Artificial intelligence in traumatology

a comparative study between conventional and AI-aided diagnostic performance in distal radius fractures

From Ludwig Boltzmann Institute, Vienna, Austria

Cite this article:
Bone Joint Res 2024;13(10):
588–595.

DOI: 10.1302/2046-3758.
1310.BJR-2023-0275.R3

Correspondence should be sent to Rosmarie Breu
rosmarie.breu@oss.at

R. Breu,[1,2,3] C. Avelar,[4] Z. Bertalan,[4] J. Grillari,[3,5,6] H. Redl,[3,6] R. Ljuhar,[4,5] S. Quadlbauer,[2] T. Hausner[2,3,6,7]

[1]Orthopedic Hospital Vienna-Speising, Vienna, Austria
[2]AUVA Trauma Hospital Lorenz Böhler, Vienna, Austria
[3]Ludwig Boltzmann Institute for Traumatology, the Research Center in Cooperation with AUVA, Vienna, Austria
[4]ImageBiopsy Lab, Vienna, Austria
[5]Institute of Molecular Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
[6]Austrian Cluster for Tissue Regeneration, Vienna, Austria
[7]Department for Orthopedic Surgery and Traumatology, Paracelsus Medical University, Salzburg, Austria

Aims
The aim of this study was to create artificial intelligence (AI) software with the purpose of providing a second opinion to physicians to support distal radius fracture (DRF) detection, and to compare the accuracy of fracture detection of physicians with and without software support.

Methods
The dataset consisted of 26,121 anonymized anterior-posterior (AP) and lateral standard view radiographs of the wrist, with and without DRF. The convolutional neural network (CNN) model was trained to detect the presence of a DRF by comparing the radiographs containing a fracture to the inconspicuous ones. A total of 11 physicians (six surgeons in training and five hand surgeons) assessed 200 pairs of randomly selected digital radiographs of the wrist (AP and lateral) for the presence of a DRF. The same images were first evaluated without, and then with, the support of the CNN model, and the diagnostic accuracy of the two methods was compared.

Results
At the time of the study, the CNN model showed an area under the receiver operating curve of 0.97. AI assistance improved the physician's sensitivity (correct fracture detection) from 80% to 87%, and the specificity (correct fracture exclusion) from 91% to 95%. The overall error rate (combined false positive and false negative) was reduced from 14% without AI to 9% with AI.

Conclusion
The use of a CNN model as a second opinion can improve the diagnostic accuracy of DRF detection in the study setting.

Article focus
- Train a convolutional neural network (CNN) model to detect the presence of a distal radius fracture (DRF) to provide a second opinion to physicians.
- Compare the diagnostic accuracy of DRF detection of physicians with and without software support.

Bone & Joint Research

Key messages

- Artificial intelligence (AI) software can be trained to detect DRF with a high accuracy.
- The use of this CNN model as a second opinion improved the diagnostic accuracy of physicians' DRF detection in the study setting.

Strengths and limitations

- A faster and safer diagnostic accuracy might be expected in Emergency Departments using this software.
- Each radiograph was primarily reviewed by two orthopaedic trauma surgeons to ensure high quality diagnosis.
- This study had a small sample size of 11 readers and there was no clinical examination of the patients, as retrospective radiographs were used.

Introduction

Distal radius fractures (DRFs) are defined as juxta-articular fractures up to 3 cm proximal to the radiocarpal joint.[1] They are among the most common fracture types in humans. The risk for a woman aged over 50 years suffering a DRF is 15%.[1,2] Chung and Spilson,[3] and MacIntyre and Dewan,[4] found that DRFs are accountable for about 1% of all presentations and 15% of all fractures in Emergency Departments (EDs). The standard treatment is either a cast immobilization and/or open reduction and volar locking plate fixation.[5,6]

The number of DRFs is expected to rise in the future, as population, life expectancy, and activity levels in elderly people will increase.[7,8] A higher number of patients in EDs has already been observed: Polinder et al[9] found an age-adjusted increase of upper limb injuries of 13% between 1986 and 2008 in the Netherlands. Existing data predict a similar outcome for patient numbers in Austria, with an estimated increase of over 65-year-olds from 18% in 2018 to over 27% in 2050.[10]

These figures underline the high prevalence of DRFs and increasing workload in EDs. Furthermore, the large direct and indirect costs associated with these fractures should be considered. The cost of all osteoporotic fractures in 2010 in the European Union amounted to €37.4 billion, with forearm fractures accounting for 2%.[11,12]

Currently, diagnoses of fractures are primarily based on clinical examination and visual assessment of conventional radiographs. Although most DRFs are not difficult to identify and occult fractures are rare, some non-displaced fractures, especially of the radial styloid, can be challenging to see on plane radiographs.[13,14] The causes of error are multifactorial, and can arise from subtlety of a particular fracture in a radiograph, high workload, and inexperienced or fatigued clinicians, especially during night shifts.[15,16] Unidentified fractures are the most common diagnostic errors in EDs, comprising up to 80%.[17–20] Among these, the distal radius is one of the most frequent locations.[20,21] Misdiagnosis or overlooked fractures may result in delayed, deprived, or prolonged therapy, pain or malfunction with decreased quality of life or inability to work, unnecessary medical and economical costs, as well as avoidable exposure to radiation, when performing repeated x-ray imaging or CT to confirm uncertain image findings.[14,22]

