject of continued discourse. This study aims to address two main objectives. Firstly, it seeks to investigate whether the G&P atlas, as applied by the AI software, is still relevant for healthy children in Germany today. Secondly, the study aims to assess the performance of the AI software in handling non-strict posterior-anterior (p. a.) projections of the hand and wrist.

Materials and Methods The AI software retrospectively estimated the BA in children who had undergone radiographs of a single hand using posterior-anterior and oblique planes. The primary purpose was to rule out any osseous injuries. The prediction error of BA in relation to CA was calculated for each plane and between the two planes.

Results A total of 1253 patients (aged 3 to 16 years, median age 10.8 years, 55.7 % male) were included in the study. The average error of BA in posterior-anterior projections compared to CA was 3.0 (\pm 13.7) months for boys and 1.7 (\pm 13.7) months for girls. Interestingly, the deviation from CA tended to be even slightly lower in oblique projections than in posterior-anterior projections. The mean error in the posterior-anterior projection plane was 2.5 (\pm 13.7) months, while in the oblique plane it was 1.8 (\pm 13.9) months (p = 0.01).

Conclusion The AI software for BA generally corresponds to the age of the contemporary German population under study, although there is a noticeable prediction error, particularly in younger children. Notably, the software demonstrates robust performance in oblique projections.

Key Points

1. Bone age, as determined by artificial intelligence, aligns with the chronological age of the contemporary German cohort under study.
2. As determined by artificial intelligence, bone age is remarkably robust, even when utilizing oblique X-ray projections.

Citation Format

- Pape J, Hirsch F, Deffaa O et al. Applicability and robustness of an artificial intelligence-based assessment for Greulich and Pyle bone age in a German cohort. *Fortschr Röntgenstr* 2024; 196: 600–606

ZUSAMMENFASSUNG

Ziel Die Bestimmung des Knochenalters (BA) anhand der Hand und des Handgelenks unter Verwendung des 70 Jahre alten Atlas von Greulich und Pyle (G&P) ist auch heute noch eine weit verbreitete Praxis in verschiedenen Einrichtungen. Ein neuerer Ansatz, basierend auf dem Einsatz künstlicher Intelligenz (KI), ermöglicht eine automatische BA-Schätzung auf der Grundlage des G&P-Atlas. Allerdings stoßen KI-basierte Methoden an ihre Grenzen, wenn es um Bilder geht, die von den Standardprojektionen der Hand und des Handgelenks abweichen. Nach wie vor ist umstritten, inwieweit das mit dem G&P-Atlas ermittelte BA dem chronologischen Alter (CA) der heutigen deutschen Bevölkerung entspricht. Mit dieser Studie werden zwei Hauptziele verfolgt. Erstens soll untersucht werden, ob der G&P-Atlas, wie er von der KI-Software angewendet wird, für gesunde Kinder in Deutschland heute noch relevant ist. Zweitens zielt die Studie darauf ab, die Leistung der KI-Software bei der Handhabung nicht-strikter posterior-anteriorer (p. a.) Projektionen der Hand und des Handgelenks zu bewerten.

Materialien und Methoden Die AI-Software schätzte retrospektiv die BA bei Kindern, die sich Röntgenaufnahmen einer einzelnen Hand unter Verwendung von posterior-anterioren und schrägen Ebenen unterzogen hatten. Der Hauptzweck bestand darin, knöcherne Verletzungen auszuschließen. Der Vorhersagefehler des BA im Verhältnis zum CA wurde für jede Ebene und zwischen den beiden Ebenen berechnet.

Ergebnisse Insgesamt wurden 1253 Patienten (im Alter von 3 bis 16 Jahren, medianes Alter 10,8 Jahre, 55,7 % männlich) in die Studie aufgenommen. Die durchschnittliche Abweichung des BA in posterior-anterioren Projektionen im Vergleich zum CA betrug bei Jungen 3,0 ($\pm 13,7$) Monate und 1,7 ($\pm 13,7$) Monate bei Mädchen. Interessanterweise war die Abweichung des BA vom CA in den schrägen Projektionen tendenziell etwas geringer als in den posterior-anterioren Projektionen. Der mittlere Fehler in der posterior-anterioren Projektionsebene betrug 2,5 ($\pm 13,7$) Monate, während er in der schrägen Ebene bei 1,8 ($\pm 13,9$) Monaten lag ($p = 0,01$).

Schlussfolgerung Das mittels KI-Software ermittelte BA entspricht im Allgemeinen dem Alter der deutschen Untersuchungspopulation, obwohl es einen merklichen Vorhersagefehler gibt, insbesondere bei jüngeren Kindern. Insbesondere bei schrägen Projektionen zeigt die Software eine robuste Leistung.

