ARTIFICIAL INTELLIGENCE (AI) IN CARPAL BONE AGE ASSESSMENT VERSUS GREULICH AND PYLE METHOD EVALUATED BY YOUNG AND EXPERIENCED PAEDIATRIC RADIOLOGISTS. A TWO-CENTRE EXPERIENCE.

Davide Alaimo[1], Manfredi De Angelis[1], Ettore Palizzolo[1], Domenica Matranga[2], Giuseppe Lo Re[1], Giuseppe Paviglianiti[3], Sergio Salerno[1]

[1]Dipartimento Bi.N.D. – Università degli studi di Palermo, Italy

[1]Dipartimento PROMISE – Università degli studi di Palermo, Italy

[3]UOC Radiologia Pediatrica – A.R.N.A.S. Civico – Di Cristina, Palermo, Italy

ESPR
European Society of Paediatric Radiology

INTRODUCTION

Bone age assessment is useful to evaluate maturity and to diagnose many pediatric disorders, including endocrinological, orthodontic and orthopedical disorders [1, 2].

The “Radiographic Atlas of Skeletal Development of the Hand and Wrist” by Greulich and Pyle is the most widely method for age estimation.

The x-ray of the wrist of the patient is compared to the standard x-rays in the atlas, bone age is determined for both female and male patients [3].

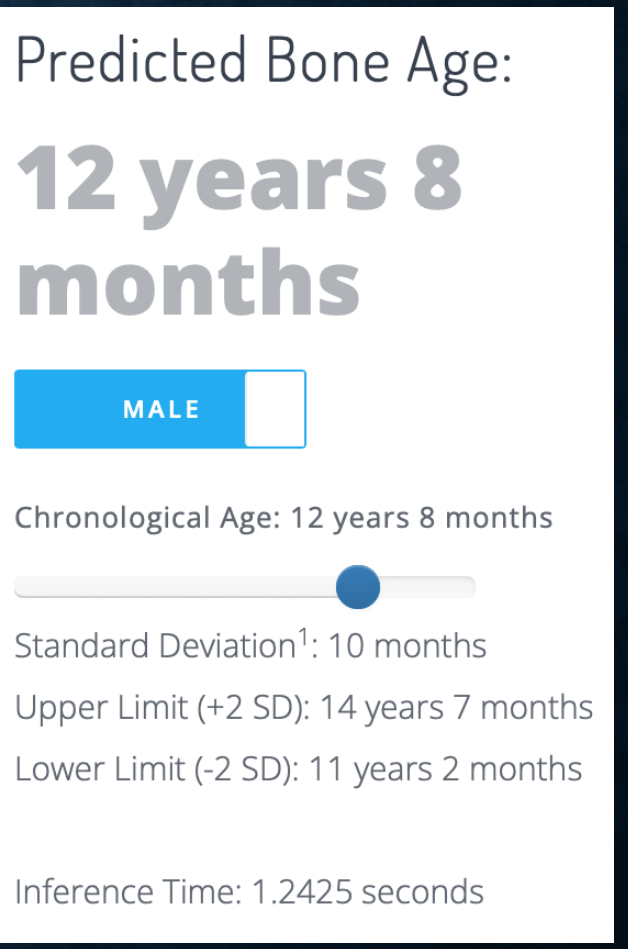
1. Satoh, M., Bone age: assessment methods and clinical applications. Clinical Pediatric Endocrinology, 2015. 24(4): p. 143-152.
2. Mansourvar, M., et al., Automated bone age assessment: motivation, taxonomies, and challenges. Comput Math Methods Med, 2013. 2013: p. 391626.
3. Greulich WW, P.S., Radiographic atlas of skeletal development of the hand and wrist. 1st edition ed. 1959, Stanford: Stanford University Press.

INTRODUCTION

In the last decades some machine learning system (AI) have been released to determine bone age on standard x-ray of the wrist.

Since HANDX system to recently, the Physis[tm] software (16bit, Toronto, Canada).

Bone age images are an ideal dataset for training a deep learning solution, as there is a single image of the left hand and wrist and relatively standardized findings [4].



