ORIGINAL ARTICLE

Artificial Intelligence (AI) in a Singaporean Emergency Department: Detecting fractures and reducing recalls

Chan Hiok Yang, FRCR[1], Tang Yee Peng, MD[1], Wong Zi Yang, MD[2], Koh Shao Hui, MBBS[2], Oliver Nickalls, FRCR, Wong Bak Siew Steven, MMed[1], Tan Min-On, FRCR[1]

[1]Department of Radiology, Sengkang General Hospital, 110 Sengkang E Way, Singapore, [2]Department of Emergency Medicine, Sengkang General Hospital, 110 Sengkang E Way, Singapore

ABSTRACT

Introduction: There has been rapid increase in the number of artificial intelligence and machine learning (ML) algorithms in recent years. In our local emergency department (ED), after-hours, radiographs are read by the ED doctor, with formal reporting by the radiology department performed on the subsequent day. Discrepant diagnoses between the ED doctor and radiologist potentially result in recalls of discharged patients for additional treatment, leading to greater monetary and manpower costs. To the authors' knowledge, no Singapore based study has utilized local data to analyse the performance of an AI fracture detection solution in the Singapore ED. The objective of this study is to evaluate the diagnostic performance of an AI radiograph fracture tool compared to ED doctors.

Materials and Methods: A retrospective study was conducted on 42 discrepant radiographic studies. In these studies, the final radiology report by the radiology department (the "ground truth") had a different diagnosis from bedside radiographic assessment by an ED Doctor.

Results: There were 20 studies with fractures and 22 studies with no fractures. The AI solution correctly diagnosed 15 fractures (75.0% of cases with fracture) (Figure 1), missed 5 fractures (25.0% of cases with fracture) and overcalled 1 fracture (4.5% of cases with no fracture) (Figure 2). The AI solution sensitivity is 75.0%, specificity is 95.5%, positive predictive value (PPV) is 93.8% and the negative predictive value (NPV) is 80.8%.

Conclusion: Having a fracture detection AI solution has the potential of reducing discrepant cases by up to 73.7% in the ED setting. Further large-scale studies should be performed to quantify the economic, manpower and healthcare outcome benefits of such an AI solution.

KEYWORDS:
Artificial Intelligence, Discrepancy, Fracture, Radiograph

INTRODUCTION

There has been rapid increase in the number of artificial intelligence (AI) and machine learning (ML) algorithms in recent years due to advances in computational power and increase in legislative approval.[1-14] This is especially evident in radiology where AI has shown great potential to triage, detect and classify abnormalities on imaging. Within the emergency department (ED), injuries to extremities account for up to 50% of non-fatal injuries, with almost all of these patients requiring radiographic imaging.[15] In our local ED, after-hours, radiographs are read by the ED senior physicians/consultants, with formal reporting by a radiology resident or consultant being performed only on the subsequent day. Discrepant diagnoses between the ED doctor and radiologist potentially result in recalls of discharged patients for additional treatment, leading to greater monetary and manpower costs on the healthcare system.[16] Apart from having 24 hours radiology radiographic reporting coverage which is labour intensive, some institutions have tried to decrease discrepancy rates by having monthly teaching/training meetings between the radiology and ED.[17] However, with the advancement in AI diagnostic algorithms, there exists now an additional method of increasing accuracy and efficiency in detecting and diagnosing fractures.[1-5,7-9,12-14] To the authors' knowledge, no Singapore based study has utilized local data to analyse the performance of an AI fracture detection solution in the Singapore ED. Moreover, while most AI articles focus on the diagnostic performance of AI in routine radiographs, this article focus on discrepant radiographs between ED physicians and radiologists, which are inherently more difficult.

Aim

The objective of this study is to evaluate the diagnostic performance of an AI radiograph fracture detection support tool compared to radiographic assessment by emergency department doctors based on discrepant radiographs between ED physicians and radiologists.