Kernaussagen

1. Das von der künstlichen Intelligenz ermittelte Knochenalter stimmt mit dem chronologischen Alter der untersuchten deutschen Alterskohorte überein.
2. Das durch künstliche Intelligenz ermittelte Knochenalter ist bemerkenswert stabil, auch bei der Verwendung schräger Röntgenprojektionen.

Introduction

Determining bone age (BA) holds significant importance in the clinical evaluation of childhood growth and maturation [1]. In clinical practice, BA is a standardized parameter for diagnosing and monitoring pediatric endocrine diseases, metabolic conditions, and growth disorders and is also used for legal and forensic age determination [1–3]. The assessment of BA relies on the typical sequence of ossification in the hand and wrist over time [1, 3]. The determination of hand BA predominantly depends on the Greulich and Pyle (G&P) method [4, 5]. The technique compares age-specific developmental markers on hand and wrist X-rays with reference images from the G&P atlas, categorized by age and gender [1, 4]. While the G&P method is easier to implement and faster in clinical practice than alternatives like Tanner and Whitehouse [6], it does exhibit susceptibility to significant inter and intra-observer variability [1, 6, 7].

The G&P atlas originated from the analysis of bone ages in North American children from 1931 to 1942 [4]. In recent decades, an emerging trend towards earlier skeletal maturation among children has been attributed to improved socioeconomic conditions and better healthcare and nutrition [8, 9]. Consequently, questions have arisen regarding the applicability of the G&P atlas to the skeletal maturation of modern children and its

suitability as a reference. Various studies have already noted potential disparities between G&P-based BA and chronological age (CA) [2, 10, 11], along with indications of variations across genders and ethnicities [2, 9, 12]. However, many of these studies had small sample sizes or other limitations [2, 9].

To enhance the precision and objectivity of BA assessment, the integration of artificial intelligence (AI) has gained prominence in clinical practice [13–15]. Numerous studies have demonstrated the accuracy and efficiency of AI [14–17]. Remarkably, the fully automated software IB Lab PANDA (IB Lab GmbH, Vienna, Austria) has proven reliable in providing BA data. Notably, the accuracy of IB Lab PANDA has shown no significant differences compared to assessments conducted by experienced pediatric radiologists [18]. Conventionally, strict p. a. projections of the hand and wrist are used for BA determination [1]. However, AI-based software encounters limitations when interpreting images that deviate from the standard position or exhibit altered bone morphology [19].

This study addresses whether the G&P atlas, as interpreted by the AI software, remains applicable to contemporary healthy children in Germany. Its secondary aim is to quantify the AI software's capability to handle non-strict p. a. projections of the hand and wrist.

Materials and methods

Patients

The retrospective study was conducted with patients who had undergone a hand X-ray between 2012 and 2022 at an anonymous hospital. Ethical approval for the retrospective evaluation of the study was obtained from the local ethics committee. Patients ranging from birth to 18 years old were identified using the hospital's Picture Archiving and Communication System (PACS).

Image selection

Patients with a known bone age were excluded. Only cases with both p. a. and oblique views of the hand were considered. These radiographs were primarily taken to assess trauma sequelae. If multiple X-rays were available for a patient at different times, only one was included. While both the right and left hands were eligible, the image of the left hand was selected if both hands were X-rayed during the same visit. Exclusion criteria included traumatic injuries (fractures and dislocations), deformities (polydactyly and syndactyly), and technically suboptimal image quality, such as incomplete hand depiction due to overlays, e. g., overlying dressing material, or only one radiographic projection, p. a. or oblique, was available. Images showing pathological changes, such as abnormal bone texture or masses, were also excluded. The exact number of subjects included and excluded is shown in ► **Fig. 1**. Radiographs were screened by two radiologists (blinded) with 15 and 3 years of experience in pediatric radiology, respectively. A total of 1703 patients with two radiographs (p. a. and oblique) devoid of pathological findings were identified.

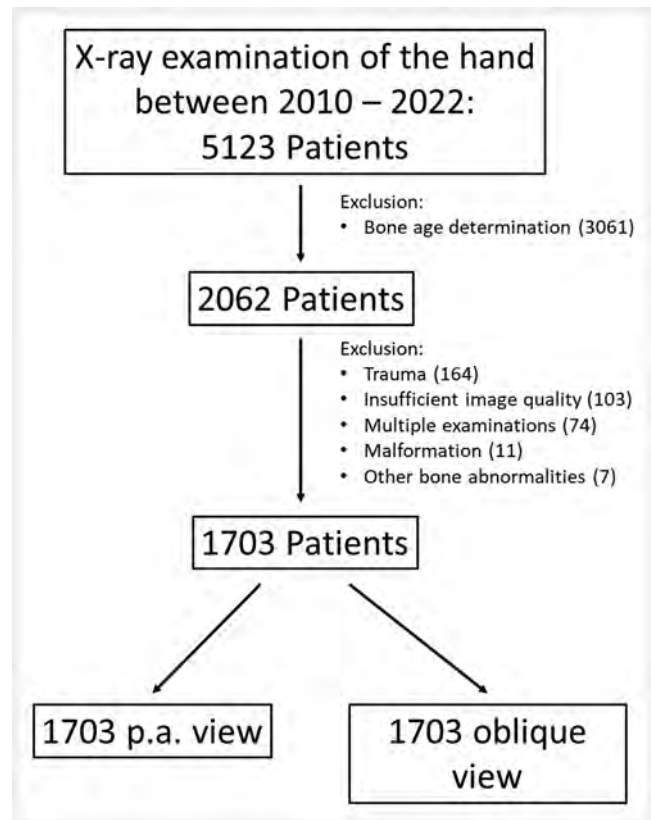
AI model for automated bone age assessment