Computer-assisted detection systems are a potential solution for these problems, identifying regions on radiographs that are highly likely to contain a pathology and providing the clinician with a quick and reliable second opinion. Recent advances in deep learning models, especially in convolutional neural networks (CNNs) that specialize in processing grid-like data such as images, have allowed for the creation of computer models evaluating radiographs. In previous years, the potential of such models has been analyzed by assigning them various tasks, such as the assessment of osteoarthritis or chest pathologies and detection of cancer or dental caries, showing great promise.[23–27] The algorithms are trained on large datasets of radiographs and learned by example.[28]

With a certain number of labelled examples, a well-designed model can be trained to assess fractures. In previous studies, deep learning models achieved sensitivity rates for detecting fractures on radiographs between 90% and 95%.[29–33] In comparison, the sensitivity of physicians is described as only 71% to 82%. However, some highly specialized traumatologists show a sensitivity of up to 93%.[34] Several studies have even compared the sensitivity of fracture detection of physicians with and without AI assistance, and found a statistically significant improvement.[29,35,36]

The aim of this study was to evaluate the ability of AI software to assess plain radiographs for the presence of DRF, and to compare the diagnostic performance of physicians detecting such fractures with and without software support.

Methods

General

This retrospective study was approved by the local ethical review board of the Austrian Workers' Compensation Board (AUVA), (26/2019), and the Clinical Artificial Intelligence Research (CAIR)[37] checklist has been used for this paper.

Dataset

The model dataset consisted of 26,121 anonymized anterior-posterior (AP) and lateral digital radiographs of the wrist, which were randomly sampled from five AUVA trauma hospitals around Austria between 2015 and 2019. Overall, 49.5% (12,934) showed a DRF and 50.5% (13,187) were inconspicuous images showing no fracture. The radiographs were taken from patients aged over 18 years, sustaining an injury to the wrist. Patients who had any other diagnosis than DRF at the time of examination were excluded. Ground truth was defined as the diagnosis that was set during the initial patient contact with the patient being present, which allowed cross-checking of the set diagnosis after performing a clinical examination and follow-up checks. Each radiograph was primarily reviewed by two orthopaedic trauma surgeons: first by the doctor in the ED, which could have been either an orthopaedic trauma resident or specialist; and subsequently by an orthopaedic trauma specialist. Additionally, ground truth was checked again by two surgeons during the labelling process for the study.

The model's dataset was further randomly split into a training set (85%), tune set (7%), and test set (8%). The training set and tune set were used during software training, and the test set was exclusively used to analyze the performance of the model. This data split was skewed in favour of model training, in order to avoid losing high-quality data for hyper-parameter

Fig. 1
Model prediction. The classification score above the image determines the diagnosis (fracture = 1, no fracture = 0). The segmentation mask determines the location of the fracture (drawn polygon).

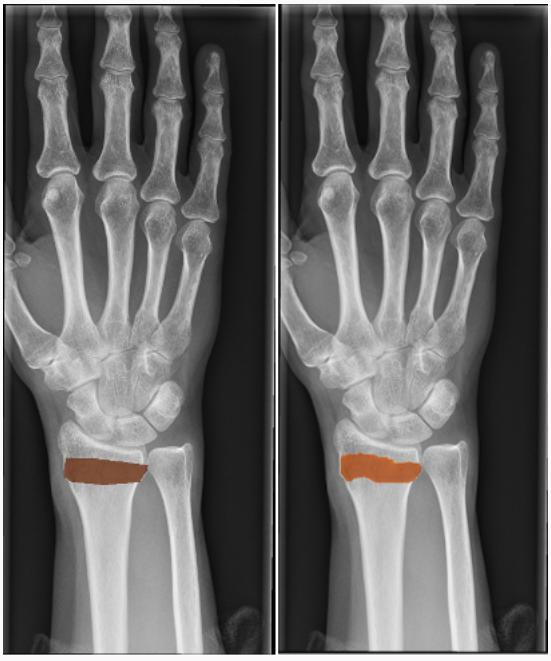

Fig. 2
Manually labelled fracture site (left) and model prediction (right).

tuning. This approach was deemed appropriate, since the final model performance was tested in the follow-up reader study. For blinding purposes, and to prevent bias during the test and validation processes, the test set was kept separate from the training set and tune set. Additionally, images from three different hospitals were used for the training set, and images from four different hospitals were used for the test set (two hospitals overlapping). This means that in the reader study, half of the images were from a completely different data source compared with those used for the training process.

Model training process

The AI model was trained to detect the presence of a DRF by comparing the radiographs containing a fracture to those without a fracture. Prior to training, each fracture was labelled by two surgeons (RB, see Acknowledgements), one orthopaedic trauma surgeon and one hand surgeon, drawing bounding boxes to help the model concentrate on the point of interest. The images were pre-processed by randomly resizing, rotating, or flipping them horizontally and vertically. An object detector network – a RetinaNet with a ResNet50 backbone (Facebook AI Research (FAIR), USA) – was used to identify the wrist area in all radiographs with AP and lateral views. A CNN was used for the task of classifying the presence or absence of a fracture on the image. The chosen architecture was a modified U-Net, with a classification branch appended to the end of the feature-encoding part. Thus, the classification model had two outputs, the region of the potential fracture and a score that is related to the probability of the presence of a fracture. This was chosen since the model was developed for a double task: the classification score that determines the diagnosis (fracture = 1, no fracture = 0); and the segmentation mask, which determines the location of the fracture (Figures 1 and 2).

