

Automated bone age assessment across multi-site study in USA:

Agreement Between AI And Expert Readers

Wood, Jonathan¹, Chung, Tek Sin², Caudill, Karen A.¹, Hulett-Bowling, Rebecca², Kelsey, Matt², Cho, Woonchan², Hileman, Michael², DiFranco, Matthew D.¹, Mintz, Aaron²

¹ Washington University of St. Louis, USA

² ImageBiopsy Lab GmbH, Vienna, Austria



Disclosures & Ethics

- Chung, T.S. and DiFranco, M.D. are employees of ImageBiopsy Lab
- All other authors have nothing to disclose
- Ethics committee approval was obtained

Background

- Bone age from left-hand x-rays according to Greulich & Pyle (G&P) remains a common reference standard for skeletal maturity assessment.
- G&P bone age assessment is time consuming and prone to intra- and inter-reader variability.
- Artificial intelligence (AI) algorithms have shown promise in MSK x-ray assessment.
- IB Lab PANDA¹ is a CE-marked AI-based software that automates bone age estimation according to G&P.

[1] Not for sale in the U.S.

Patient and sampling

- 5541 hand radiographs taken for bone age assessment from 2011 to 2020 from multiple sites affiliated with the Washington University in St. Louis (WUSTL) School of Medicine were available for the study.
- 345 bone age x-rays were selected using stratified random sampling
 - boys 2-17 years
 - girls 2-16 years
 - stratified by years
- No patients with more than one image

Patients - Conditional Indications

Conditional Indication	Relative Distribution	Conditional Indication	Relative Distribution
short stature	96 (28.07%)	premature thelarche	10 (2.92%)
scoliosis	52 (15.2%)	growth hormone deficiency	6 (1.75%)
none	39 (11.4%)	fracture (unspecified)	5 (1.46%)
premature adrenarche	31 (9.06%)	failure to thrive	4 (1.17%)
precocious puberty	31 (9.06%)	NF1	3 (0.88%)
developmental delay	13 (3.8%)	accelerated growth	3 (0.88%)
poor growth	11 (3.19%)	others	31 (9.06%)
leg length discrepancy	10 (2.92%)		
Total	345 (100%)		



ESPR

European Society of
Paediatric Radiology



Bone Age Estimation

- Three pediatric radiologists (6, 19 and 27 years post-fellowship) read G&P bone age blinded to chronological age, each other and AI results.
- Ground truth was established using:
 - Mean of the three readers, or
 - Consensus of the three readers when any two initial read differed by more than 6 months.
- All images were subsequently processed with IB Lab PANDA v1.06

Comparison with automated bone age estimation

- Statistical analysis was performed to compare automated BA with the ground truth using
 - **Accuracy** in estimating the correct G&P reference image
 - Is AI output within 1 G&P plate of the ground truth?
 - **Mean absolute difference (MAD)**
 - **Root mean squared error (RMSE)**
 - **Interchangeability** of PANDA with pediatric radiologist readers

Results

	Comparison to Ground Truth			Comparison with Readers
	MAD [months]	RMSE [months]	Plate Accuracy [percent]	Interchangeability [months]
PANDA	5.79 (5.30; 6.28)	7.46 (6.86; 8.06)	89.70 (86.77; 92.77)	-5.8 (-7.1; -4.8)
Reader 1	5.67 (4.90; 6.44)	9.26 (8.01; 10.52)		
Reader 2	3.95 (3.36; 4.52)	6.81 (5.84; 7.75)		
Reader 3	4.94 (4.30; 5.55)	7.78 (6.87; 8.64)		

GT = Ground Truth, MAD = Mean absolute deviation, RMSE = Root mean squared error

Plate Accuracy measures whether the AI output is within 1 G&P plate of the ground truth G&P plate.

Inteachangeability shows the mean change in inter-reader differences when interchanging PANDA with a reader.

