Use of AI in radiology: Automatic fracture detection

Cruising into the future

Koen Mermuys MD, Pieter Vandaele MD

06/12/2023 VIVES Brugge

Radiology AZ Sint-Lucas - Brugge, BELGIUM
Radiology AZ St-Jan - Brugge BELGIUM

j. delanote, p. dekimpe, t. de schrijver, a. favril, b. geerts, k. geldof, j. ghekiere, k. gieraerts
h. goris, j. kager, c. kenis, j. marrannes, k. mermuys, k. ramboer, n.salamon, k. vanslambrouck,
p. vandaele, k. van de moortele, a. van holsbeeck, l. verhaeghe, m.-s. walgraeve, c. werbrouck

Missed fractures in the ID are not uncommon

Increasingly busy emergency departments
Increasing flood of radiographs
Faster accessibility of health care data / results
Readouts frequently during night
Trainees commonly involved in initial reading

2012 – 2017, Nijmegen, teaching hospital
25.957 fractures, 289 missed

49% of missed fractures between 4 – 7PM
(occupational fatigue) and around 3AM

AI tool to improve fracture detection ?
AI tool to triage radiographs into those that need immediate attention ?

Mattijssen-Horstink *et al. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* (2020) 28:38
https://doi.org/10.1186/s13049-020-00727-8

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine

ORIGINAL RESEARCH Open Access

Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency department: a retrospective analysis

Laura Mattijssen-Horstink[1*], Judith Joëlle Langeraar[2], Gert Jan Mauritz[2], William van der Stappen[3], Maarten Baggelaar[4] and Edward Camillus Thwan Han Tan[5,6]

Abstract

Background: Missed fractures in the emergency department (ED) are common and may lead to patient morbidity.

Aim: To determine the rate and nature of radiographic discrepancies between ED treating physicians, radiologists and trauma/orthopaedic surgeons and the clinical consequences of delayed diagnosis. A secondary outcome measurement is the timeframe in which most fractures were missed.

Methods: A single-centre retrospective analysis of all missed fractures in a general teaching hospital from 2012 to 2017 was performed. Data regarding missed fractures were provided by the hospital's complication list and related database. Additional data were retrieved from the electronic medical records as required for the study.

Results: A total of 25,957 fractures were treated at our ED. Initially, 289 fractures were missed by ED treating physicians (1.1%). The most frequently missed fractures were the elbow (28.6%) and wrist (20.8%) in children, the foot (17.2%) in adults and the pelvis and hip (37.3%) in elderly patients. Patients required surgery in 9.3% of missed fractures, received immobilization by a cast or brace in 45.7%, had no treatment alterations during the first week in 38.1%. Follow-up data were lacking for 6.9% of cases. 49% of all missed fractures took place between 4 PM and 9 PM. There is a discrepancy in percentages of correctly diagnosed fractures and missed fractures between 5 PM and 3 AM.

Conclusion: Adequate training of ED treating physicians in radiographic interpretation is essential in order to increase diagnostic accuracy. A daily multidisciplinary radiology meeting is very effective in detecting missed fractures.

Keywords: Emergency department, Fracture, Radiographs, Radiologic discrepancies, Missed diagnosis, Diagnostic error

24 physicians varying background
(non radiologists and trainees)

480 radiographs

AI improved sens / spec

Sensitivity from 64,8% to 75,2% = 10,4% higher
Specificty from 90,6% tot 95,6% = 5% higher

Increase of 10,4% in sensitivity

Radiology

ORIGINAL RESEARCH · MUSCULOSKEL

Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence

Ali Guermazi, MD, PhD • Chadi Tannoury, MD • Andrew J. Kompel, MD • Akira M. Murakami, MD • Alexis Ducarouge, MSc • André Gillibert, MD, MSc • Xinning Li, MD • Antoine Tournier, MSc • Youmna Lahoud, MD • Mohamed Jarraya, MD • Elise Lacave, MSc • Hamza Rahimi, MD • Aloïs Pourchot, MSc • Robert L. Parisien, MD • Alexander C. Merritt, MD • Douglas Comeau, DO • Nor-Eddine Regnard, MD, MSc • Daichi Hayashi, MD, PhD

From the Departments of Radiology (A. Guermazi, A.J.K., A.M.M., H.R., A.C.M., D.H.), Orthopaedic Surgery (C.T., X.L.), and Family Medicine (D.C.), Boston University School of Medicine, Boston, Mass; Department of Radiology, VA Boston Healthcare System, 1400 VFW Parkway, Suite 1B105, West Roxbury, MA 02132 (A. Guermazi); Gleamer, Paris, France (A.D., A.T., E.L., A.P., N.E.); Department of Biostatistics, CHU Rouen, Rouen, France (A. Gillibert); Department of Rheumatology, Harvard Vanguard Medical Associates, Braintree, Mass (Y.L.); Department of Radiology, Musculoskeletal Division, Massachusetts General Hospital, Harvard Medical School, Boston, Mass 02114 (M.J.); Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, Paris, France (A.P.); Department of Orthopaedic Surgery, The Mount Sinai Hospital, New York, NY (R.L.P.); University Health Services and Primary Care Sports Medicine, Boston College, Chestnut Hill, Mass (D.C.); and Department of Radiology, Stony Brook University Renaissance School of Medicine, Stony Brook, NY (D.H.). Received April 15, 2021; revision requested June 21; revision received October 13; accepted October 28. **Address correspondence to** A. Guermazi. (e-mail: *guermazi@bu.edu*).

Supported by Gleamer, which developed the AI and software and built the study sample and design.

Conflicts of interest are listed at the end of this article.

See also the editorial by Link and Pedoia in this issue.

Radiology 2022; 302:627–636 • https://doi.org/10.1148/radiol.210937 • Content codes: MK AI

Background: Missed fractures are a common cause of diagnostic discrepancy between initial radiographic interpretation and the final read by board-certified radiologists.