A description of the detailed architecture can be found in the Supplementary Material.

Readers and reader study

Overall, 11 examiners (six orthopaedic trauma residents and five hand surgeons, each of the latter having at least ten years

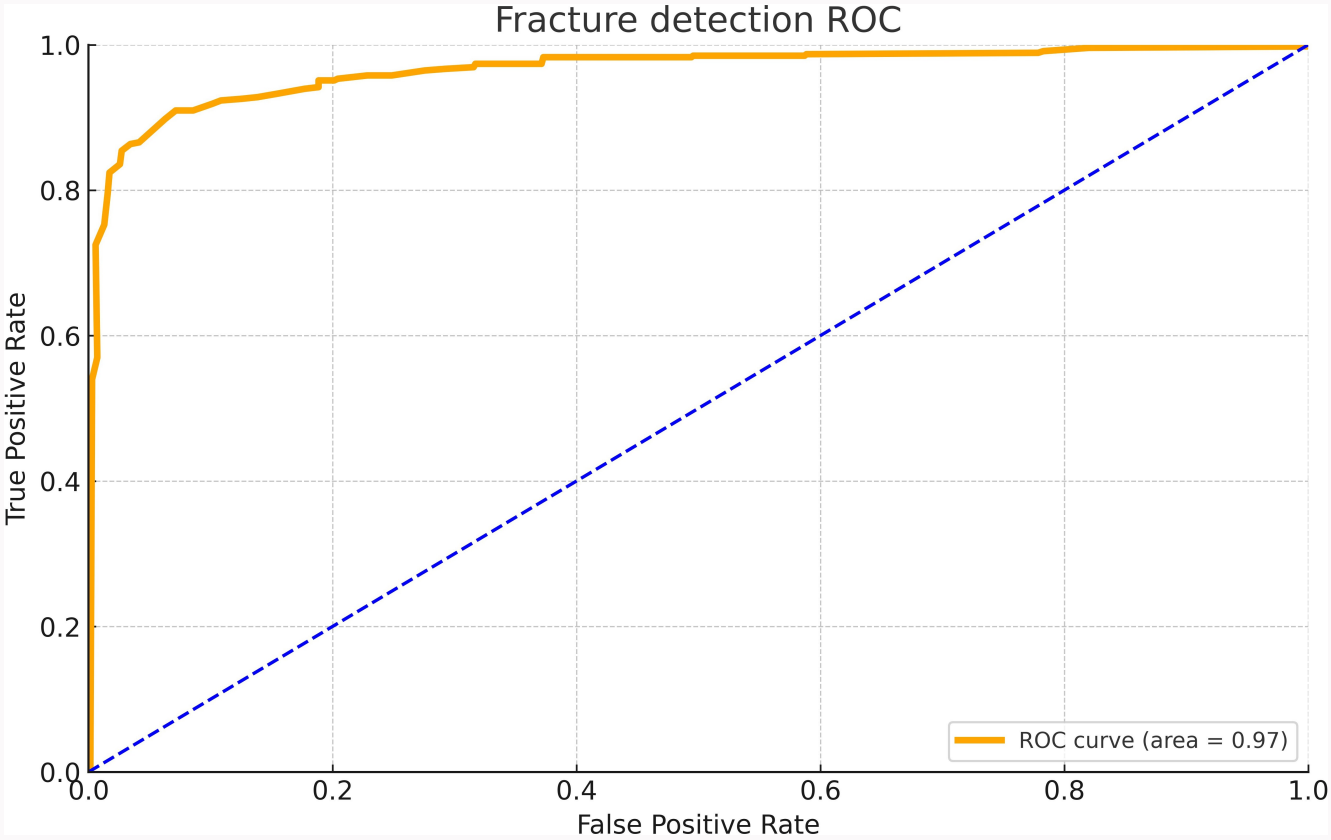

Fig. 3
Deep model training performance: area under the curve = 0.97. ROC, receiver operating characteristic.

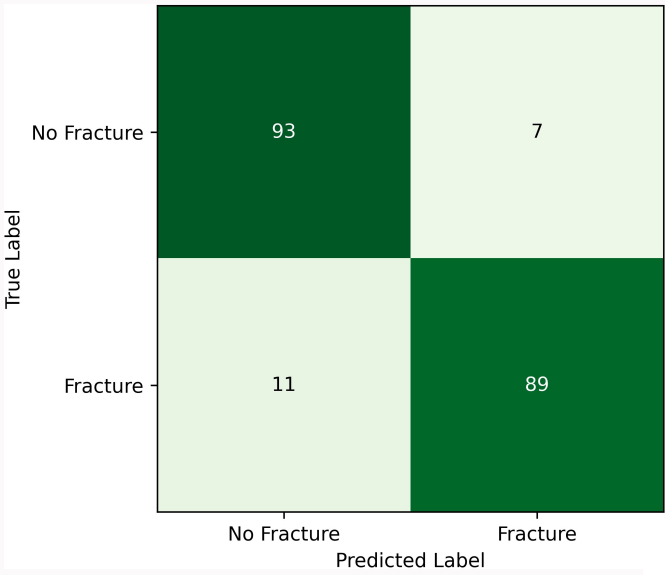

Fig. 4
Deep model training performance.