The BA for p. a. and oblique images was automatically determined separately using the Conformité Européene CE-marked commercial software IB Lab PANDA software (version 1.06), designed to assess hand radiographs according to the G&P method. IB Lab PANDA is intended for girls aged 36 to 192 months and boys aged 36 to 204 months, based on CA at the time of radiograph acquisition. The software generates a graphical display of the BA rounded to the nearest month, among other outputs. A secondary capture of the input radiograph designates the region analyzed by the software and is used for visual inspection.

Automated AI analysis of the radiographic images was performed through an internal clinical pipeline by the installation of the containerization software Docker (Docker Inc., Palo Alto, CA) containing IB Lab ZOO v.1.13.21 on a dedicated standalone PC configured as a PACS sending and receiving node.

Statistics

The test for normal distribution of the residuals was performed using a Shapiro-Wilk test and visual QQ plot analysis. The mean error, mean absolute error (MAE), and standard deviation of the prediction error were calculated to measure the discrepancies between BA in p. a. and oblique views and CA. The statistical significance level was set at 0.05. RStudio 2022.07.2 (PBC, Boston, MA) was used for statistical analysis.



► **Fig. 1** Inclusion and exclusion criteria with the respective absolute number of patients.

► **Abb. 1** Ein- und Ausschlusskriterien mit Angabe der jeweiligen absoluten Patient*innenanzahl.

Results

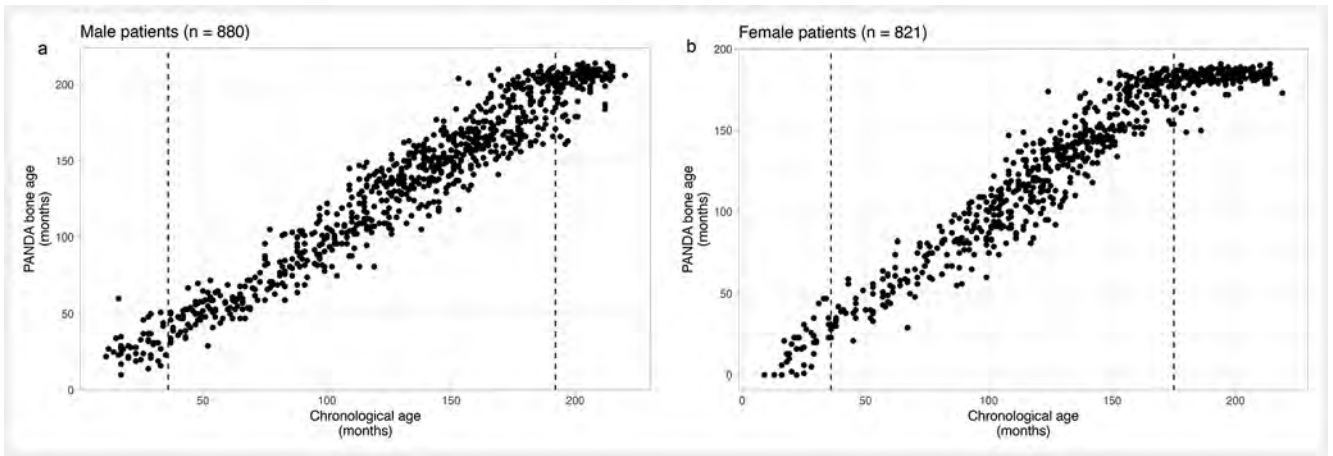
Patient cohort

A total of 1703 patients with two radiographs (p. a. and oblique) free from pathological findings were initially identified (► **Fig. 2**). Patients younger than 36 months were excluded due to the AI software's usage restrictions. The recommended upper threshold for applying the AI software (204 months for boys and 192 months for girls) was adjusted based on the plotted distribution to 192 months for boys and 175 months for girls.

As a result, 1253 patients with a median age of 130 months (IQR 100–155, 55.7 % male) were retained for subsequent analysis (► **Fig. 3**).