MATERIALS AND METHODS

A retrospective study was conducted on a set of discrepant radiographic studies collected over 3 consecutive months in 2019. These studies were from patients that were either 1) initially discharged from the ED without a diagnosis of fracture and later recalled for treatment following the diagnosis of fracture or 2) initially discharged from the ED with a diagnosis of fracture and later informed that no fracture was evident on radiograph. In these studies, the final radiology report by a radiology consultant or resident had a different diagnosis from bedside radiographic assessment by an ED Doctor. This retrospective study was approved by our institutional review board.

This article was accepted: 23 June 2025
Corresponding Author: Chan Hiok Yang
Email: hiokyang@gmail.com

A total of 53 discrepant radiographic studies were acquired. Patient details and demographics were not recorded for the purpose of this study. Of these 53 discrepant radiographs, 11 studies were excluded from the study as these could not be processed by the AI solution – 4 were chest or spine radiographs and 7 were of skeletally immature patients.

The remaining 42 studies were processed by a commercial AI deep learning-based solution – RB Fracture (Radiobotics, Copenhagen, Denmark) on an AI orchestration platform - CARPL (CARPL.AI, US). RB fracture is an AI technology tool for pelvic and limb radiographs based on advanced computer vision and machine learning methods. It is E-marked as Class IIa according to (EU) MDR, commercially available AI solution that is FDA approved, and has been deployed in clinical settings in Europe and US. When a fracture has been detected, the AI solution will highlight the site of fracture with a square ROI (Figures 1 and 2).

All the 42 studies were also newly and independently reviewed by a pair of radiology consultants (one with 2 years and another with 7 years of specialty experience). The radiology consultants were blinded to the clinical history during assessment. If there was a disagreement between the two radiology consultants, the radiograph was to be reviewed by a third radiology consultant.

There were no financial or non-financial support by the vendors for this study. This study has been approved by the Institutional Review Board (IRB) (IRB registration 2023/2159).

Ground Truth

The ground truth was defined as "fracture" or "no fracture" diagnosed on radiograph by both radiology consultants.

RESULTS

Both the radiology consultants concurred with the initial radiologist for all 42 cases. Of the 42 discrepant radiographic studies, there were 20 studies with fractures and 22 studies with no fractures diagnosed by the pair of radiology consultants – "ground truth". Radiographs with fractures only had one fracture each. All studies had 100% concordance between both radiology consultants.

The AI solution correctly diagnosed 15 fractures (75.0% of cases with fracture) (Figure 1), missed 5 fractures (25.0% of cases with fracture) and overcalled 1 fracture (4.5% of cases with no fracture) (Figure 2). The AI solution sensitivity is 75.0%, specificity is 95.5%, positive predictive value (PPV) is 93.8% and the negative predictive value (NPV) is 80.8%. For the 5 missed fractures, 4 fractures were located in the radial head of neck, 1 fracture was located in the talus. For the 1 fracture overcalled, the os naviculare was overcalled as a fracture.

DISCUSSION

Globally, the EDs are facing increasing workload with ED attendance doubling in the past 20 years.[7] Locally, total ED attendance at public hospitals has grown approximately 5.57% per year between 2005 and 2016.[18] This results in an increasing number of radiographic studies ordered in the ED and poses a great challenge to ED physicians, requiring them to read a greater volume of radiographs per ED shift.[7,17,19] However, compared to radiologists, ED physicians are at a disadvantage in reading such radiographic studies due to the utilization of non-dedicated diagnostic monitors, suboptimal room lighting, and significant time and patient turnaround pressures. Studies have reported a large variety of ED physician radiographic discrepancy rates ranging from 1.1%-23%.[19-22] Up to 24.2% of these patients were recalled back to the ED for further treatment.[20] The process of recalling is laborious and time intensive process which requires the patient to be seen a second time by the ED physician, incurring additional monetary and manpower costs. Recalling will also exacerbate overcrowding within the ED. Our study has shown that the AI solution can potentially reduce 75.0% recalls due to discrepant radiograph findings. In addition, the AI solution has a high specificity of 95.5% and PPV of 93.8%, implying that patients will unlikely undergo unnecessary fracture treatment. In a separate local study, the same AI solution has also proven to be effective in non-discrepant radiographs, with a sensitivity of 98.9%, an accuracy of 85.9%.[23]