Table I. Demographic characteristics of patients used for the reader study.

Patient group	Number of women, n (%)	Mean age, yrs (SD)
With fracture	66/100 (66)	58.2 (16.8)
Without fracture	50/100 (50)	41.0 (20.3)
Total	116/200 (58)	49.0 (20.5)

of experience) were asked to evaluate the radiographs. All examiners participated independently from one another.

A total of 200 pairs of radiographs (200 AP and 200 lateral of the same wrist, respectively) were randomly sampled by using Python library "random" (Python Software Foundation, USA). To avoid any kind of bias, all images showing a cast, internal/external fixation, any artefacts, or containing inappropriate body parts (such as the elbow) were discarded. Of the cleaned images, 100 pairs of radiographs were randomly selected showing a DRF, and 100 pairs of inconspicuous radiographs were randomly selected showing no fracture. The examiners were not aware of any of these proportions. They were asked to evaluate all 200 cases regarding the presence of a DRF, first without the aid of AI software and then after a three-week washout period with the software's diagnostic opinion as assistance. The software showed the proposed diagnosis (fracture/no fracture) for each radiograph. The images were presented in a random order, and the readers had to participate independently from one another. Skipped images were taken as wrong diagnosis.

A demo version of the software used for the reader study can be found here, after creating an account: http://demo.imagebiopsy.com/.

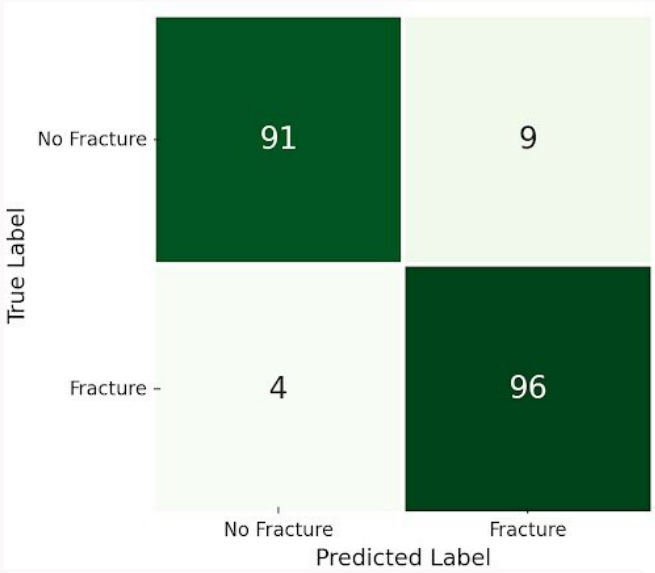

Fig. 5
Deep model standalone performance in the reader study.

Outcome measures and statistical analyses

For descriptive purposes, demographic data including sex and age were analyzed. The outcome measure was binary (fracture or no fracture) and compared to the ground truth. The standalone diagnostic performance of the AI model was measured by using receiver operating characteristic (ROC) and area under the curve (AUC), the latter being a standard method to describe a ROC curve, whereby an AUC of 1.0 means a perfect prediction of the reference standard, while an AUC of 0.5 means a random outcome. Furthermore, sensitivity (correct fracture detection), specificity (correct fracture exclusion), Youden Index, which monitors the overall diagnostic performance and is defined as sensitivity+specificity-1, and overall error rate were described.

The diagnostic performance of fracture detection of the surgeons was assessed by the following parameters: AI-aided and unaided sensitivity, specificity, Youden Index, and overall error rate. Unaided and aided diagnostic performance was compared using paired *t*-test after confirming normal distribution with Shapiro-Wilk test, otherwise using Wilcoxon signed-rank test. Furthermore, the diagnostic improvement of the trainees was compared to the experts by using the independent-samples *t*-test, to analyze if the trainees benefited more from the software's assistance more than the experts. Statistical analyses were performed using SPSS version 27 (IBM, USA). All tests were two-tailed with significance set at the 5% level.

Results

CNN model final training performance

The training dataset consisted of 9,017 (46%) images showing a DRF and 10,766 (54%) inconspicuous images without fracture. At the time of the study, the CNN model reached an AUC of 0.97, a sensitivity of 89%, and a specificity of 93% on the test set (Figures 3 and 4). This algorithm was used for the following reader study.

Table II. Reader study results. Unaided and aided diagnostic performance was compared using paired *t*-test, a = 0.05.

Variable	AI alone	Reader without AI	Reader with AI	p-value*
Sensitivity, %	96	80	87	0.065
Specificity, %	91	91	95	0.036
Youden Index	0.87	0.72	0.82	0.007
Experts		0.77	0.83	0.223
Trainees		0.67	0.82	0.020
Error rate, %	7	14	9	0.007

*Reader's results without AI versus with AI (paired *t*-test, $\alpha = 0.05$).
AI, artificial intelligence.