Purpose: To assess the effect of assistance by artificial intelligence (AI) on diagnostic performances of physicians for fractures on radiographs.

Materials and Methods: This retrospective diagnostic study used the multi-reader, multi-case methodology based on an external multicenter data set of 480 examinations with at least 60 examinations per body region (foot and ankle, knee and leg, hip and pelvis, hand and wrist, elbow and arm, shoulder and clavicle, rib cage, and thoracolumbar spine) between July 2020 and January 2021. Fracture prevalence was set at 50%. The ground truth was determined by two musculoskeletal radiologists, with discrepancies solved by a third. Twenty-four readers (radiologists, orthopedists, emergency physicians, physician assistants, rheumatologists, family physicians) were presented the whole validation data set (n = 480), with and without AI assistance, with a 1-month minimum washout period. The primary analysis had to demonstrate superiority of sensitivity per patient and the noninferiority of specificity per patient at −3% margin with AI aid. Stand-alone AI performance was also assessed using receiver operating characteristic curves.

Results: A total of 480 patients were included (mean age, 59 years ± 16 [standard deviation]; 327 women). The sensitivity per patient was 10.4% higher (95% CI: 6.9, 13.9; $P < .001$ for superiority) with AI aid (4331 of 5760 readings, 75.2%) than without AI (3732 of 5760 readings, 64.8%). The specificity per patient with AI aid (5504 of 5760 readings, 95.6%) was noninferior to that without AI aid (5217 of 5760 readings, 90.6%), with a difference of +5.0% (95% CI: +2.0, +8.0; $P = .001$ for noninferiority). AI shortened the average reading time by 6.3 seconds per examination (95% CI: −12.5, −0.1; $P = .046$). The sensitivity by patient gain was significant in all regions (+8.0% to +16.2%; $P < .05$) but shoulder and clavicle and spine (+4.2% and +2.6%; $P = .12$ and .52).

"The interpretation of radiographs suffers from an ever-increasing workload in emergency and radiology departments while missed fractures represent up to 80% of diagnostic errors in the emergency department."

600 patients
Radiologists and emergency physicians

Improved sensitivity in fracture detection of 8,7%

Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study

Loïc Duron, MD, MSc • Alexis Ducarouge, MSc • André Gillibert, MD, MSc • Julia Lainé, MD, MSc • Christian Allouche • Nicolas Cherel, MSc • Zekun Zhang, MSc • Nicolas Nitche, MSc • Elise Lacave, MSc • Aloïs Pourchot, MSc • Adrien Felter, MD • Louis Lassalle, MD, MSc • Nor-Eddine Regnard, MD, MSc • Antoine Feydy, MD, PhD

From the Department of Radiology, Hôpital Fondation A. de Rothschild, 25 rue Manin, 75019 Paris, France (L.D.); Faculty of Medicine, Université de Paris, Paris France (L.D., A. Feydy); Gleamer, Paris, France (A.D., C.A., N.C., Z.Z., N.N., E.L., A.P., N.E.R.); Department of Biostatistics, CHU Rouen, Rouen, France (A.G.) Department of Radiology, Hôpital Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, Paris, France (J.L.); Department of Radiology, Hôpital Ambroise-Paré, Assistanc Publique-Hôpitaux de Paris, Boulogne-Billancourt, France (A. Felter); Department of Radiology, Hôpital Raymond-Poincaré, Assistance Publique-Hôpitaux de Paris Garches, France (A. Felter); and Department of Radiology B, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France (L.L., N.E.R., A. Feydy). Received September 30, 2020; revision requested December 23; revision received January 26, 2021; accepted March 4. **Address correspondence to** L.D. (e-mail: *lduron@for.paris*)

This study was funded by Gleamer.

Conflicts of interest are listed at the end of this article.

Radiology 2021; 300:120–129 • https://doi.org/10.1148/radiol.2021203886 • Content codes: MK AI

Background: The interpretation of radiographs suffers from an ever-increasing workload in emergency and radiology departments, while missed fractures represent up to 80% of diagnostic errors in the emergency department.

Purpose: To assess the performance of an artificial intelligence (AI) system designed to aid radiologists and emergency physicians in the detection and localization of appendicular skeletal fractures.

Materials and Methods: The AI system was previously trained on 60 170 radiographs obtained in patients with trauma. The radiographs were randomly split into 70% training, 10% validation, and 20% test sets. Between 2016 and 2018, 600 adult patients in whom multiview radiographs had been obtained after a recent trauma, with or without one or more fractures of shoulder, arm, hand, pelvis, leg, and foot, were retrospectively included from 17 French medical centers. Radiographs with quality precluding human interpretation or containing only obvious fractures were excluded. Six radiologists and six emergency physicians were asked to detect and localize fractures with (n = 300) and fractures without (n = 300) the aid of software highlighting boxes around AI-detected fractures. Aided and unaided sensitivity, specificity, and reading times were compared by means of paired Student t tests after averaging of performances of each reader.

Results: A total of 600 patients (mean age ± standard deviation, 57 years ± 22; 358 women) were included. The AI aid improved the sensitivity of physicians by 8.7% (95% CI: 3.1, 14.2; P = .003 for superiority) and the specificity by 4.1% (95% CI: 0.5, 7.7; P < .001 for noninferiority) and reduced the average number of false-positive fractures per patient by 41.9% (95% CI: 12.8, 61.3; P = .02) in patients without fractures and the mean reading time by 15.0% (95% CI: −30.4, 3.8; P = .12). Finally, stand-alone performance of a newer release of the AI system was greater than that of all unaided readers, including skeletal expert radiologists, with an area under the receiver operating characteristic curve of 0.94 (95% CI: 0.92, 0.96).