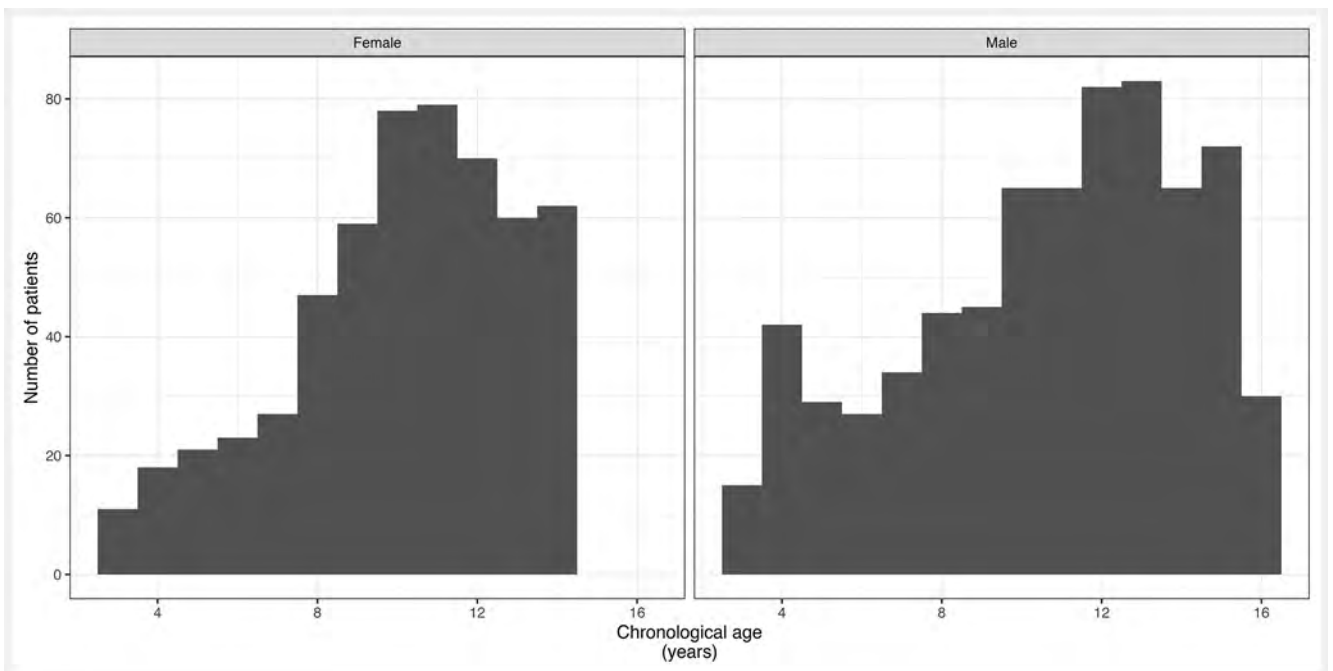
Deviation of bone age from chronological age

The AI software exhibited a mean prediction error of (3.0 ± 13.7) (standard deviation of the prediction error) months in boys and (1.7 ± 13.7) (standard deviation of the prediction error) months in girls. The CA of patients tended to be underestimated by the AI software in boys below eight years of age and overestimated above that age. For girls, the age cutoff was approximately ten years (► **Fig. 4**). The residuals between BA and CA displayed a normal distribution for males and females (**Supplementary Material 1**).



► **Fig. 2** IB Lab PANDA bone age related to chronological age in males (a) and females (b). According to IB Lab PANDA's intended use population and distribution of data, gender-specific upper and lower limits were set (dashed lines). **a** There is no more correlation between BA and CA in boys after age 16. **b** In girls, the correlation between BA and CA already ends at about 14.5 years.

► **Abb. 2** Knochenalter nach IBLab PANDA im Vergleich zum chronologischen Alter bei Jungen (a) und Mädchen (b). Entsprechend der Zulassung von IBLab PANDA und der Altersverteilung der Daten wurden geschlechtsspezifische Ober- und Untergrenzen festgelegt (gestrichelte Linien). **a** Bei Jungen besteht nach dem 16. Lebensjahr keine Korrelation zwischen BA und CA. **b** Bei Mädchen endet die Korrelation zwischen BA und CA bei etwa 14,5 Jahren.