Reader study results

In the reader study, the mean patient age was 49.6 years (SD 20.5; 18 to 103), and 116 of the 200 included patients (58%) were female. Patients with a fracture had a mean higher age than those without a fracture (58.2 years (SD 16.8) vs 41.0 years (SD 20.3)), and were more often female (66% vs 50%) (Table I).

The standalone performance of the CNN model in the reader study showed a sensitivity of 96%, a specificity of 91%, a Youden Index of 0.87, and an error rate of 7% (Figure 5). The reliability of the model was calculated with the help of a reliability diagram. All model outputs are within the error bars (Figure 6).

Unaided and aided diagnostic performance was compared using paired *t*-test. AI assistance improved the physician's specificity statistically significantly from 91% to 95% (3.8% increase; 95% CI -1.1 to 8.7; p = 0.036), and Youden Index from 0.72 to 0.82 (increase 0.1; 95% CI 0.0 to 0.2; p = 0.007). Sensitivity also improved from 80% to 87% with AI assistance, but without statistical significance (6.7% increase; 95% CI 0.5 to 14.0; p = 0.065). The overall error rate (combined false positive and false negative) was reduced from 14% to 9% with statistical significance (5.3% decrease; 95% CI 1.8 to 8.8; p = 0.007) (Table II).

Trainees benefited more from AI assistance than experts, as the Youden Index in the trainee subgroup improved statistically significantly from 0.67 to 0.82 (increase 0.15; 95% CI 0.03 to 0.25; p = 0.020, paired *t*-test), but Youden Index in the expert subgroup showed improvement from 0.77 to 0.83 without statistical significance (increase 0.06; 95% CI -0.06 to 0.18; p = 0.223, paired *t*-test). The difference in improvement between the trainees and the experts was not statistically significant (mean difference 0.08; 95% CI -0.22 to 0.05; p = 0.200, independent-samples *t*-test) (Figure 6). Detailed results of all 11 readers can be found in the Supplementary Material.

Discussion

An AI algorithm was trained and tested to detect DRFs on radiographs of the wrist with an equal or better accuracy than that of a trauma surgeon. The standalone performance of the CNN model with an AUC of 0.97 is comparable to the literature (AUC 0.94 to 0.97).[29–32,35] In the reader study, the sensitivity could be improved from 80% to 87% with AI

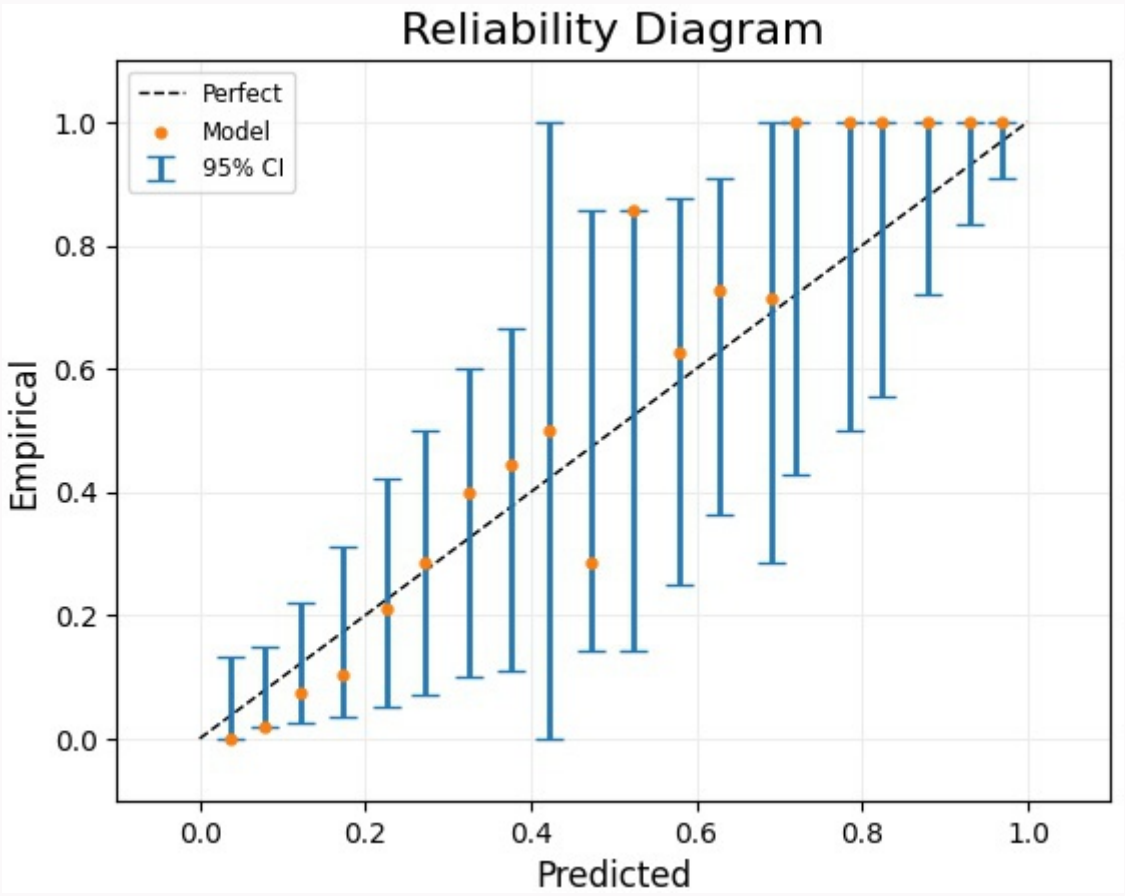

Fig. 6
Reliability diagram of the model's output on the study test set.

