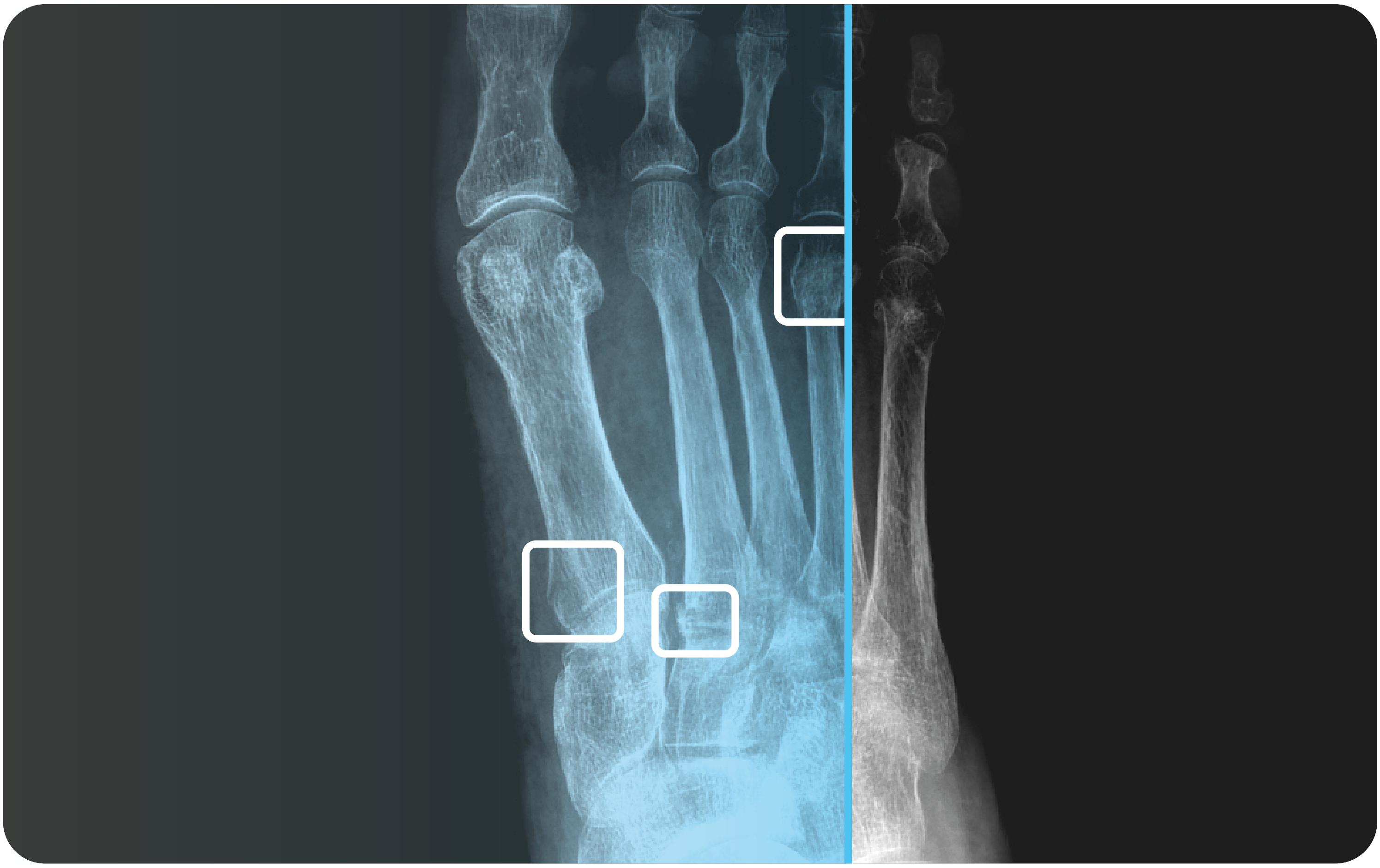
Radiobotics®



Results from Clinical Trial

RBfracture™ reduces the number of missed fractures on X-rays for emergency physicians and experienced reporters

Radiobotics has shown that their fracture decision support tool, RBfracture™ reduces the number of missed fractures by 42% whilst using 40% less time, thus improving the diagnostic performance of emergency and radiology staff. There is a great potential for improving patient care when deployed in a clinical setting.