Our study demonstrates a sensitivity of 75.0% likely because these were discrepant radiographs that were, by nature, more difficult radiographs that ED doctors and radiology trainees have misdiagnosed. On further analysis of the false negative cases (i.e fractures that were missed), majority of them were radial head/neck fractures which are notoriously difficult to diagnose.[19,24]

In the local Singapore context, AI fracture detection solutions also have the potential to reduce major discrepancy rates of radiology residents/trainees as a large proportion of hospital radiographs are formally reported by them. Notwithstanding reducing discrepancy rates, AI solutions have additional benefits of use in healthcare settings with limited resources. There is great potential to improve patient care by increasing diagnostic accuracy and reducing the burden on healthcare resources.[3,5] AI radiograph fracture detection support tools have all had significant published positive results, though they have different sensitivity and specificity depending on factors (definition of ground truth, the type of training and validation sets and the model function and algorithm). Some AI solutions have been shown to be superior in accuracy to radiology trainees (3), non-musculoskeletal trained radiologists (7), non-radiologists (2), and orthopaedic surgeons.[14] Kuo et al reported having a pooled sensitivity of 91% in fracture detection based on a meta-analysis of 37 studies which did not include RB fracture as an AI tool.[5] Guermazi et al. reported an increase in sensitivity of fracture detection by 10.4% and shortened average reading time by 6.3 seconds per radiograph.[12] Meynet et al. reported AI improving the sensitivity of radiologists for fracture detection by 20%, specificity by 0.6% and area under the ROC curve increased up to 10.6%.[1]

RB Fracture, the AI solution that was used in this study, has a reported relative reduction of missed fractures by the ED doctors of 43%.[25] The majority of published AI fracture detection tools focus on one anatomic area.[15] The major

Table I: Summary table of AI performance in all 42 radiographs

Site	Ground Truth - Fracture	AI Accuracy	Ground Truth.- No Fracture	AI Accuracy
Shoulder	3	3/3	1	1/1
Elbow	8	4/8	5	5/5
Pelvis	1	1/1	NA	NA
Femur	NA	NA	2	2/2
Knee	3	3/3	3	3/3
Ankle	4	3/4	7	7/7
Foot	NA	NA	4	3/4
Toes	1	1/1	NA	NA

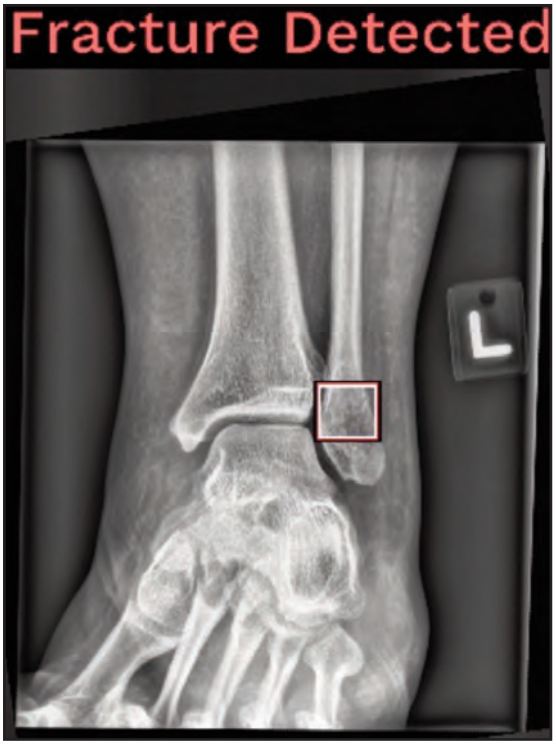

Fig. 1: Left ankle radiograph in Mortise view. Left lateral malleolus fracture that was correctly diagnosed. The fracture site ROI was labelled by the AI fracture detection tool