assistance, which is also comparable to the literature, where improvement between 7% and 11% is described.[29,35,36] In this study, the improvement of sensitivity was not statistically significant, possibly because two of 11 readers (one resident and one specialist) decreased their sensitivity with AI help from 97% to 84% and from 91% to 83%, respectively, due to unknown reasons. The AI standalone sensitivity exceeded the unaided and the aided performance of all 11 physicians. Regarding specificity, the AI standalone performance was similar to the unaided performance of the physicians. The unaided specificity of the physicians (91%) was the same as the AI standalone performance (91%). With AI help, the physician's specificity could be improved to 95%; the second opinion might increase the confidence of the physicians and therefore improve their performance. Youden Index increased with statistical significance in the trainee group, but not in the expert group. When comparing the residents to the specialists, there was no statistically significant difference, which may have resulted from the small sample size (5 vs 6 physicians). The three readers that improved the most were all in the residents group (Youden Index 0.58 vs 0.86, 0.64 vs 0.83, and 0.71 vs 0.9 without and with AI assistance, respectively).

Demographic data in the reader study reflected real-life demographics. There were more women (58%, 116/200), who tend to suffer from osteoporosis more often than men, and patients with fracture diagnosis tended to be older than patients with inconspicuous images.[1,2] Additionally, the 26,121 images were randomly chosen from different hospitals around Austria, which ensured a sufficient sample size and also allowed the model to learn from several image sources to reflect real-world clinical conditions as closely as possible.

Limitations of this study were the small sample size of 11 readers, and the missing clinical examination of the patients. Moreover, retrospective radiographs were used, and there are no data for the diagnostic accuracy of the readers in a real-life setting, which would include the help of patients' history and clinical examinations. Moreover, in this study we used a version of the CNN model, where only the proposed diagnosis was shown to the readers in the reader study. A later version of the algorithm also showed the fracture location on the radiographs, as described in the training process, but had not yet been used for this reader study. With the updated version, we expect even better results.

In conclusion, AI may improve the diagnostic accuracy of trauma residents and specialists. Not only might the number of missed fractures be reduced, but the tendency of trainees to over-report apparent abnormalities in radiographs might be prevented – Williams et al[38] observed a false-positive rate of 18% in training radiologists compared to more experienced radiologists. Furthermore, a faster and safer diagnostic accuracy in EDs can be expected. This might be helpful in choosing the correct treatment of DRF, which is often a complex decision.[39–41] AI algorithms could also be used as triage systems, prioritizing patients with radiographs that show a potential fracture. However, taking the limitations of this study and general reservations against AI in mind, it is

important to emphasize that the use of the software can only serve as a second opinion, and that physicians will always have to stay in charge.[25,42]

Supplementary material

The supplementary material contains more information regarding the training process and more detailed results of the reader study.