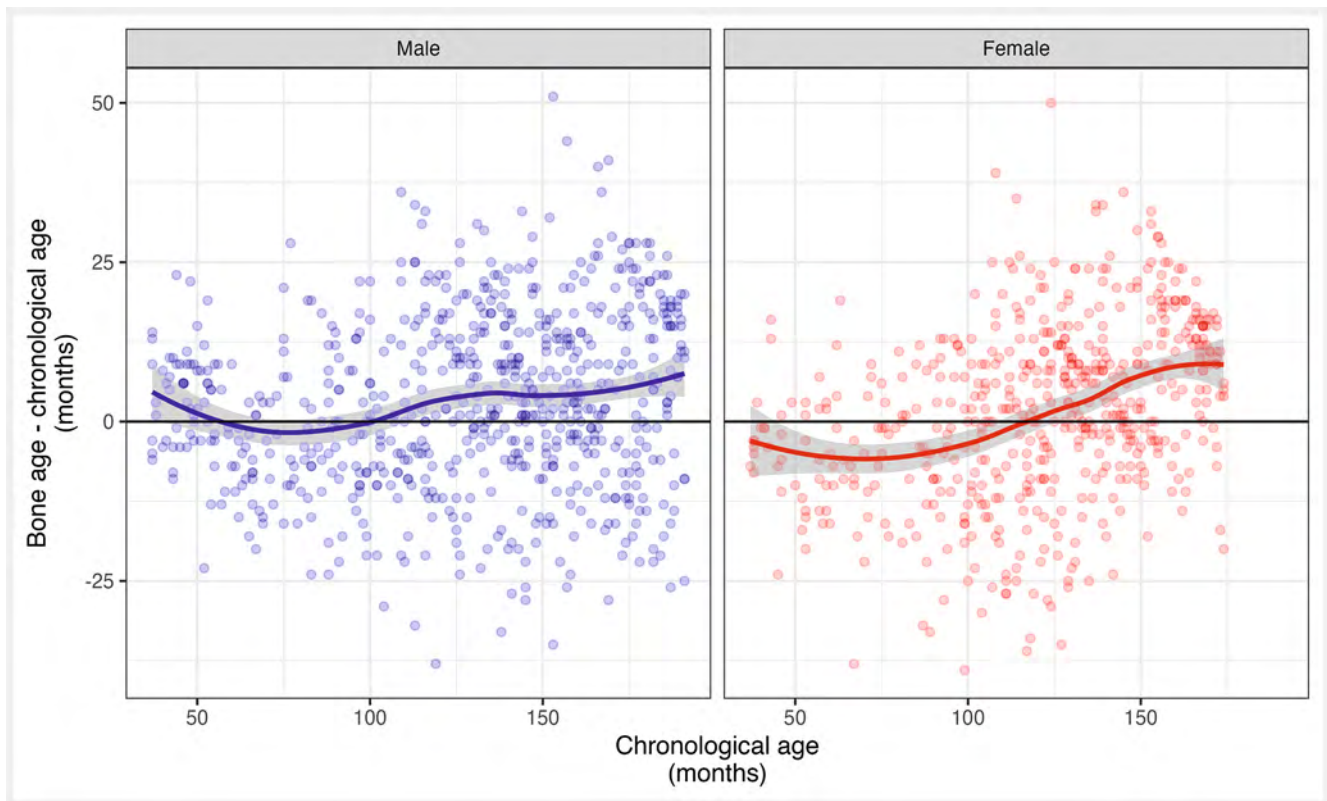
► **Fig. 3** Histogram illustrating the age of the 1253 patients classified by gender. The data does not adhere to normal distribution because it is not present.

► **Abb. 3** Histogramm der Altersverteilung der 1253 Patient*innen, getrennt nach Geschlecht. Eine Normalverteilung liegt nicht vor.

Impact of Oblique Projection on AI Bone Age

Concerning the prediction error regarding CA, the AI software's determination of BA using oblique images showed minimal deviation from that derived from p. a. images (► **Fig. 5**). For the entire cohort, the MAE was 11.1 months for p. a. images and 11.0 months for the oblique images ($p < 0.38$, ► **Table 1**). Notably,

oblique projections in girls demonstrated an even lower error than p. a. projections ($p < 0.001$, ► **Table 1**). The variance of BA between oblique and p. a. projections was less than that between oblique BA and CA (**Supplementary Material 2**).



► **Fig. 4** Gender-specific prediction error of the AI software bone age from chronological age. Solid lines: mean error, shaded area: 95 % confidence interval. Under the age of 8, BA, as estimated by the AI software, tends to be lower than CA. BA is higher than CA from age 8 to 10.

► **Abb. 4** Geschlechtsspezifischer Vorhersagefehler des Knochenalters nach der KI-Software zum chronologischen Alter. Die Linien entsprechen dem mittleren Fehler, die schattierte Fläche dem 95 %-Konfidenzintervall. Im Alter von unter 8 Jahren ist das von der KI-Software bestimmte BA tendenziell niedriger als das CA. Im Alter von 8 bis 10 Jahren ist das BA höher als das CA.