Conclusion: The artificial intelligence aid provided a gain of sensitivity (8.7% increase) and specificity (4.1% increase) without loss of

A.I. fracture detection

Rayvolve by AZmed

The powerful computer-aided diagnosis tool that detects all types of fractures on X-rays

20 % errors avoided

36 % time saved

Used in az sint lucas since septembre 2021

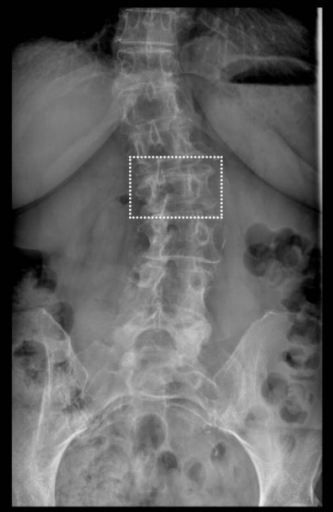
Fusion radiology departments st jan – st lucas in since 2021.
Large investment of both hospitals in a new overall ONE PACS system.
Go live st jan – st lucas in Q1 2024.

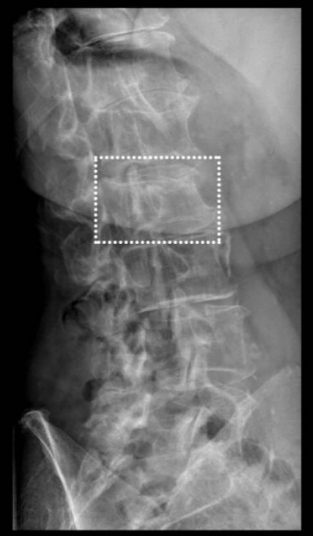
Attention required

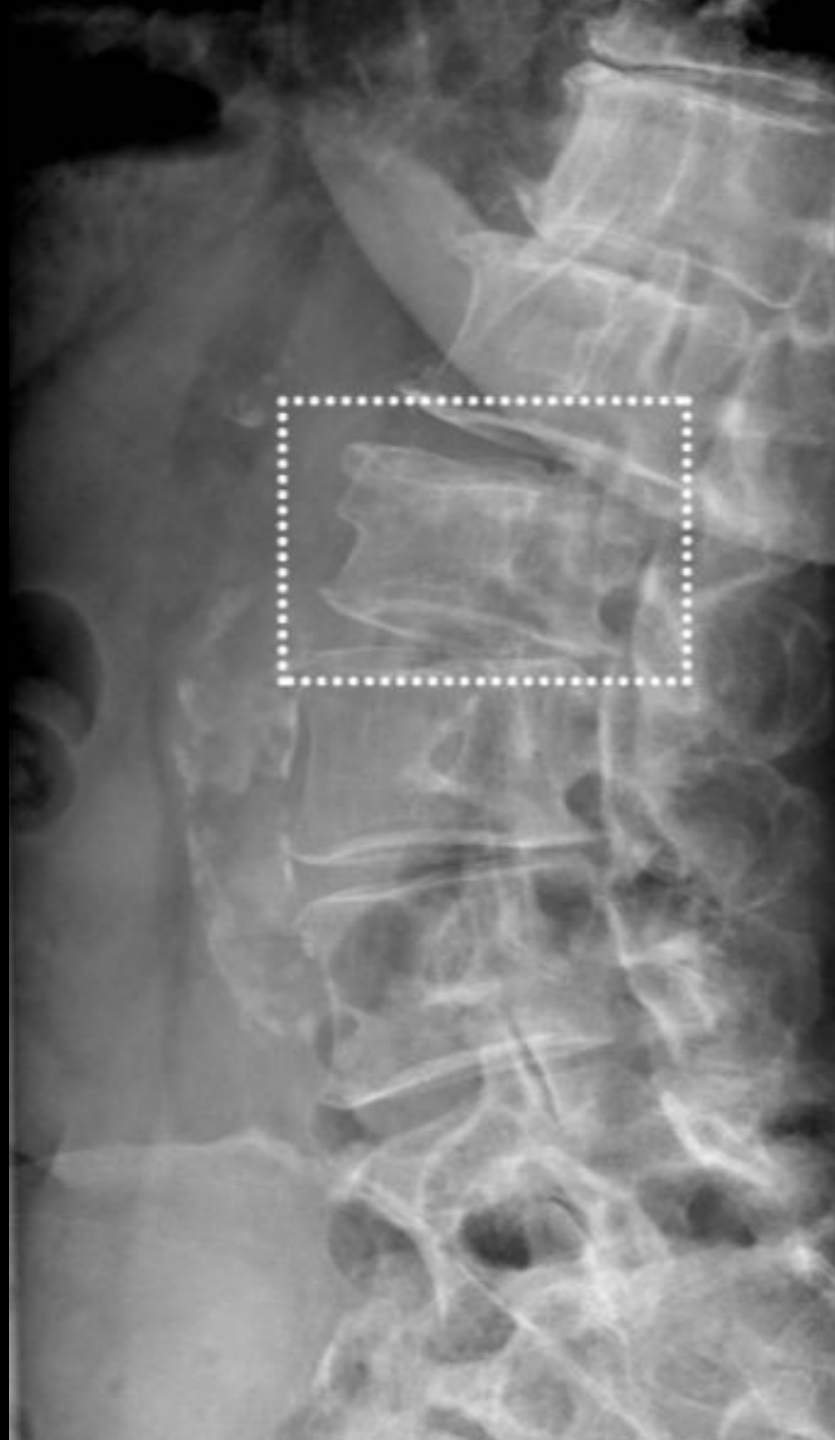
AZMED

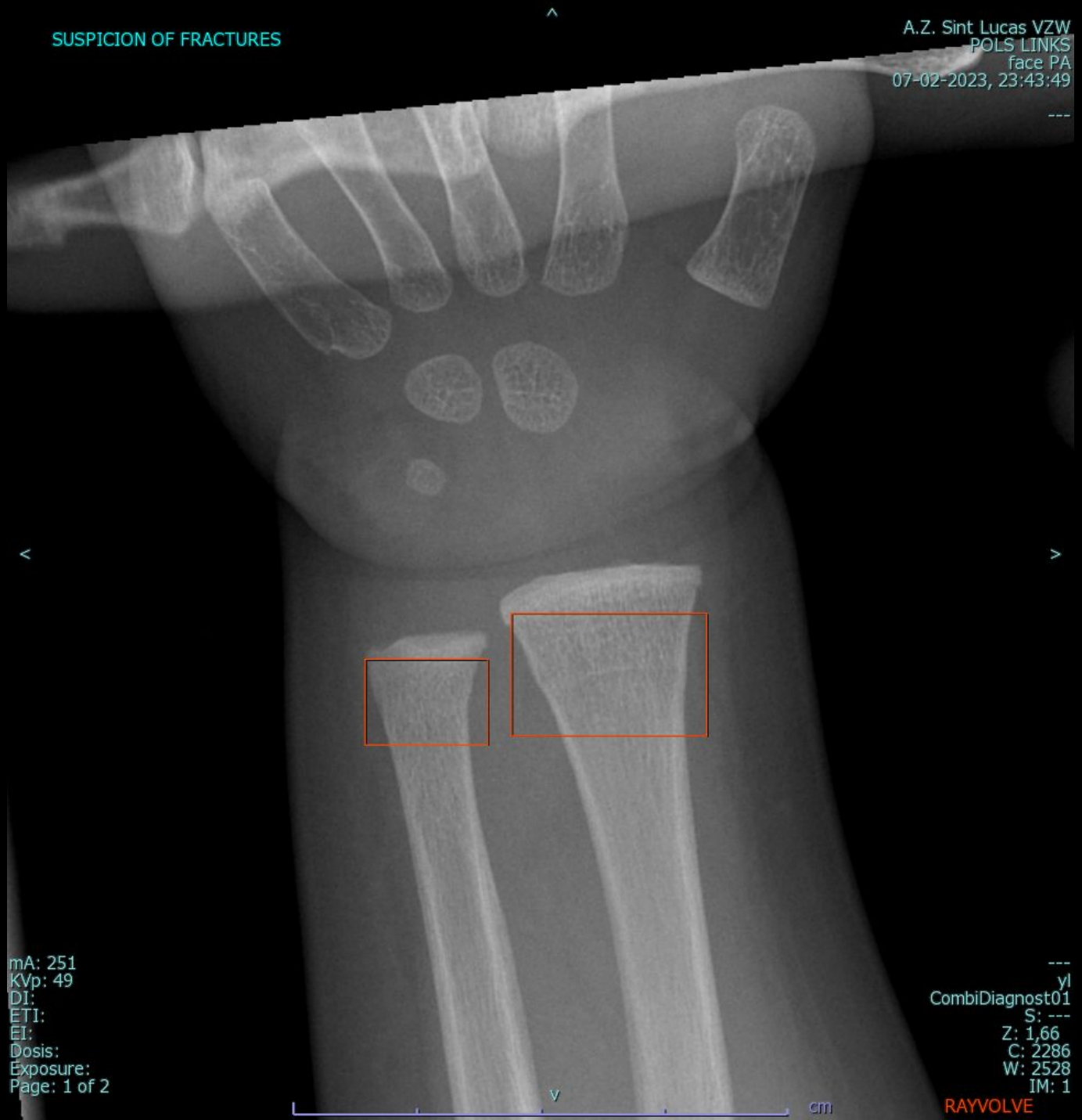
RAYVOLVE
TRAUMA v.3.0.1

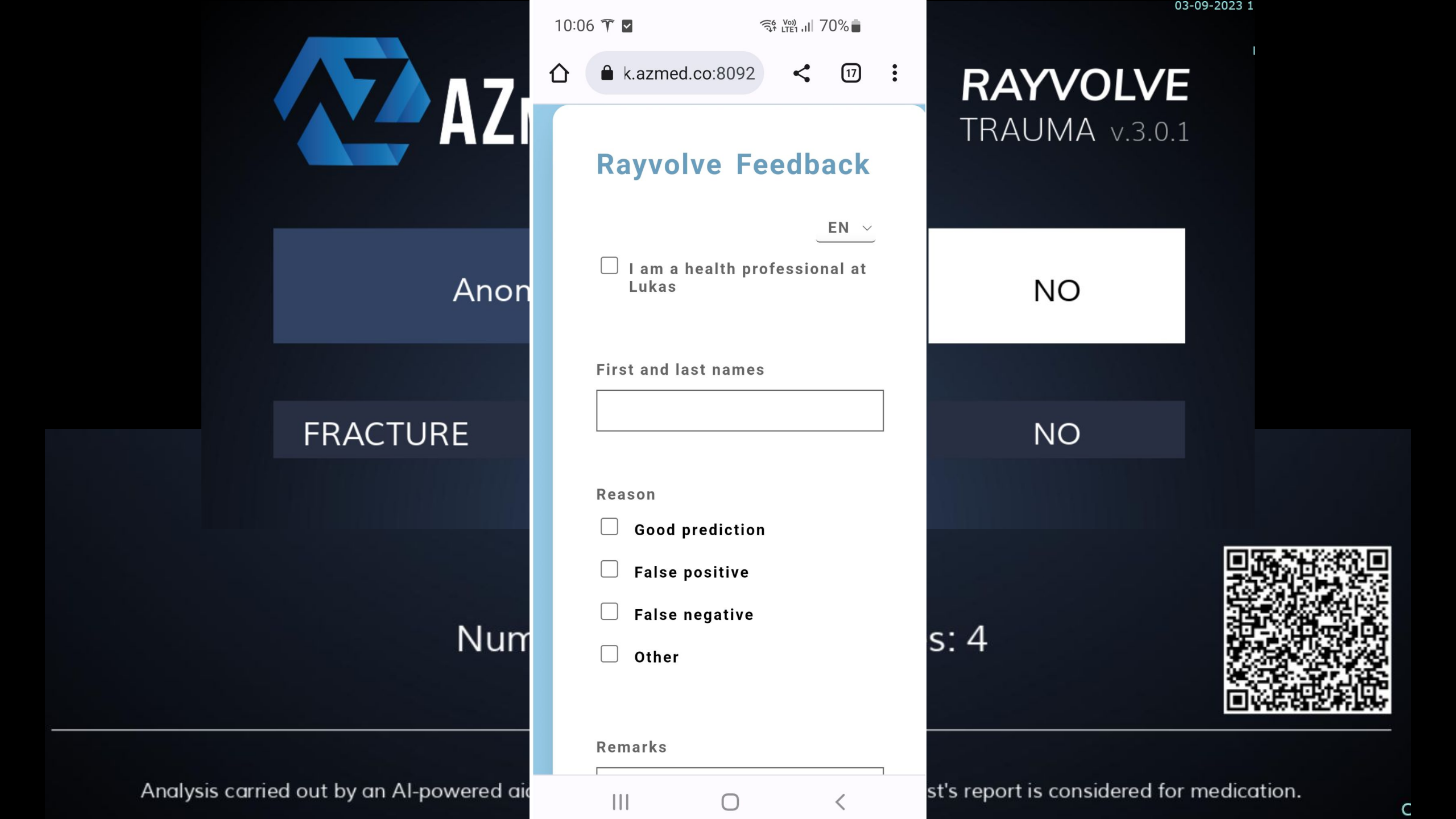
LOW SUSPICION



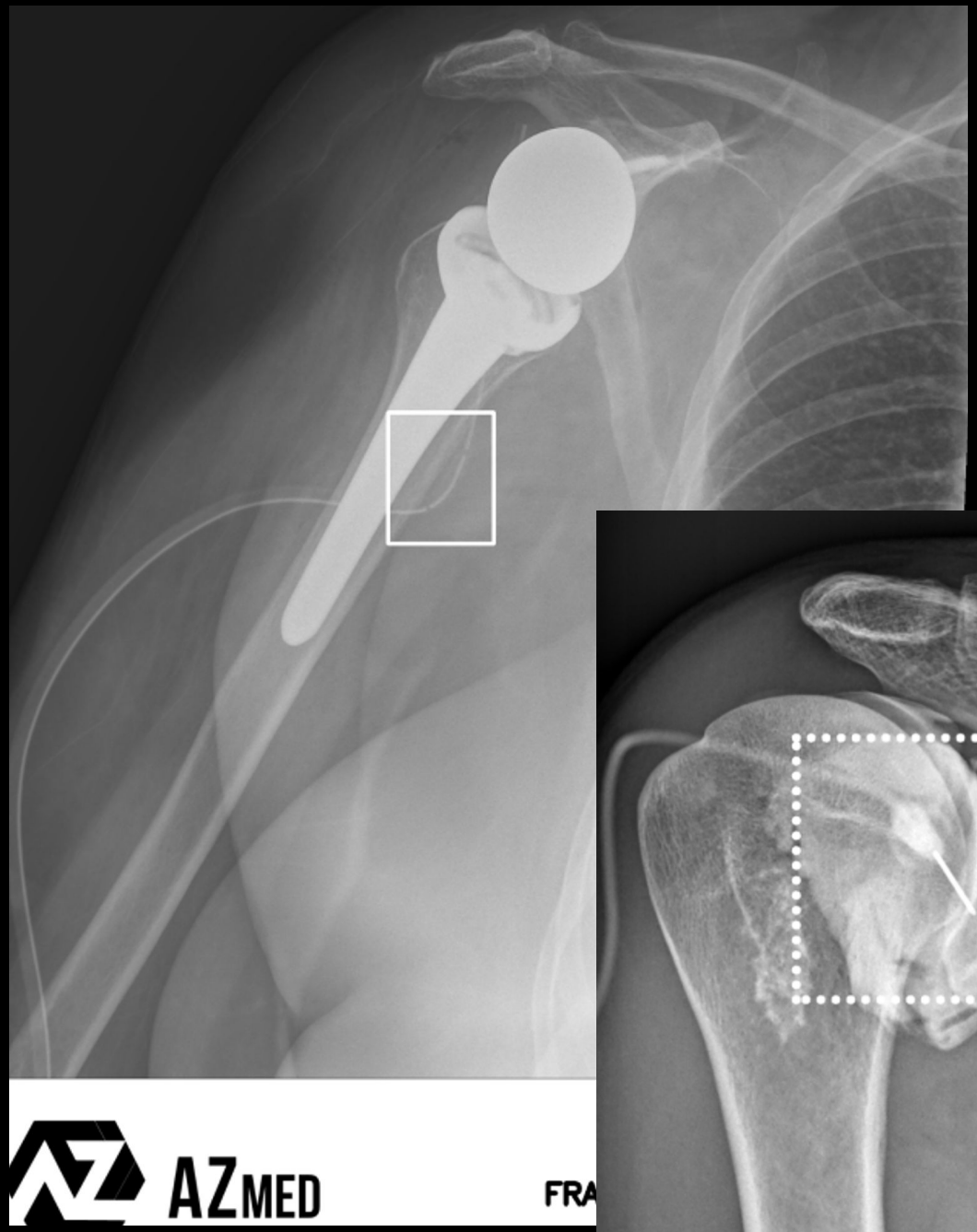