References

1. **Larsen CF**, **Lauritsen J**. Epidemiology of acute wrist trauma. *Int J Epidemiol*. 1993;22(5):911–916.
2. **Baron JA**, **Karagas M**, **Barrett J**, **et al**. Basic epidemiology of fractures of the upper and lower limb among Americans over 65 years of age. *Epidemiology*. 1996;7(6):612–618.
3. **Chung KC**, **Spilson SV**. The frequency and epidemiology of hand and forearm fractures in the United States. *J Hand Surg Am*. 2001;26(5):908–915.
4. **MacIntyre NJ**, **Dewan N**. Epidemiology of distal radius fractures and factors predicting risk and prognosis. *J Hand Ther*. 2016;29(2):136–145.
5. **Quadlbauer S**, **Pezzei C**, **Jurkowitsch J**, **et al**. Immediate mobilization of distal radius fractures stabilized by volar locking plate results in a better short-term outcome than a five week immobilization: a prospective randomized trial. *Clin Rehabil*. 2022;36(1):69–86.
6. **Quadlbauer S**, **Pezzei C**, **Jurkowitsch J**, **et al**. Functional and radiological outcome of distal radius fractures stabilized by volar-locking plate with a minimum follow-up of 1 year. *Arch Orthop Trauma Surg*. 2020;140(6):843–852.
7. **Court-Brown CM**, **Caesar B**. Epidemiology of adult fractures: a review. *Injury*. 2006;37(8):691–697.
8. **Nellans KW**, **Kowalski E**, **Chung KC**. The epidemiology of distal radius fractures. *Hand Clin*. 2012;28(2):113–125.
9. **Polinder S**, **Iordens GIT**, **Panneman MJM**, **et al**. Trends in incidence and costs of injuries to the shoulder, arm and wrist in The Netherlands between 1986 and 2008. *BMC Public Health*. 2013;13:531.
10. **No authors listed**. Independent statistics for evidence-based decision making. Statistics Austria. 2019. https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/demographische_prognosen/bevoelkerungsprognosen/027308.html (date last accessed 5 May 2020).
11. **Al-Hourani K**, **Tsang S-TJ**, **Simpson AHRW**. Osteoporosis: current screening methods, novel techniques, and preoperative assessment of bone mineral density. *Bone Joint Res*. 2021;10(12):840–843.
12. **Hernlund E**, **Svedbom A**, **Ivergård M**, **et al**. Osteoporosis in the European Union: medical management, epidemiology and economic burden. *Arch Osteoporos*. 2013;8(1–2):136.
13. **Pinto A**, **Berritto D**, **Russo A**, **et al**. Traumatic fractures in adults: missed diagnosis on plain radiographs in the Emergency Department. *Acta Biomed*. 2018;89(1-S):111–123.
14. **Tyson S**, **Hatem SF**. Easily missed fractures of the upper extremity. *Radiol Clin North Am*. 2015;53(4):717–736.
15. **Hallas P**, **Ellingsen T**. Errors in fracture diagnoses in the emergency department--characteristics of patients and diurnal variation. *BMC Emerg Med*. 2006;6:4.
16. **Pinto A**, **Reginelli A**, **Pinto F**, **et al**. Errors in imaging patients in the emergency setting. *Br J Radiol*. 2016;89(1061):20150914.
17. **Fernholm R**, **Pukk Härenstam K**, **Wachtler C**, **Nilsson GH**, **Holzmann MJ**, **Carlsson AC**. Diagnostic errors reported in primary healthcare and emergency departments: a retrospective and descriptive cohort study of 4830 reported cases of preventable harm in Sweden. *Eur J Gen Pract*. 2019;25(3):128–135.
18. **Guly HR**. Diagnostic errors in an accident and emergency department. *Emerg Med J*. 2001;18(4):263–269.
19. **Leeper WR**, **Leeper TJ**, **Vogt KN**, **Charyk-Stewart T**, **Gray DK**, **Parry NG**. The role of trauma team leaders in missed injuries: does specialty matter? *J Trauma Acute Care Surg*. 2013;75(3):387–390.
20. **Wei CJ**, **Tsai WC**, **Tiu CM**, **Wu HT**, **Chiou HJ**, **Chang CY**. Systematic analysis of missed extremity fractures in emergency radiology. *Acta Radiol*. 2006;47(7):710–717.
21. **Kung JW**, **Melenevsky Y**, **Hochman MG**, **et al**. On-call musculoskeletal radiographs: discrepancy rates between radiology residents and musculoskeletal radiologists. *AJR Am J Roentgenol*. 2013;200(4):856–859.
22. **Moonen PJ**, **Mercelina L**, **Boer W**, **Fret T**. Diagnostic error in the Emergency Department: follow up of patients with minor trauma in the outpatient clinic. *Scand J Trauma Resusc Emerg Med*. 2017;25(1):13.
23. **Dunnmon JA**, **Yi D**, **Langlotz CP**, **Ré C**, **Rubin DL**, **Lungren MP**. Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology*. 2019;290(2):537–544.
24. **Lee JH**, **Kim DH**, **Jeong SN**, **Choi SH**. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent*. 2018;77:106–111.
25. **Lisacek-Kiosoglous AB**, **Powling AS**, **Fontalis A**, **Gabr A**, **Mazomenos E**, **Haddad FS**. Artificial intelligence in orthopaedic surgery. *Bone Joint Res*. 2023;12(7):447–454.
26. **von Schacky CE**, **Sohn JH**, **Liu F**, **et al**. Development and validation of a multitask deep learning model for severity grading of hip osteoarthritis features on radiographs. *Radiology*. 2020;295(1):136–145.
27. **Xue Y**, **Zhang R**, **Deng Y**, **Chen K**, **Jiang T**. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS One*. 2017;12(6):e0178992.
28. **Kalmet PHS**, **Sanduleanu S**, **Primakov S**, **et al**. Deep learning in fracture detection: a narrative review. *Acta Orthop*. 2020;91(2):215–220.
29. **Duron L**, **Ducarouge A**, **Gillibert A**, **et al**. Assessment of an AI aid in detection of adult appendicular skeletal fractures by emergency physicians and radiologists: a multicenter cross-sectional diagnostic study. *Radiology*. 2021;300(1):120–129.
30. **Gan K**, **Xu D**, **Lin Y**, **et al**. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthop*. 2019;90(4):394–400.
31. **Jones RM**, **Sharma A**, **Hotchkiss R**, **et al**. Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs. *NPJ Digit Med*. 2020;3:144.
32. **Kim DH**, **MacKinnon T**. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol*. 2018;73(5):439–445.
33. **Urakawa T**, **Tanaka Y**, **Goto S**, **Matsuzawa H**, **Watanabe K**, **Endo N**. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol*. 2019;48(2):239–244.
34. **Chung SW**, **Han SS**, **Lee JW**, **et al**. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468–473.
35. **Guermazi A**, **Tannoury C**, **Kompel AJ**, **et al**. Improving radiographic fracture recognition performance and efficiency using artificial intelligence. *Radiology*. 2022;302(3):627–636.
36. **Lindsey R**, **Daluiski A**, **Chopra S**, **et al**. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A*. 2018;115(45):11591–11596.
37. **Olczak J**, **Pavlopoulos J**, **Prijs J**, **et al**. 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*. 2021;92(5):513–525.
38. **Williams SM**, **Connelly DJ**, **Wadsworth S**, **Wilson DJ**. Radiological review of accident and emergency radiographs: a 1-year audit. *Clin Radiol*. 2000;55(11):861–865.
39. **Leixnering M**, **Rosenauer R**, **Pezzei C**, **et al**. Indications, surgical approach, reduction, and stabilization techniques of distal radius fractures. *Arch Orthop Trauma Surg*. 2020;140(5):611–621.
40. **Quadlbauer S**, **Pezzei C**, **Jurkowitsch J**, **et al**. Early complications and radiological outcome after distal radius fractures stabilized by volar angular stable locking plate. *Arch Orthop Trauma Surg*. 2018;138(12):1773–1782.
41. **Rosenauer R**, **Pezzei C**, **Quadlbauer S**, **et al**. Complications after operatively treated distal radius fractures. *Arch Orthop Trauma Surg*. 2020;140(5):665–673.
42. **Clement ND**, **Simpson AHRW**. Artificial intelligence in orthopaedics. *Bone Joint Res*. 2023;12(8):494–496.