► **Table 1** Variation of residuals (given in months) between the AI software bone age and chronological age, each based on either a p. a. scan or an oblique scan. Comparison between variation of p. a. and oblique, significance level was set at $p < 0.05$. In girls, the mean error and mean absolute error are significantly lower for oblique projections than for p. a. projections.

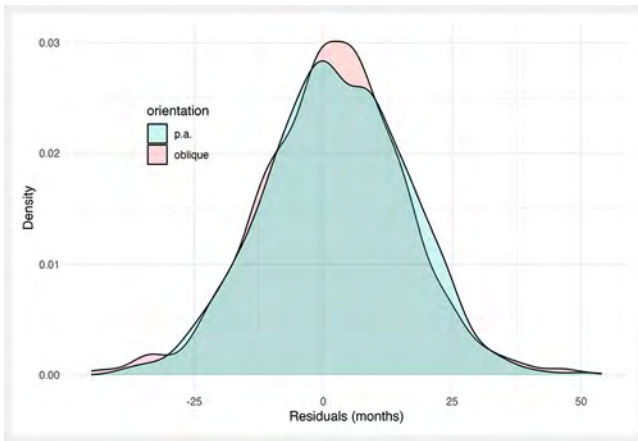
► **Tab. 1** Abweichung der Residuen zwischen dem KI-Software-Knochenalter und dem chronologischen Alter, jeweils basierend auf der p. a.-Aufnahme und der Schrägaufnahme. Vergleich der Abweichung von p. a. und oblique, signifikante Unterschiede bestanden bei $p < 0,05$. Der mittlere Fehler und der mittlere absolute Fehler sind bei schräger Projektion nicht höher als bei p. a.-Projektion. Bei Mädchen sind diese bei schräger Projektion signifikant geringer als bei p. a.-Projektion.

	Overall			Female			Male		
	<i>p. a.</i>	<i>oblique</i>	<i>p-value</i>	<i>p. a.</i>	<i>oblique</i>	<i>p-value</i>	<i>p. a.</i>	<i>oblique</i>	<i>p-value</i>
Mean prediction error	2.5	1.8	0.01	1.7	0.3	<0.001	3.0	3.1	0.91
Mean absolute prediction error	11.1	11.0	0.38	11.0	9.9	<0.001	11.2	11.8	0.08
Standard deviation of the prediction error	13.7	13.9		13.7	12.9		13.7	14.6	

Discussion

This study assessed the applicability of a novel automated AI interpretation of the G&P atlas to healthy children in present-day Germany while also investigating the effect of oblique X-ray projections on the estimated BA results.

The G&P atlas, introduced in the 1950s, outlines skeletal maturation stages in children from that era [4]. Multiple studies have highlighted the continued utility of the G&P atlas for age determination in modern times. Nevertheless, some studies have shown that BA tends to be more advanced than CA, particularly



► **Fig. 5** Difference (residuals) between bone age (each determined in p. a. and oblique projection via AI software) and chronological age, shown as a smoothed histogram of frequencies (density). The agreement between oblique projection and chronological age is not lower than that between p. a. projection and chronological age.

► **Abb. 5** Differenz (Residuals) zwischen Knochenalter (jeweils in p. a. und schräger Projektion, bestimmt durch KI-Software) und chronologischem Alter, dargestellt als geglättetes Histogramm der Häufigkeiten (Density). Die Übereinstimmung zwischen schräger Projektion und chronologischem Alter ist nicht geringer als die Übereinstimmung zwischen p. a. Projektion und chronologischem Alter.

in pubertal children [20, 21], due to accelerated skeletal maturation [8, 22]. Schmidt et al. recommended the Thiemann-Nitz method over the G&P method due to a possible age overestimation [23].

In our study, no significant differences were observed between the AI software BA and CA across the age range evaluated. However, the AI software showed a tendency to underestimate the patient's age before puberty and overestimate it after. This agrees with Hwang et al., who found that deep learning-based software tended to estimate BA lower in younger children and higher in older children [10]. The deviation of BA from CA, especially in older boys, might stem from the methodological effects of the G&P atlas [9]. Notably, the atlas's annual radiographs conclude at 18 years for girls and 19 years for boys [4]. However, as those ages do not mark the end of skeletal maturity, higher BAs, no longer represented by the atlas with their stage of skeletal maturity, are assigned to the last radiograph [9], potentially leading to lower BA estimations and homogeneity in older adolescents [9]. This study, however, focused on the AI software's intended age range.

The mean error in our study was 2.5 months, which was well below the natural standard deviation of the BA. This is already higher at the age of 1.5 years and reaches up to 15.4 months in adolescence [4]. However, for intra-individual prediction, the MAE is more relevant than the mean error. The MAE between AI-estimated BA and CA was 11.1 months in our patient cohort and thus higher than the natural standard deviation in children below ten years of age, implying clinical significance. However, the AI software's MAE is lower when compared to human expert readers' assessments. The software BoneXpert, as an example,

achieved an MAE of 4.1 months [14]; another BA algorithm (using the Tanner-Whitehouse method instead of G&P) reached as low as 0.2 months [24].

