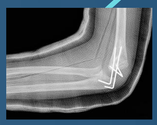
KK Women's and Children's Hospital
SingHealth

ELBOWING IN ON ARTIFICIAL INTELLIGENCE: FLEXING ChatGPT DIAGNOSTIC MUSCLE IN *PAEDIATRIC ELBOW FRACTURES*

Authors: Jonathan Kia Sheng Phua, Eu Jo Wong, Timothy Shao Ern Tan

Department of Diagnostic and Interventional Imaging, KK Women's and Children's Hospital, Singapore



Introduction

- Elbow fractures are the commonest injuries in children and are frequently evaluated with plain radiographs in the acute setting.
- Radiographic assessment of the paediatric elbow can however be challenging due to:
  - Complex changing anatomy of growing bones.
  - Varied appearance of ossification centres, which can confound interpretation.
- **As dedicated paediatric radiology services are not widely available**, timely reporting of these fractures could be delayed.
  - **Imperative for prompt and accurate diagnosis and management of these injuries** to prevent significant complications.
- Since 2023, ChatGPT-4 has offered image analysis capabilities, which may also be used for radiographic analysis.
- **This study aims to validate ChatGPT-4's diagnostic performance in detecting bony fractures on paediatric elbow radiographs.**

Materials and Methods

- A curated set of 200 paediatric elbow radiographs:
  - 100 normal, 100 abnormal with at least one fracture site.
  - 105 right elbow and 95 left elbow radiographs.
  - Performed between ***October 2023 and March 2024*** at the largest tertiary paediatric hospital in Singapore.
- These anonymized radiographs were uploaded onto the ChatGPT-4 platform, and subsequently prompted for fracture detection.
- ChatGPT-4's outputs were then compared against verified radiology reports (ground-truth).



ChatGPT

- "Is there a fracture?"
- "Where is the fracture?"

Results

Table 1: Demographics + Fracture Sites	N=200 (100 Normal, 100 Abnormal)
Males	109 (54%)
Females	91 (46%)
Median Age (years)	7 (Range: 8 months to 19 years)
Fractures Sites (100 radiographs)	
• Supracondylar	71 (71%)
• Humeral Epicondyle	5 (5%)
• Proximal Radius	12 (12%)
• Proximal Ulna	14 (14%)
• Monteggia Fracture-Dislocation	1 (1%)
Radiographs with at >1 fracture site	5 (5%)

Table 2: Performance		ChatGPT Output		
		Normal (No #)	Abnormal (#)	Total
Radiologist Reports (Ground truth)	Normal (No #)	82	18	100
	Abnormal (#)	13	87	100
	Total	95	105	200

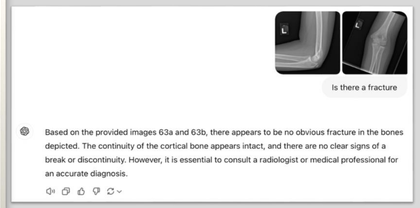
Concordance:
- ✓ 169 (85%) – 87/100 cases with fractures, 82/100 normal cases.
- ✓ ChatGPT identified fracture site accurately in 68 (78%) of 87 cases which it had detected fractures correctly.

Discordance: 31(15%) – 13/100 cases with fractures, 18/100 normal cases.

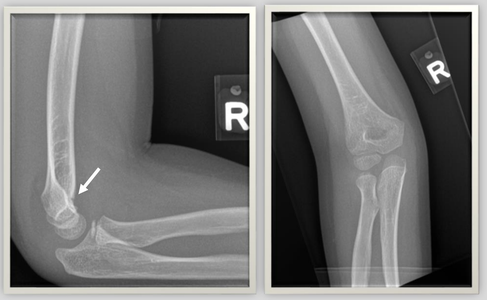
AUROC	F1-Score	Sensitivity	Specificity	PPV	NPV
0.85	0.85	0.87	0.82	0.83	0.87

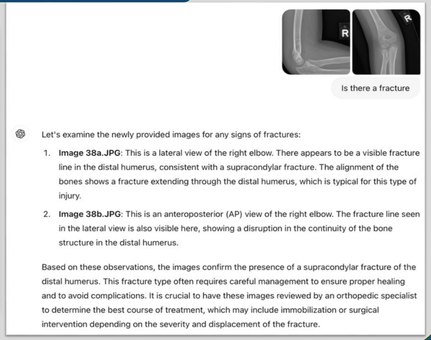
Results- Sample Output (Normal Radiograph)





Results- Sample Output (Radiograph with Fracture)





Conclusion

This study highlights the **utility and potential applications of ChatGPT-4** as a valuable **point-of-care tool** in **aiding detection of paediatric elbow fractures in** emergency settings:

- ✓ Where availability of paediatric radiology services may be limited.
- ✓ Larger multi-centre datasets are required to validate the performance of the algorithm further.

References

1. Guermazi A, Tannoury C, Kompel AJ et al (2022) Improving Radiographic fracture Recognition Performance and Efficiency using Artificial Intelligence. Radiology 302(3):627–636.
2. Barash Y, Klang E, Konen E, Sorin V (2023) ChatGPT-4 assistance in optimizing Emergency Department Radiology referrals and Imaging Selection. J Am Coll Radiol 20(10):998–1003.
3. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD (2023) Evaluating ChatGPT as an Adjunct for Radiologic decision-making. medRxiv.
4. Oka K, Shiode R, Yoshii Y, Tanaka H, Iwahashi T, Murase T (2021) Artificial intelligence to diagnosis distal radius fracture using biplane plain X-rays. J Orthop Surg Res 16(1):694.
5. Tobler P, Cyriac J, Kovacs BK et al (2021) AI-based detection and classification of distal radius fractures using low-effort data labeling: evaluation of applicability and effect of training set size. Eur Radiol 31(9):6816–6824.
6. Zech JR, Carotenuto G, Igbinoba Z, Tran CV, Insley E, Baccarella A, Wong TT (2023) Detecting pediatric wrist fractures using deep-learning-based object detection. Pediatr Radiol 53(6):1125–1134.
7. Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundström C, Hedlund J, Gordon M (2021) Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop 92(5):513–525.
8. Anttila TT, Karjalainen TV, Mäkelä TO, Waris EM, Lindfors NC, Leminen MM, Ryhänen JO (2023) Detecting Distal Radius fractures using a segmentation-based Deep Learning Model. J Digit Imaging 36(2):679–687.
9. Gan K, Xu D, Lin Y et al (2019) Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthop 90(4):394–400
10. Kim DH, MacKinnon T (2018) Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 73(5):439–445
11. Thian YL, Li Y, Jagmohan P, Sia D, Chan VEY, Tan RT (2019) Convolutional Neural Networks for Automated Fracture Detection and localization on wrist radiographs. Radiol Artif Intell 1(1):e180001.
12. Lindsey R, Daluiski A, Chopra S et al (2018) Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 115(45):11591–11596.