1. Satoh, M., Bone age: assessment methods and clinical applications. Clinical Pediatric Endocrinology, 2015. 24(4): p. 143-152.
4. 16-bit. Physis[tm]. Available from: http://physis.16bit.ai/.
5. Lee BD, Lee MS. Automated Bone Age Assessment Using Artificial Intelligence: The Future of Bone Age Assessment. Korean J Radiol. 2021;22(5):792-800. doi:10.3348/kjr.2020.0941

AIM OF THE STUDY

In our study, we tried to assess the performance and concordance between the bone age evaluated with the Physis^{tm} software and with the Greulich-Pyle method in:

- ✓ Comparing distribution of evaluated ages;
- ✓ Inference time;
- ✓ Difference in GP method between senior and young radiologist.

MATERIALS AND METHODS

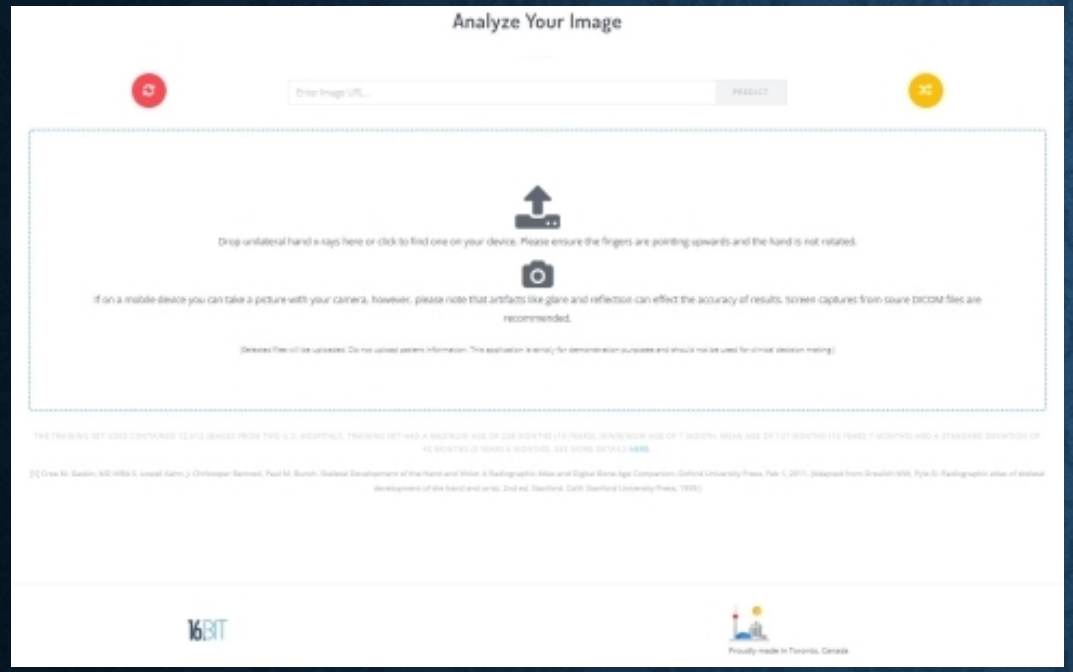
- ✓ 181 patients (95 males, 86 females) enrolled between October 2018 and October 2021.
- ✓ Age between 1 and 16 years old, mean age of 9.36.
- ✓ Two hospitals, "Policlinico Paolo Giaccone" and "A.R.N.A.S. Ospedali Civico Di Cristina Benfratelli - Ospedale Di Cristina" in Palermo.
- ✓ First, we used the Greulich-Pyle atlas.
- ✓ Second, we applied the Physis^{tm} software.