In contrast, some studies show differences between BA and CA in boys in Asia and girls in Africa [2, 9, 25]. It should be noted that African studies are underrepresented in relation to countries with a high socioeconomic status [9]. A large proportion of the patients in the cohort of this study were of Caucasian descent. In other ethnicities, such as Asian boys and African girls, discrepancies between BA and CA were demonstrated [2, 12]. Overall, it is still unclear to what extent ethnicity [25, 26] and socioeconomic factors affect BA [2, 9, 27].

The impressive robustness of the automatic BA determination against deviations from the strict p. a. hand projection is remarkable. In fact, in some instances, the variation in BA compared to CA was either on par or even slightly lower in the oblique projection than in the p. a. projection. This observation offers some reassurance when dealing with slightly tilted images or instances of less than fully extendable fingers.

A potential limitation of the AI software employed in this study for bone age evaluation is its inability to distinguish between appropriate and inappropriate input images. Striking a balance between achieving the lowest possible rejection rate while minimizing the risk of erroneous outcomes due to inadequate data is a recognized challenge in AI-assisted diagnosis [28]. The input verification of the software used in this study is limited to analyzing DICOM headers, thus enabling us to analyze oblique hand projections. This underscores the necessity of human validation to confirm proper input data for a complete hand in p. a. projection since the downstream processing up to the output of the bone age is not transparent, similar to a "black box" [29]. This underlines the relevance of one of our results that an oblique projection of the hand does not affect reliability compared to a p. a. projection.

The study has several limitations due to its retrospective nature. On the one hand, it cannot be ruled out that the study population might also include patients with growth disorders. Nevertheless, assuming a normal distribution of growth in the cohort, it would affect approximately 62 out of the 1253 patients. Additionally, it cannot be ensured that the ethnic distribution of the collective is representative of Germany. It should be noted that the size of our patient population is significantly larger than in most previous studies [2, 23]. While the patients' age distribution is not uniform, this factor should have minimal impact on the current statistical evaluation given the substantial patient cohort. Finally, it is essential to acknowledge that the BA estimated by the AI software may show some variance compared to the "true" BA of the G&P atlas.

In summary, according to the G&P atlas and as estimated by the AI software, BA is very similar to a contemporary German population on average. However, depending on age, the individual prediction error may exceed the natural standard deviation. Notably, the determination of BA by the AI software demonstrates remarkable resilience to non-standard p. a. X-ray projections.

CLINICAL RELEVANCE OF THE STUDY

1. The bone age estimation conducted through AI, following the Greulich and Pyle methodology, remains in correspondence with the chronological age of a contemporary German cohort.
2. However, the prediction error between BA and CA does, in certain cases, surpass the inherent natural standard deviation of bone age.
3. The AI software consistently produces reliable results, even for oblique projections.

Conflict of Interest

Matthew DiFranco was an employee of IB Lab GmbH. The other authors declare no conflicts of interest.