Background and context

The X-ray is usually firstline interpreted by the clinician in the A&E department and later reported by staff in the radiology department. A missed fracture is the most common diagnostic error made in the A&E department, especially during out-of-hours where assistance from a radiologist or a more senior colleague might be unavailable.[1,2]

Recent studies have shown that AI algorithms for fracture detection are as good as humans and even perform better than non-expert readers.[3] The use of AI as decision support for fracture detection has shown great promise to improve the diagnostic performance of both experienced and non-experienced readers.[4]

In this study, we investigated how the use of RBfracture as a decision support tool changed the diagnostic performance of human readers working in the A&E and radiology department. The hypothesis was that RBfracture would significantly improve the sensitivity without compromising the specificity for detection of appendicular fractures in adult patients.

Methods

Data and study design

The study was a retrospective multireader multicase study where several human readers interpreted X-rays suspected of fractures with and without the support of RBfracture. Adult patients referred for an X-ray examination of the appendicular skeleton following a recent trauma were eligible for inclusion. Patients with orthopedic hardware, cast or splint and patients referred for followup were excluded. On a data level, images outside the intended use of RBfracture, e.g., spine or rib X-rays, and images with too poor quality were excluded.

Reference standard and fracture readings

Two experts with 11 and 8 years of MSK reporting experience independently evaluated all patient exams with all projections available. The original radiology reports were available to the experts who marked all visible fractures with a smallest possible bounding box while still ensuring that the entire fracture location was included. Acute and healing fractures were considered positive findings. The reference standard was defined by randomly selecting one of the bounding boxes, if the intersection over union (IoU) was greater than 25%. In case the IoU was less than 25%, the reference standard was adjudicated by a third expert with 18 years of reporting experience.

Four reporting radiographers and three A&E residents evaluated all patient exams on a digital platform in two separate sessions with at least eight weeks in between. In each session, all the radiographic views in an exam were available to the readers. All individual fractures were marked by placing a dot in the center of the fracture. No clinical information was available to the readers. In the first session, only the original X-rays were available for the readers (unaided session). In the second session, the original X-rays plus the RBfracture outputs were available to the readers (aided session). The time spent evaluating each patient exam was automatically registered in both sessions.

Statistical analysis

A true positive (TP) finding was defined as a reader dot located within a reference standard bounding box. A reader dot outside a reference standard bounding box was considered a false positive (FP).

The fracture-wise sensitivity was defined as the proportion of TPs amongst all fractures, counting multiple fractures per patient where appropriate. The average number of false-positive fractures per patient was defined as the total number of FP's divided by the number of patients. The patient-wise sensitivity was defined as the proportion of patients in whom all fractures are detected (each unique fracture in at least one radiograph). Note that this metric is not influenced by any potential false positive predictions. The patient-wise specificity was defined as the proportion of patients in whom no fracture dot was marked amongst patients without any fracture. Confidence intervals (CI) at the 95% level were estimated by resampling with replacement (bootstrap) 1000 times, where the size of the drawn sample was equal to the original sample size. The changes in reader sensitivity, specificity and average number of false positives were evaluated using one-sided, paired t-tests. The average time spent per patient exam was calculated for each reader in both the unaided and aided sessions. Differences in reading time were compared by means of two-sided, paired t-tests.

For all tests, p-values less than 0.05 were considered statistically significant.

 Study Results

Results

Data and demographics

In total 194 patient exams with an average of 3 views per exam were included. The mean age was 47 (range: 21-99) and 48% were females. A total of 89 unique fractures were found among the 76 fracture-positive patients. The distribution into different body parts are visualised in Table 1.

Body part	Positive Studies	Negative Studies	Total Studies
Shoulder/clavicle	8	7	15
Arm/Elbow	8	10	18
Wrist/Hand/Fingers	26	34	60
Hip/Pelvis	6	7	13
Leg/Knee	8	22	30
Ankle/Foot/Toes	20	38	58
Total	**76**	**118**	**194**

Table 1. Body parts and fracture prevalence

Fracture detection performance

The overall patient-wise sensitivity increased from 0.70 (CI: 0.66; 0.74) in the unaided readings to 0.83 (CI: 0.80; 0.86) when aided by RBfracture. The change between the unaided and aided session was +12.8% (CI: 4.5%; 20.8%) for the average pairwise difference in patient-wise sensitivity (p=0.004). The specificity remained unchanged with 0.88 (CI: 0.86; 0.90) and 0.90 (CI: 0.88; 0.92) in the unaided and aided session, respectively. The change in sensitivity and specificity is summarised in Table 2.
For the fracture-wise sensitivity a similar increase was observed with 0.72 (CI: 0.68;0.76) in the unaided session compared to 0.84 (CI: 0.81;0.87) in the RBfracture aided session (p=0.007). The number of false positives per patient remained unchanged, see Table 2.