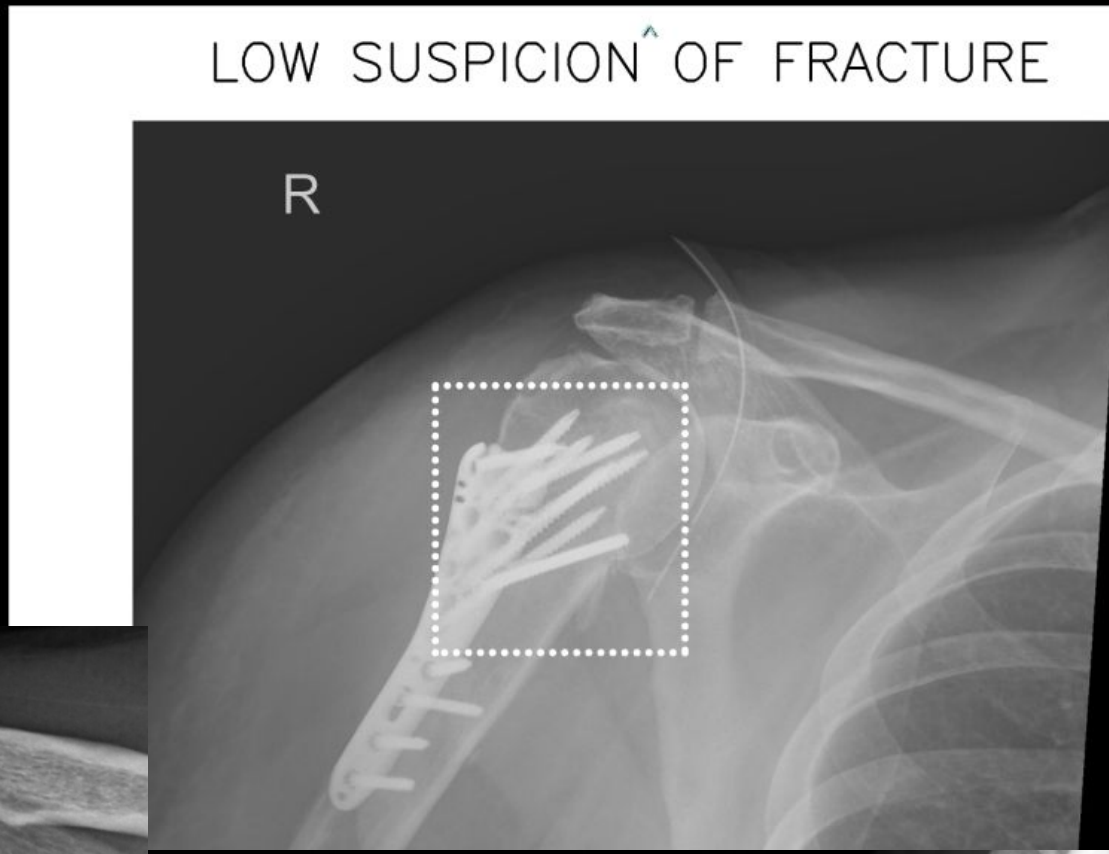


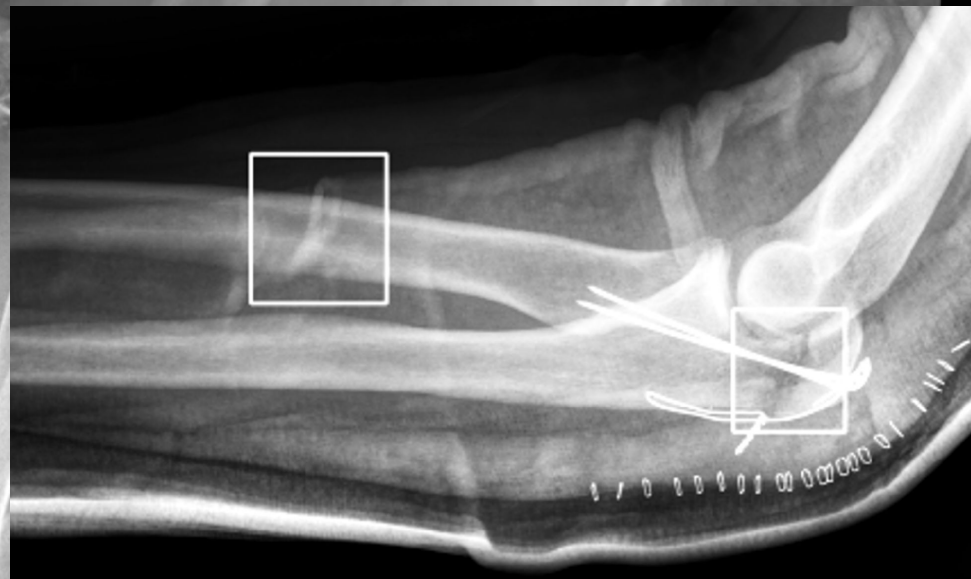

SUSPICION OF FRACTURES
A.Z. Sint Lucas VZW
POLS LINKS
face PA
07-02-2023, 23:43:49
---
mA: 251
KVp: 49
DI:
ETI:
EI:
Dosis:
Exposure:
Page: 1 of 2
---
yl
CombiDiagnost01
S: ---
Z: 1,66
C: 2286
W: 2528
IM: 1
RAYVOLVE
cm