References

- [1] Satoh M. Bone age: assessment methods and clinical applications. *Clin Pediatr Endocrinol* 2015; 24: 143–152. doi:10.1297/cpe.24.143
- [2] Alshamrani K, Messina F, Offiah AC. Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. *Eur Radiol* 2019; 29: 2910–2923. doi:10.1007/s00330-018-5792-5
- [3] Manzoor Mughal A, Hassan N, Ahmed A. Bone Age Assessment Methods: A Critical Review. *Pak J Med Sci* 2014; 30: 211–215. doi:10.12669/pjms.301.4295
- [4] Greulich and Pyle. Radiographic Atlas of Skeletal Development of the Hand and Wrist. *Am J Med Sci* 1959; 238 (3): 393. doi:10.1097/0000441-195909000-00030
- [5] Breen MA, Tsai A, Stamm A et al. Bone age assessment practices in infants and older children among Society for Pediatric Radiology members. *Pediatr Radiol* 2016; 46: 1269–1274. doi:10.1007/s00247-016-3618-7
- [6] Bull RK, Edwards PD, Kemp PM et al. 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. doi:10.1136/adc.81.2.172
- [7] King DG, Steventon DM, O'Sullivan MP et al. Reproducibility of bone ages when performed by radiology registrars: an audit of Tanner and Whitehouse II versus Greulich and Pyle methods. *Br J Radiol* 1994; 67: 848–851. doi:10.1259/0007-1285-67-801-848
- [8] Boeyer ME, Sherwood RJ, Deroche CB et al. Early Maturity as the New Normal: A Century-long Study of Bone Age. *Clin Orthop Relat Res* 2018; 476: 2112–2122. doi:10.1097/CORR.0000000000000446
- [9] Dahlberg PS, Mosdøl A, Ding Y et al. A systematic review of the agreement between chronological age and skeletal age based on the Greulich and Pyle atlas. *Eur Radiol* 2019; 29: 2936–2948. doi:10.1007/s00330-018-5718-2
- [10] Hwang J, Yoon HM, Hwang J-Y et al. Re-Assessment of Applicability of Greulich and Pyle-Based Bone Age to Korean Children Using Manual and Deep Learning-Based Automated Method. *Yonsei Med J* 2022; 63: 683–691. doi:10.3349/yjm.2022.63.7.683
- [11] Kim JR, Lee YS, 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. doi:10.3348/kjr.2015.16.1.201
- [12] Ontell FK, Ivanovic M, Ablin DS et al. Bone age in children of diverse ethnicity. *Am J Roentgenol* 1996; 167: 1395–1398. doi:10.2214/ajr.167.6.8956565
- [13] Thodberg HH, Kreiborg S, Juul A et al. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging* 2009; 28: 52–66. doi:10.1109/TMI.2008.926067
- [14] Booz C, Yel I, Wichmann JL 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. doi:10.1186/s41747-019-0139-9
- [15] Thodberg HH, Thodberg B, Ahlkvist J et al. Autonomous artificial intelligence in pediatric radiology: the use and perception of BoneXpert for bone age assessment. *Pediatr Radiol* 2022; 52: 1338–1346. doi:10.1007/s00247-022-05295-w
- [16] Martin DD, Calder AD, Ranke MB et al. Accuracy and self-validation of automated bone age determination. *Sci Rep* 2022; 12. doi:10.1038/s41598-022-10292-y
- [17] Thodberg HH. Clinical review: An automated method for determination of bone age. *J Clin Endocrinol Metab* 2009; 94: 2239–2244. doi:10.1210/jc.2008-2474
- [18] DiFranco M, Chung TS, Mintz A et al. Automated Bone Age Assessment Across Multi-site U.S. Study: Agreement between AI and Expert Readers. *Semin Musculoskelet Radiol* 2022; 26: A117. doi:10.1055/s-0042-1750637
- [19] Offiah AC. Current and emerging artificial intelligence applications for pediatric musculoskeletal radiology. *Pediatr Radiol* 2022; 52: 2149–2158. doi:10.1007/s00247-021-05130-8
- [20] Calfee RP, Sutter M, Steffen JA et al. Skeletal and chronological ages in American adolescents: current findings in skeletal maturation. *J Child Orthop* 2010; 4: 467–470. doi:10.1007/s11832-010-0289-z
- [21] Hackman L, Black S. The reliability of the Greulich and Pyle atlas when applied to a modern Scottish population. *J Forensic Sci* 2013; 58: 114–119. doi:10.1111/j.1556-4029.2012.02294.x
- [22] Himes JH. An early hand-wrist atlas and its implications for secular change in bone age. *Ann Hum Biol* 1984; 11: 71–75. doi:10.1080/03014468400006911
- [23] Schmidt S, Koch B, Schulz R et al. Comparative analysis of the applicability of the skeletal age determination methods of Greulich-Pyle and Thiemann-Nitz for forensic age estimation in living subjects. *Int J Legal Med* 2007; 121: 293–296. doi:10.1007/s00414-007-0165-7
- [24] Gong P, Yin Z, Wang Y et al. Towards Robust Bone Age Assessment: Rethinking Label Noise and Ambiguity. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., Hrsg. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Cham: Springer International Publishing; 2020: 621–630. doi:10.1007/978-3-030-59725-2_60
- [25] Mansourvar M, Ismail MA, Raj RG et al. The applicability of Greulich and Pyle atlas to assess skeletal age for four ethnic groups. *J Forensic Leg Med* 2014; 22: 26–29. doi:10.1016/j.jflm.2013.11.011
- [26] Zhang A, Sayre JW, Vachon L et al. Racial Differences in Growth Patterns of Children Assessed on the Basis of Bone Age1. *Radiology* 2009; 250: 228–235. doi:10.1148/radiol.2493080468
- [27] Schmeling A, Schulz R, Danner B et al. The impact of economic progress and modernization in medicine on the ossification of hand and wrist. *Int J Legal Med* 2006; 120: 121–126. doi:10.1007/s00414-005-0007-4
- [28] Yi PH, Arun A, Hafezi-Nejad N et al. Can AI distinguish a bone radiograph from photos of flowers or cars? Evaluation of bone age deep learning model on inappropriate data inputs. *Skeletal radiology* 2021; 1–6. doi:10.1007/s00256-021-03880-y
- [29] Reyes M, Meier R, Pereira S et al. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiol Artif Intell* 2020; 2: e190043. doi:10.1148/ryai.2020190043