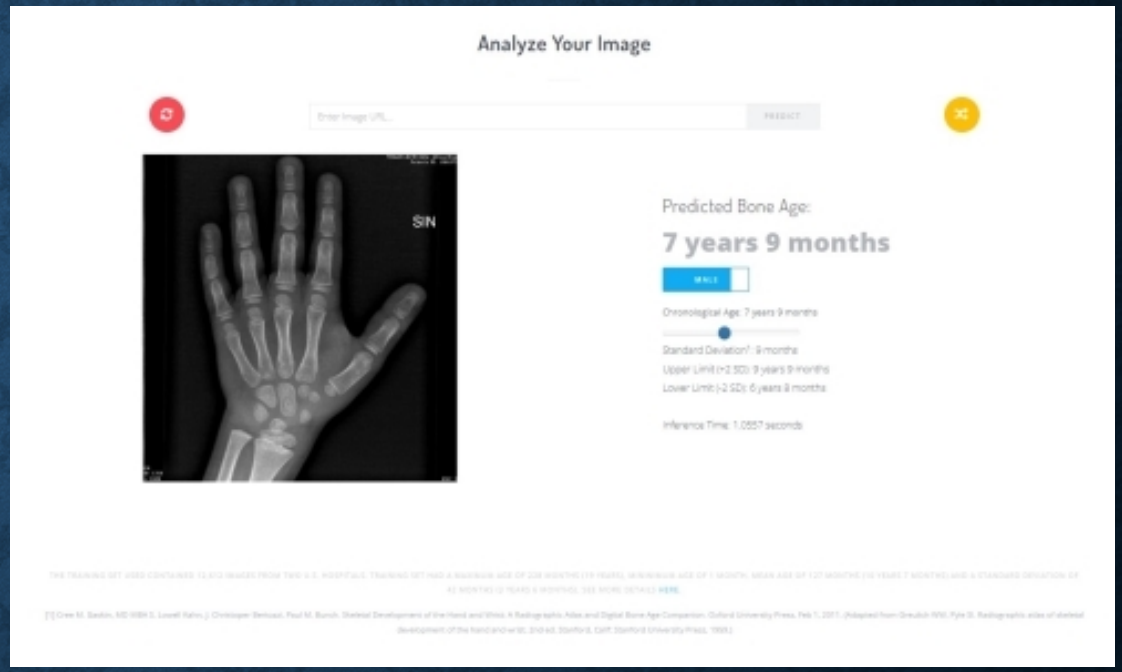
MATERIALS AND METHODS

- ✓ One expert and three residents analyzed each radiograph, knowing only the sex of the patient, and estimated the bone age on standard Greulich-Pyle atlas.
- ✓ We randomly measured the inference time.
- ✓ Second, we exported the DICOM images into jpeg format and uploaded the jpeg image on the server, which automatically resizes it to 500 x 500 pixels and, after a few seconds, supplied the estimated bone age, the standard deviation and the inference time [4, 6].

4. 16-bit. Physis[tm]. Available from: http://physis.16bit.ai/.
6. Halabi, S.S., et al., The RSNA Pediatric Bone Age Machine Learning Challenge. Radiology, 2019. 290(2): p. 498-503.

MATERIALS AND METHODS





- ✓ By the «drag&drop» system we can upload the picture on the server; after this the software analyses the radiograph and shows predicted bone age, inference time and standard deviation.
- ✓ It is possible to edit the real chronological age and the sex to assess correctly the standard deviation and the predicted bone age, respectively.