An example of a fracture that was missed by readers in the unaided session, but diagnosed in the aided session is shown in Figure 1.

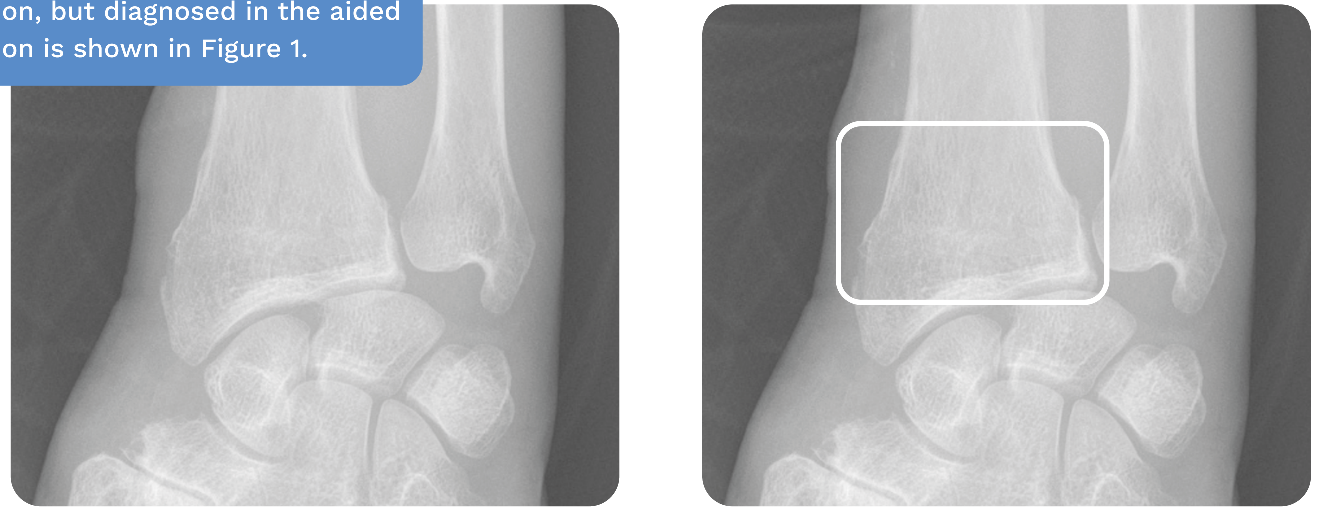


Figure 1. Wrist fracture.
The radiograph (left) shows a non-dislocated distal radius fracture. The fracture was missed by 6 readers in the unaided session. RBfracture correctly identifies the fracture and highlights it with a bounding box (right).

Figure 2 shows an example of toe trauma, where one out of two fractures was missed by all readers in the unaided session.