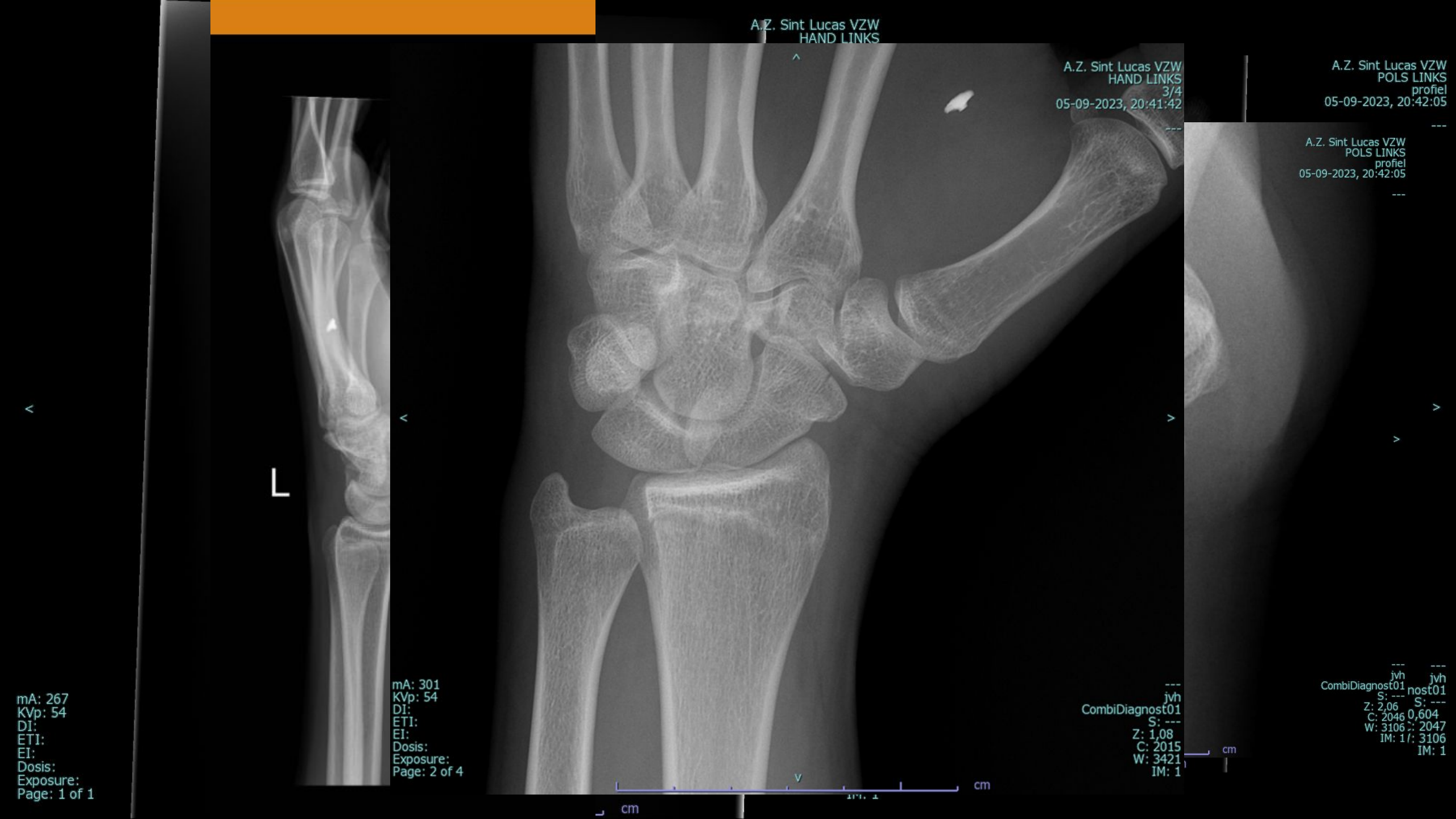

03-09-2023 1
RAYVOLVE
TRAUMA v.3.0.1
NO
FRACTURE
NO
Num
s: 4
Analysis carried out by an AI-powered aid
st's report is considered for medication.
10:06
70%
k.azmed.co:8092
Rayvolve Feedback
EN
I am a health professional at Lukas
First and last names
Reason
Good prediction
False positive
False negative
Other
Remarks

A.I. Learning by feed back








A.Z. Sint Lucas VZW
HAND LINKS
A.Z. Sint Lucas VZW
HAND LINKS
3/4
05-09-2023, 20:41:42

A.Z. Sint Lucas VZW
POLS LINKS
profiel
05-09-2023, 20:42:05

A.Z. Sint Lucas VZW
POLS LINKS
profiel
05-09-2023, 20:42:05

L
mA: 267
KVp: 54
DI:
ETI:
EI:
Dosis:
Exposure:
Page: 1 of 1
mA: 301
KVp: 54
DI:
ETI:
EI:
Dosis:
Exposure:
Page: 2 of 4
---
jvh
CombiDiagnost01
S: ---
Z: 1,08
C: 2015
W: 3421
IM: 1
---
jvh
CombiDiagnost01
S: ---
Z: 2,06
C: 2046
W: 3106
IM: 1
cm
cm
cm