STATISTICAL ANALYSIS

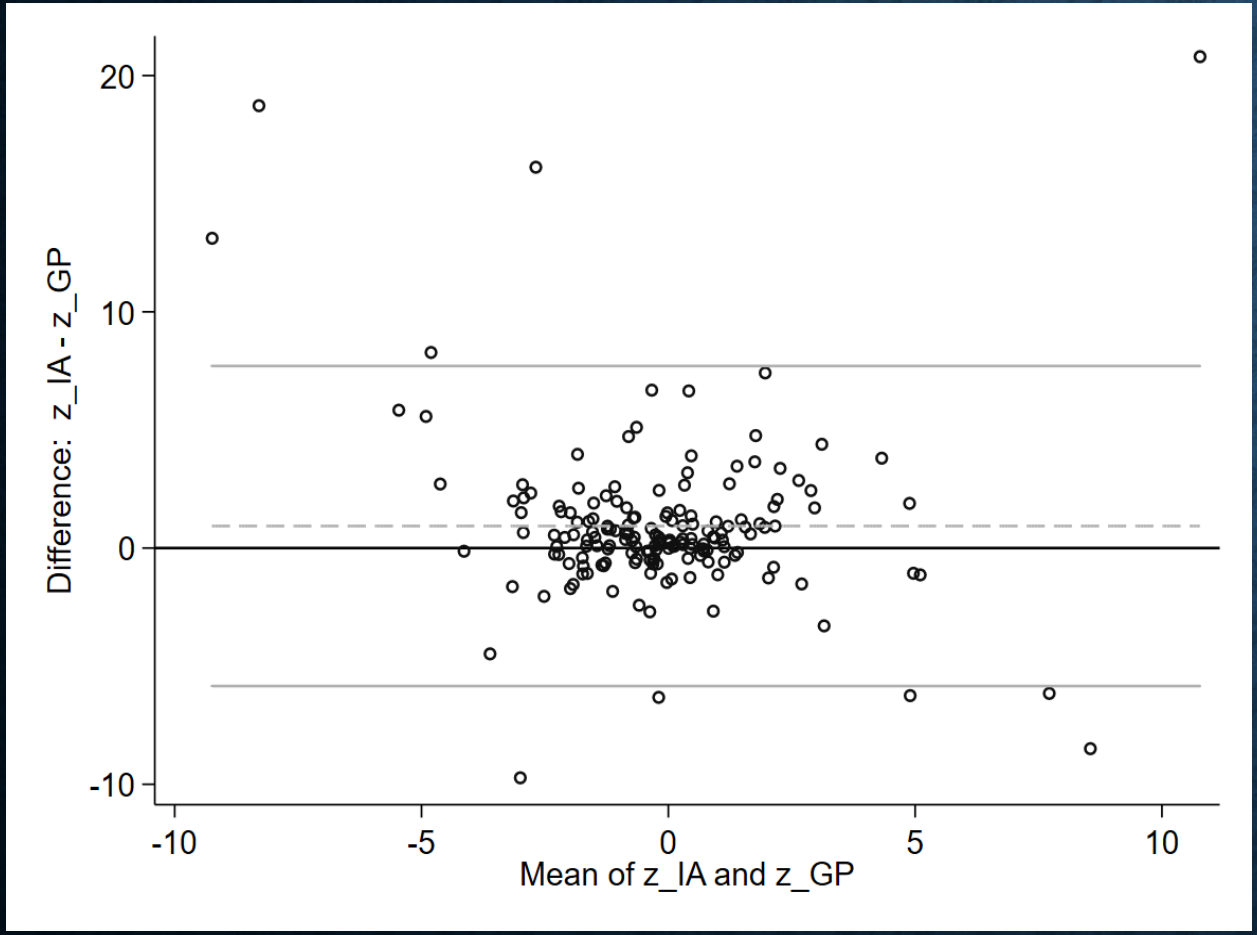
- ✓ AI is expressed in terms of $Mean_{AI}$ (SD_{AI}).
- ✓ To compare AI with GP, we have designed a repeated measures study with four radiologists (one expert and three residents) evaluating the same observation.
- ✓ After assessing normality through the Shapiro-Wilks test, we calculated Pearson's correlation between the expert and each of the three residents to assess reproducibility.
- ✓ Then, we used the mean and the SD of these four measurements to express GP in terms of $Mean_{GP}$ (SD_{GP}).