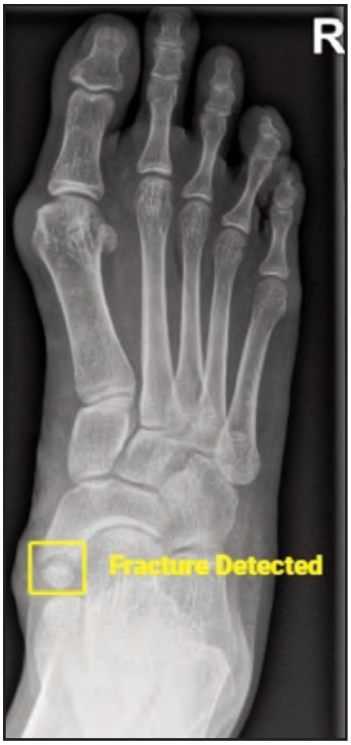

Fig. 2: Right foot radiograph in Dorsal-Plantar view. Right os naviculare that was overcalled as a fracture. The "fracture" site ROI was labelled by the AI fracture detection tool

strength of RB Fracture is its ability to analyse multiple anatomic areas for fractures, reducing the need to have a separate solution for each region or body part.

Despite the positive findings of our study, we acknowledge some limitations and biases. Our analysis is limited by a small sample size. The retrospective assessment of these cases from an already categorized "discrepant case" poses selection biases in the analysis. We were unable to obtain additional cases for processing due to dataset issues and manpower exigencies during the study. Taking such a solution to full live implementation will require further testing on a larger local dataset with internal and external validation.[3] The "ground truth" in our study may also be subject to contention due to the human readers being blinded to clinical history. Extending this, some AI tools may be trained on such "subjective" ground truths resulting in selection and training biases.[13] In addition, currently RB fracture can only analyse pelvic and limb radiographs though this study has shown great potential for future improvements to the algorithm. Regarding fractures, AI may be useful for majority of fractures but limited in pathological fractures which has great complications.[14] Needless to say, there remain outstanding ethical and legal issues and disclosures that require resolution and refinement before such solutions can be fully integrated into daily practice.[3]

Despite the present limitations, the future of AI Diagnostic solutions in the field of radiology is bright. Our study provides a good snapshot of AI's potential in enabling EDs to operate more safely and cost-effectively. Besides improving afterhours discrepancy rate, if the fracture detection tool is proven to be as accurate as radiology residents or non-subspecialised radiology consultant, ED physicians can potentially rely on AI during office hours and decide on disposition before the final report by the radiology resident or consultant. Increasing efficiency is key in the emergency department as it helps to minimize overcrowding. In addition to fracture

detection, there are other applications for AI in musculoskeletal trauma such as occult fracture prediction, diagnostic worklist prioritisation and image quality improvement.[3,15] These will be useful solutions that aid in the resource planning and prioritization/triage of appropriate care. With regards to AI fracture detection tools, few published studies at present have looked at human-algorithm performance through prospective clinical trials (5). Further studies should be undertaken to analyse large scale data prospectively to further ascertain the benefits and utility of having a fully capable AI solution for fracture detection.

CONCLUSION

Having a fracture detection AI solution has the potential of reducing discrepant or recall cases in the ED by up to 75.0% while maintaining a high specificity of 95.5%. This is especially useful in times of reduced manpower hours for both radiologists and ED physicians. Whilst the results may vary across centres and depend on the exact type of AI solution deployed, the potential benefit of having a fracture detection AI solution in the ED setting is apparent.

Further large-scale studies should be performed to quantify the economic, manpower and healthcare outcome benefits of such an AI solution.