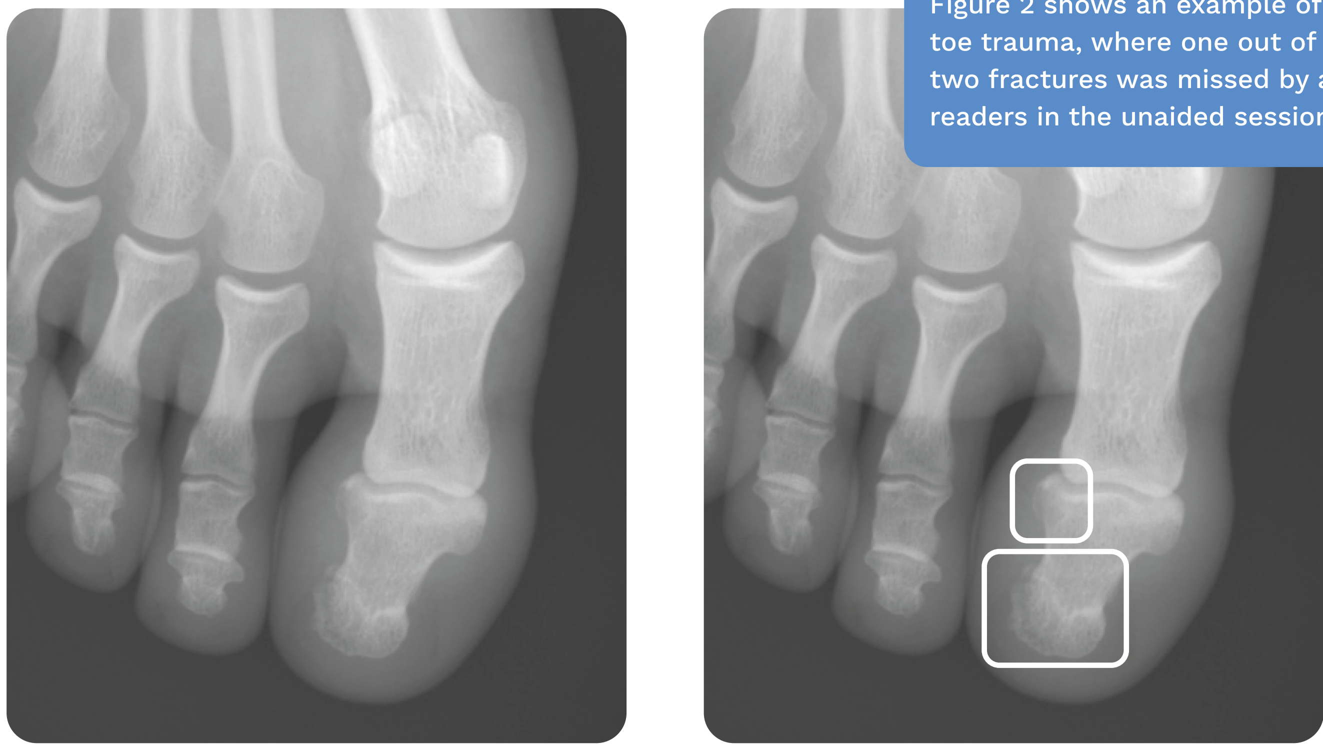


Figure 2. Toe fracture.
The radiograph (left) shows a comminuted fracture of the toe tuft and a non-dislocated fracture of the basis of the distal phalanx. The non-dislocated fracture was missed by all readers in the unaided session and is a classical example of the “satisfaction of search”-bias.

Study Results

The average number of missed fractures and relative change in number of missed fractures are shown in Table 3. A substantial reduction in missed fractures of 42% is observed for all readers.

The average reading time across all readers for the unaided session was 90±29 seconds. The average reading time in the aided session was 55±17 seconds. A consistent reduction in reading time was observed for all readers for aided exams with an average reduction in reading time of 36±23 seconds.

Subgroup analysis

The ability to detect fractures improved for both A&E residents and reporting radiographers, while the biggest improvement was observed for the junior doctors working in the A&E department. The average gain in absolute sensitivity was 19% for the A&E residents, whereas the gain for the reporting radiographers was 8%. See Table 2. The absolute reduction in number of missed fractures for reporting radiographers and A&E residents are shown in Table 3.

The reporting radiographers spent more time reading exams in the unaided session when compared to the A&E residents (109 seconds vs. 66 seconds). The difference was reduced for exams read in the aided mode (59 seconds vs. 49 seconds for reporting radiographers and A&E residents, respectively).