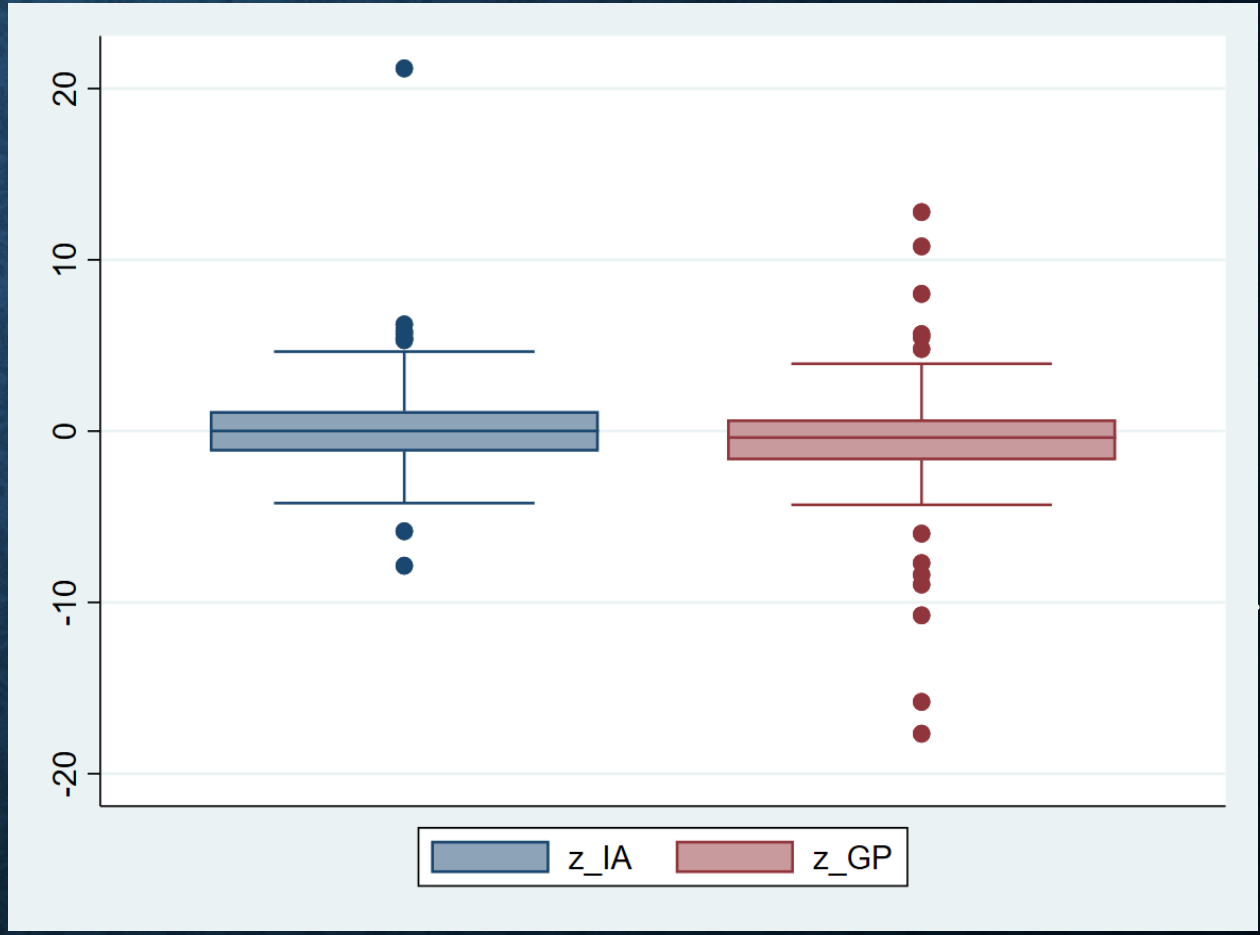
STATISTICAL ANALYSIS

- ✓ After assessing normality of AI, we calculated the z-scores for AI and GP.

- ✓ To assess the agreement between GP and AI, we obtained the Bland-Altman plot for z_{AI} and z_{GP}. Each couple of measurements is represented as a couple of coordinates on a cartesian system with their difference on the y-axis and their mean on the x-axis.

RESULTS


Difference: z_IA - z_GP
20
10
0
-10
-10
-5
0
5
10
Mean of z_IA and z_GP


20
10
0
-10
-20
z_IA
z_GP

DISCUSSION

- ✓ The bone age assessment suffers from an intrinsic limitation: the width of the standard deviations.
- ✓ It is not possible to overcome biological variation; hence, the 95% prediction interval for chronological age is wide for each bone age developmental stage [7].
- ✓ Our analysis shows that both GP and AI methods correctly estimates bone age; the mean difference were the higher variability of measurements made with GP method and the longer inference time of the manual evaluation.

7. Ording Müller, L.S., et al., Bone age for chronological age determination - statement of the European Society of Paediatric Radiology musculoskeletal task force group. Pediatr Radiol, 2019. 49(7): p. 979-982.

DISCUSSION

Our study has also limitations:

- First, we had no clinical history of our patients at the moment of evaluation; that means that the chronological age may differ from the bone age due to different clinical conditions.

- Second, the limited experience of the three residents, with less than 2 years of experience, may have influenced the rating and the GP analyses.

- Third, a recent study from Hi P.H. et al. highlighted that the software we applied does not recognize wrong inputs, like photos of flowers or chest radiographs [8].

7. Ording Müller, L.S., et al., Bone age for chronological age determination - statement of the European Society of Paediatric Radiology musculoskeletal task force group. Pediatr Radiol, 2019. 49(7): p. 979-982.
8. Yi, P.H., 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 Radiol, 2022. 51(2): p. 401-406.

CONCLUSION

- ✓ Both AI and GP correctly estimate the bone age; however, the measurements made by AI were faster and closer one to another rather than the GP method, so that we concluded that the AI made faster and more accurate evaluations.

- ✓ As a result, the current application in the clinical practice of this AI software may speed up radiographies evaluation.

ESPR
European Society of
Paediatric Radiology