REFERENCES

1. Canoni-Meynet L, Verdot P, Danner A, Calame P, Aubry S. Added value of an artificial intelligence solution for fracture detection in the radiologist's daily trauma emergencies workflow. Diagn Interv Imaging 2022; 103(12): 594-600.
2. Twinprai N, Boonrod A, Boonrod A, Chindaprasirt J, Sirithanaphol W, Chindaprasirt P, et al. Artificial intelligence (AI) vs. human in hip fracture detection. Heliyon 2022; 8(11): e11266.
3. Lex JR, Di Michele J, Koucheki R, Pincus D, Whyne C, Ravi B. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction: A Systematic Review and Meta-analysis. JAMA Netw Open 2023; 6(3): e233391.
4. Cellina M, Ce M, Irmici G, Ascenti V, Caloro E, Bianchi L, et al. Artificial Intelligence in Emergency Radiology: Where Are We Going? Diagnostics (Basel) 2022; 12(12): 3223.
5. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. Radiology 2022; 304(1): 50-62.
6. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 2018; 73(5): 439-45.
7. Cohen M, Puntonet J, Sanchez J, Kierszbaum E, Crema M, Soyer P, et al. Artificial intelligence vs. radiologist: accuracy of wrist fracture detection on radiographs. Eur Radiol 2023; 33(6): 3974-83.
8. Michelson JD. CORR Insights(R): What Are the Applications and Limitations of Artificial Intelligence for Fracture Detection and Classification in Orthopaedic Trauma Imaging? A Systematic Review. Clin Orthop Relat Res 2019; 477(11): 2492-4.
9. Kalmet PHS, Sanduleanu S, Primakov S, Wu G, Jochems A, Refaee T, et al. Deep learning in fracture detection: a narrative review. Acta Orthop 2020; 91(2): 215-20.
10. Kim T, Moon NH, Goh TS, Jung ID. Detection of incomplete atypical femoral fracture on anteroposterior radiographs via explainable artificial intelligence. Sci Rep 2023; 13(1): 10415.
11. Chou PH, Jou TH, Wu HH, Yao YC, Lin HH, Chang MC, et al. Ground truth generalizability affects performance of the artificial intelligence model in automated vertebral fracture detection on plain lateral radiographs of the spine. Spine J 2022; 22(4): 511-23.
12. Guermazi A, Tannoury C, Kompel AJ, Murakami AM, Ducarouge A, Gillibert A, et al. Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence. Radiology 2022; 302(3): 627-36.
13. Link TM, Pedoia V. Using AI to Improve Radiographic Fracture Detection. Radiology 2022; 302(3): 637-8.
14. Langerhuizen DWG, Janssen SJ, Mallee WH, van den Bekerom MPJ, Ring D, Kerkhoffs G, et al. What Are the Applications and Limitations of Artificial Intelligence for Fracture Detection and Classification in Orthopaedic Trauma Imaging? A Systematic Review. Clin Orthop Relat Res 2019; 477(11): 2482-91.
15. Laur O, Wang B. Musculoskeletal trauma and artificial intelligence: current trends and projections. Skeletal Radiol 2022; 51(2): 257-69.
16. Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. Radiographics. 2015; 35(6): 1668-76.
17. Espinosa JA, Nolan TW. Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. BMJ 2000; 320(7237): 737-40.
18. Ansah JP, Ahmad S, Lee LH, Shen Y, Ong MEH, Matchar DB, et al. Modeling Emergency Department crowding: Restoring the balance between demand for and supply of emergency medicine. PLoS One 2021; 16(1): e0244097.
19. Petinaux B, Bhat R, Boniface K, Aristizabal J. Accuracy of radiographic readings in the emergency department. Am J Emerg Med 2011; 29(1): 18-25.
20. Tranovich MJ, Gooch CM, Dougherty JM. Radiograph Interpretation Discrepancies in a Community Hospital Emergency Department. West J Emerg Med 2019; 20(4): 626-32.
21. Mattijssen-Horstink L, Langeraar JJ, Mauritz GJ, van der Stappen W, Baggelaar M, Tan E. Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency department: a retrospective analysis. Scand J Trauma Resusc Emerg Med 2020; 28(1): 38.
22. Liu YM, O'Hagan S, Holdt FC, Lahri S, Pitcher RD. After-hour trauma-radiograph interpretation in the emergency centre of a District Hospital. Afr J Emerg Med 2022; 12(3): 199-207.
23. Quek JJX, Nickalls OJ, Wong BSS, Tan MO. Deploying artificial intelligence in the detection of adult appendicular and pelvic fractures in the Singapore emergency department after hours: efficacy, cost savings and non-monetary benefits. Singapore Med J 2025; 66(4): 202-7.
24. Kung JW, Melenevsky Y, Hochman MG, Didolkar MM, Yablon CM, Eisenberg RL, et al. On-call musculoskeletal radiographs: discrepancy rates between radiology residents and musculoskeletal radiologists. AJR Am J Roentgenol 2013; 200(4): 856-9.
25. Resources - Clinical Trials, Publications and Blog [Available from: https://www.radiobotics.com/resources.