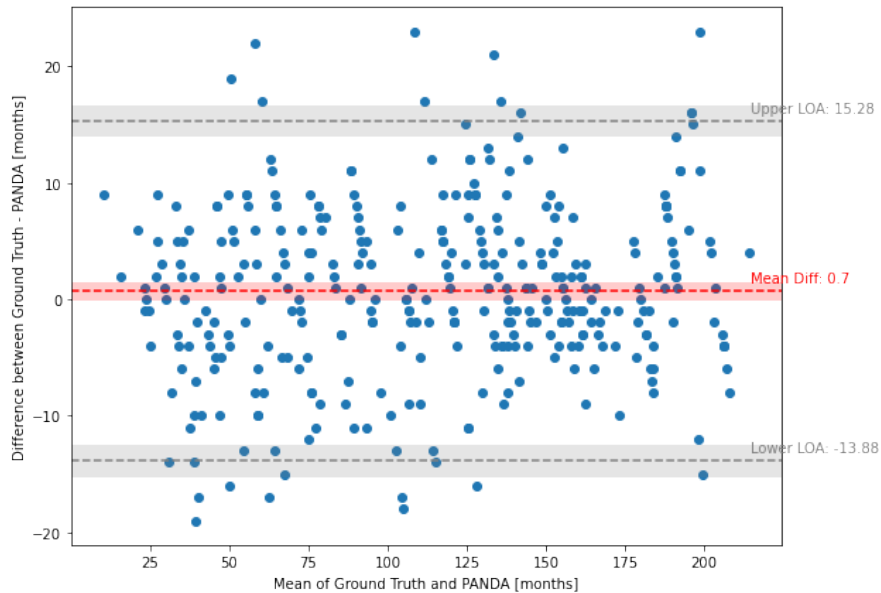


ESPR

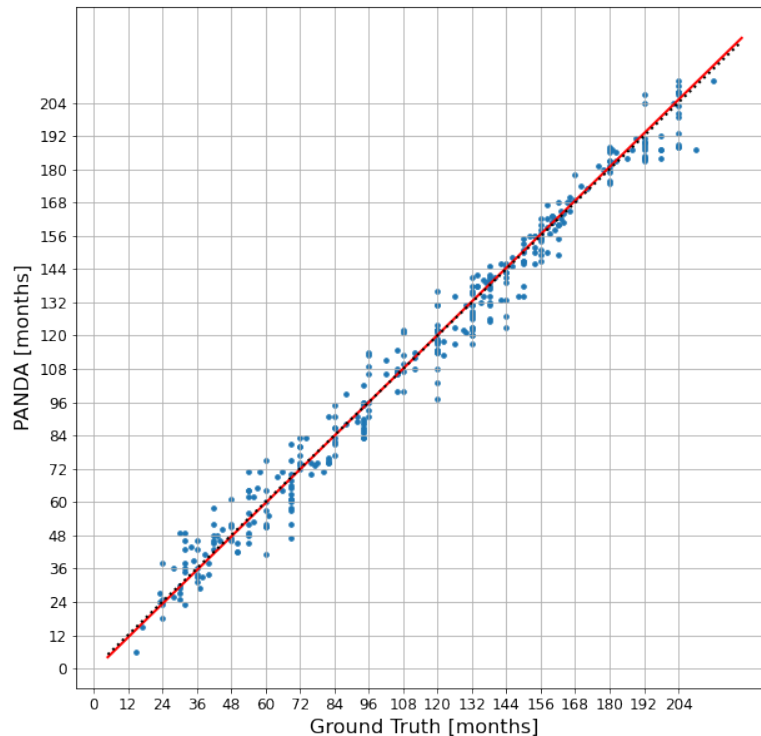
European Society of
Paediatric Radiology



Bland-Altman and Regression Plots



Shaded areas are 95% confidence intervals



ESPR

European Society of
Paediatric Radiology

Conclusions

- Good agreement between an AI and G&P ground truth on a US cohort from the clinical routine.
- Plate accuracy of nearly 90% suggests that the AI software can aid experts
- AI software demonstrated interchangeability with expert readers
- Clinical G&P bone age assessment can benefit from reliable AI automation which can be fully integrated into the reading workflow.