R

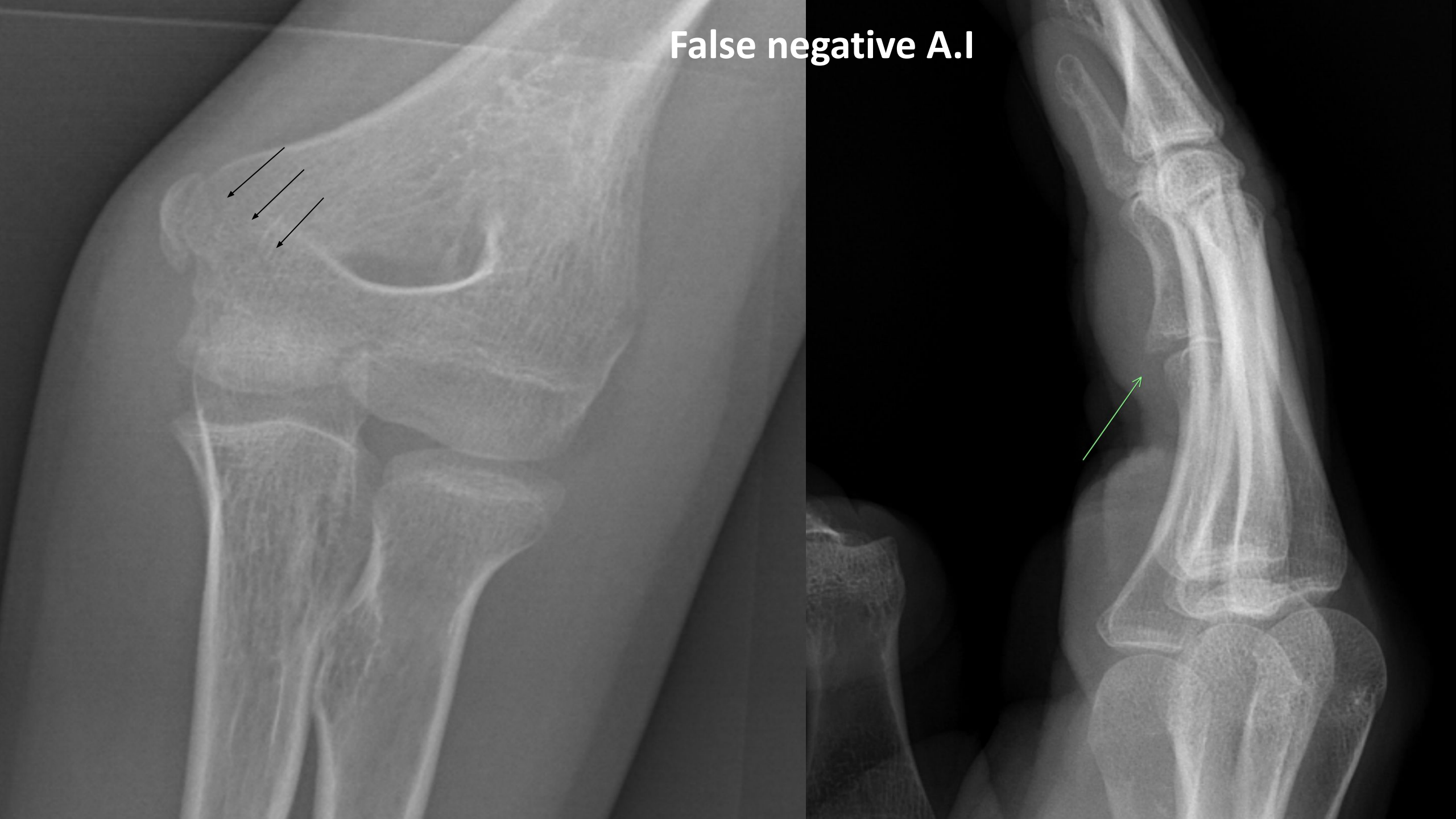

R

AZMED

Attention required

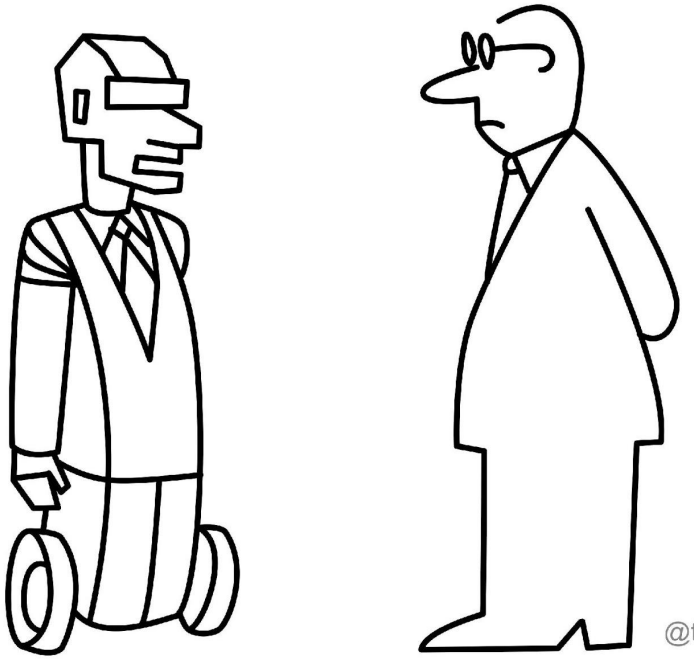

R

False negative A.I



In Conclusion:
AI adresses multiple pertinent issues:

1/ Increasing workload
2/ Occupational fatigue
3/ Helping trainees and ED physicians better detect fractures
4/ Triage of radiographs that need immediate attention
5/ Faster readouts



"The good news is I have discovered inefficiencies. The bad news is that you're one of them."

AI as stand alone?

We are still far from a "general" AI

1/ At this time numerous "one task only AI's (fractures, bone-age, brain, breast).
2/ No diagnosis of other pertinent abnormalities (tumours, infectious, inflammatory).
3/ AI does not have a estimate of uncertainty ('I do not know') which can provide a false sense of certainty.

A.I. in our department

- Bone age (BoneXpert)
- Fracture detection (Rayvolve)
- Brain stroke (Aspects/perfusion/angio) (RAPID)
- Breast (ICAD)
- Deep learning reconstruction on (DLR) techniques MR/CT

Dep. Radiology, St-Lucas Hospital, Bruges – BELGIUM
Dep. Radiology St-Jan Hospital, Bruges – BELGIUM


The role requires a high-level of soft skills. So, tell me, can you persuade, influence, empathise, inspire, build rapport, manage emotions and behaviours plus maintain a sense of humour under pressure?
Umm... no.
CV
ROBOT RR76-Q
GUY DOWNES ©
officeguycartoons.com

Thank you