Reader and parameter	Unaided	Aided	Pairwise Difference
All Readers			
Patient-wise sensitivity	0.70 (0.66;0.74)	0.83 (0.80;0.86)	**0.13 (0.05;0.21)**
Patient-wise specificity	0.88 (0.86;0.90)	0.90 (0.88;0.92)	0.02 (-0.04;0.08)
Fracture-wise sensitivity	0.72 (0.68;0.76)	0.84 (0.81;0.87)	**0.12 (0.04;0.20)**
Average false positives per patient	0.12 (0.12;0.12)	0.09 (0.09;0.10)	-0.02 (-0.09;0.04)
Average reading time per exam (seconds)	90 [48;147]	55 [24;78]	**-36 [-79;-5]**
Reporting radiographers			
Patient-wise sensitivity	0.81 (0.77;0.85)	0.89 (0.85;0.92)	**0.08 (-0.01;0.17)**
Patient-wise specificity	0.94 (0.92;0.96)	0.94 (0.92;0.96)	0.0 (-0.04;0.05)
Fracture-wise sensitivity	0.83 (0.79;0.86)	0.89 (0.86;0.93)	**0.07 (-0.02;0.15)**
Average false positives per patient	0.06 (0.05;0.06)	0.06 (0.06;0.06)	0.0 (-0.05;0.06)
Average reading time per exam (seconds)	109 [85;147]	59 [38;68]	**-50 [32;79]**
Acute and emergency doctors			
Patient-wise sensitivity	0.56 (0.49;0.62)	0.75 (0.69;0.81)	**0.19 (0;0.39)**
Patient-wise specificity	0.81 (0.76;0.85)	0.85 (0.81;0.88)	0.04 (-0.19;0.27)
Fracture-wise sensitivity	0.58 (0.52;0.65)	0.76 (0.70;0.82)	**0.18 (-0.03;0.39)**
Average false positives per patient	0.21 (0.20;0.21)	0.14 (0.14;0.14)	-0.06 (-0.31;0.19)
Average reading time per exam (seconds)	66 [48;74]	49 [24;69]	**-17 [-24;-5]**

Numbers in parentheses are the 95% confidence interval. Numbers in brackets are the range. Boldface pairwise differences indicate p<0.05

Table 2. Reader performance

	Missed fractures unaided	Missed fractures aided	Relative reduction in missed fractures
All readers	24.7 [10;48]	14.4 [7;25]	**42%**
Reporting radiographers	15.5 [10;21]	9.5 [7;14]	**39%**
Acute and emergency doctors	37 [27;48]	21 [19;25]	**43%**

Numbers reported are the average [range] across readers

Table 3. Improvement in Fracture Detection with AI Assistance

Conclusion

Support from **RBfracture™** was able to significantly increase the sensitivity for fracture detection for both A&E residents and reporting radiographers without compromising the specificity, while also significantly reducing the time needed for interpretation. The improvement in diagnostic performance and reading time may lead to improvements in patient care through better initial diagnosis and reduced workload on the radiology staff, however this warrants further research in a prospective study.

Discussion

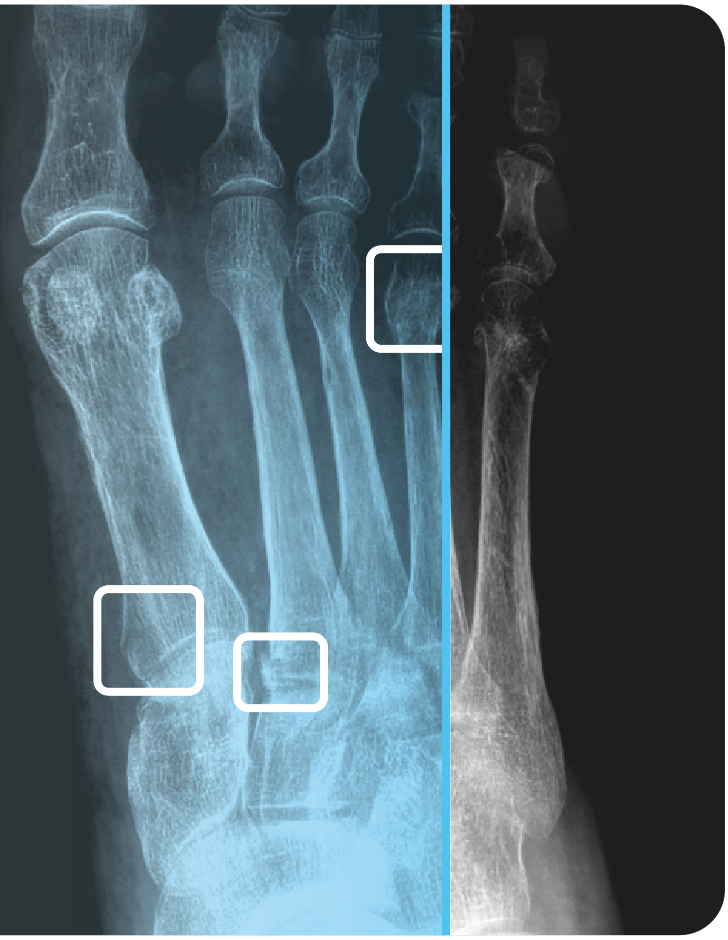
In this study we investigated the change in diagnostic performance when using RBfracture as a decision support tool for diagnosing fractures of the appendicular skeleton on X-rays. We showed significant improvement in both the sensitivity on both a patient- and fracture-level without compromising the specificity confirming the results in previously reported studies.[4,5]
The gain was most pronounced for residents working in the Acute & Emergency department with a relative reduction in missed fractures of 43%. These doctors are responsible for the clinical examination of the patients and most often doing firstline evaluation of the X-rays, even though not being radiology experts. A recent study showed that non-task expert doctors in particular benefit from correct explainable AI advice when interpreting X-rays.[6] Improving the diagnostic accuracy in the A&E department holds great promise to make sure that patients get the right diagnosis the first time, thus reducing the number of recalls and potential litigation cases.[7]
An increase in the diagnostics performance was also observed for reporting radiographers who improved their fracture-wise sensitivity from 0.83 to 0.89 without compromising the number of false-positive predictions. The increase in performance was achieved while also significantly reducing the time used to read the patient exams. The average reading time for the reporting radiographers was reduced by 46% for the patient exams aided by RBfracture. This could potentially reduce the overall turnaround time for reporting trauma X-rays and alleviate some of the urgent pressure on staff in radiology departments.[8] There are several limitations to the study.
First, no clinical information was available to the non-expert readers. In a real clinical setting, residents working in the A&E department have access to the patients and will obtain a trauma history and do clinical examination before looking at the X-rays. For the radiology staff, the clinical information is usually more sparse, although very important for accurate diagnosis, as this allows for a better correlation between clinical symptoms and X-rays findings, thus improving the diagnostic performance.[9]
Second, the reference standard was established without access to subsequent imaging, such as CT or MRI.