Author information

R. Breu, MD, Orthopaedic Surgeon,
Orthopedic Hospital Vienna-Speising, Vienna, Austria;
AUVA Trauma Hospital Lorenz Böhler, Vienna, Austria;
Ludwig Boltzmann Institute for Traumatology, the Research Center in Cooperation with AUVA, Vienna, Austria.

C. Avelar, MSc, Researcher
Z. Bertalan, PhD, Researcher
ImageBiopsy Lab, Vienna, Austria.

J. Grillari, PhD, Researcher, Ludwig Boltzmann Institute for Traumatology, the Research Center in Cooperation with AUVA, Vienna, Austria; Institute of Molecular Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria; Austrian Cluster for Tissue Regeneration, Vienna, Austria.

H. Redl, PhD, Researcher, Ludwig Boltzmann Institute for Traumatology, the Research Center in Cooperation with AUVA, Vienna, Austria; Austrian Cluster for Tissue Regeneration, Vienna, Austria.

R. Ljuhar, PhD, ImageBiopsy Lab, Vienna, Austria; Institute of Molecular Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria.

S. Quadlbauer, MD, Trauma Surgeon, Hand Specialist, AUVA Trauma Hospital Lorenz Böhler, Vienna, Austria.

T. Hausner, MD, AUVA Trauma Hospital Lorenz Böhler, Vienna, Austria; Ludwig Boltzmann Institute for Traumatology, the Research Center in Cooperation with AUVA, Vienna, Austria; Austrian Cluster for Tissue Regeneration, Vienna, Austria; Department for Orthopedic Surgery and Traumatology, Paracelsus Medical University, Salzburg, Austria.

Author contributions

R. Breu: Conceptualization, Data curation, Writing – original draft, Writing – review & editing, Formal analysis.
C. Avelar: Conceptualization, Data curation, Writing – review & editing, Software.
Z. Bertalan: Conceptualization, Software, Writing – review & editing.
J. Grillari: Supervision, Writing – review & editing.
H. Redl: Conceptualization, Supervision, Writing – review & editing.
R. Ljuhar: Conceptualization, Supervision, Writing – review & editing.
S. Quadlbauer: Supervision, Writing – review & editing.
T. Hausner: Conceptualization, Supervision, Writing – review & editing.

Funding statement

The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: funding was provided by "Wirtschaftsagentur Wien", ID 3178731 (Vienna Business Agency – Call Healthcare 2019), as reported by all authors.

ICMJE COI statement

C. Avelar, Z. Bertalan, and R. Ljuhar are employees of ImageBiopsy Lab. All other authors and the readers declare no conflict of interest.

Data sharing

The datasets generated and analyzed in the current study are not publicly available due to data protection regulations. Access to data is limited to the researchers who have obtained permission for data processing. Further inquiries can be made to the corresponding author.

Acknowledgements

The authors thank all readers: Kenneth Chen, Daniela Dziekan, Georg Garger, Tina Keuchel, Willi Müllbacher, Elena Nemecek, Josef Porta, Stefan Quadlbauer, Christoph Röder, Rudolf Rosenauer, and Sophie Wuthe. The authors also thank Stefan Salminger for helping with the labelling process, and Maximilian Kinsky for the English proofreading. None of those involved and mentioned here received financial or material support or are affiliated with ImageBiopsy Lab.

Ethical review statement

This retrospective study was approved by the local ethical review board of the Austrian Workers Compensation Board (AUVA), (26/2019).

Open access funding

The authors report that the open access funding for their manuscript was self-funded.

© 2024 Breu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See https://creativecommons.org/licenses/by-nc-nd/4.0/