The original radiology reports were available to the ground truthers and some reports referred to follow up imaging. However, it cannot be ruled out that some occult fractures were left undetected in the dataset. Third, the order of the intervention was not randomised between the two sessions and this could create a carry-over effect from the unaided to the aided session, which could affect both performance and reading time estimates.

Finally, the patient exams were assessed retrospectively in a setting not representative of a clinical environment and this will affect the absolute reading time measured in our study, which does not account for actual reporting. However, the relative change in reading time is assumed to not be affected by this. The retrospective nature of the study could also introduce a laboratory bias as observed in other radiology studies.[10]

References

1. Hallas, P. & Ellingsen, T. Errors in fracture diagnoses in the emergency department -- characteristics of patients and diurnal variation. BMC Emerg. Med. 6, 4 (2006).
2. Mattijssen-Horstink, L. et al. Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency department: a retrospective analysis. Scand. J. Trauma Resusc. Emerg. Med. 28, 38 (2020).
3. Zhang, X. et al. Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: a systematic literature review and meta-analysis. Eur. Radiol. 32, 7196–7216 (2022).
4. Kuo, R. Y. L. et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. Radiology 211785 (2022).
5. Oppenheimer, J., Lüken, S., Hamm, B. & Niehues, S. M. A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. Life 13, (2023).
6. Gaube, S. et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. Sci. Rep. 13, 1–11 (2023).
7. Atrey, A., Gupte, C. M. & Corbett, S. A. Review of successful litigation against english health trusts in the treatment of adults with orthopaedic pathology: clinical governance lessons learned. J. Bone Joint Surg. Am. 92, e36 (2010).
8. The Royal College of Radiologists. Clinical radiology census report 2021. https://www.rcr.ac.uk/sites/default/files/clinical_radiology_census_report_2021.pdf
9. Castillo, C., Steffens, T., Sim, L. & Caffery, L. The effect of clinical information on radiology reporting: A systematic review. J Med Radiat Sci 68, 60–74 (2021).
10. Gur, D. et al. The 'laboratory' effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology 249, 47–53 (2008).



About Radiobotics

Radiobotics is a multiple award-winning health tech company with their HQ in Denmark.
The company has built an innovative AI technology specialized in X-ray analysis with focus on musculo-skeletal radiology. Based on advanced computer vision and machine learning methods, Radiobotics' algorithms generate fully automated, objective text and visual reports.

Contact

info@radiobotics.com
www.radiobotics.com