ESPR
European Society of Paediatric Radiology

Assessment of an AI aid in detection of pediatric appendicular skeletal fractures by senior and junior radiologists

T. Nguyen, R. Maarek, A.-L. Hermann, H. Ducou Le Pointe
Trousseau Hospital, Paris

GLEAMER

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Conflicts of interest

- Consulting missions for Gleamer

GLEAMER

Background

- Few articles studying fracture detection in pediatric population
- Only about elbow fractures : encouraging results (Se 0,91-0,93, Sp 0,84-0,92) but
  - Low PPV : sorting role
  - Standalone studies

DATASET CONSTITUTION

1 US-based data provider

Inclusion of
- Anonymized standard X-ray exams of limbs with or without fracture(s)
- Pediatric patient, age between 2 to 21

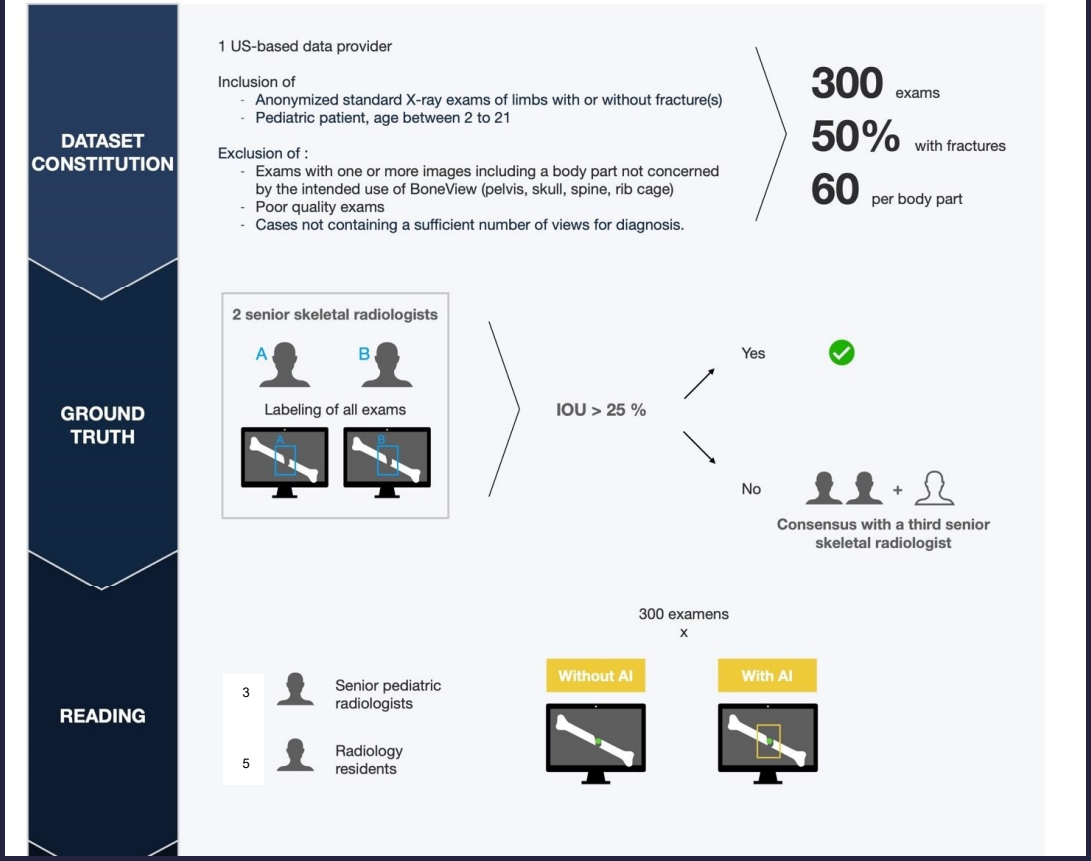
Exclusion of :
- Exams with one or more images including a body part not concerned by the intended use of BoneView (pelvis, skull, spine, rib cage)
- Poor quality exams
- Cases not containing a sufficient number of views for diagnosis.

300 exams

50% with fractures

60 per body part


DATASET CONSTITUTION
1 US-based data provider
Inclusion of
- Anonymized standard X-ray exams of limbs with or without fracture(s)
- Pediatric patient, age between 2 to 21
Exclusion of :
- Exams with one or more images including a body part not concerned by the intended use of BoneView (pelvis, skull, spine, rib cage)
- Poor quality exams
- Cases not containing a sufficient number of views for diagnosis.
300 exams
50% with fractures
60 per body part
GROUND TRUTH
2 senior skeletal radiologists
A
B
Labeling of all exams
IOU > 25 %
Yes
No
+
Consensus with a third senior skeletal radiologist


DATASET CONSTITUTION
1 US-based data provider
Inclusion of
- Anonymized standard X-ray exams of limbs with or without fracture(s)
- Pediatric patient, age between 2 to 21
Exclusion of :
- Exams with one or more images including a body part not concerned by the intended use of BoneView (pelvis, skull, spine, rib cage)
- Poor quality exams
- Cases not containing a sufficient number of views for diagnosis.
300 exams
50% with fractures
60 per body part
GROUND TRUTH
2 senior skeletal radiologists
A
B
Labeling of all exams
IOU > 25 %
Yes
No
+
Consensus with a third senior skeletal radiologist
READING
300 examens
x
Without AI
With AI
3
Senior pediatric radiologists
5
Radiology residents

DATASET CONSTITUTION

1 US-based data provider

Inclusion of
- Anonymized standard X-ray exams of limbs with or without fracture(s)
- Pediatric patient, age between 2 to 21

Exclusion of :
- Exams with one or more images including a body part not concerned by the intended use of BoneView (pelvis, skull, spine, rib cage)
- Poor quality exams
- Cases not containing a sufficient number of views for diagnosis.

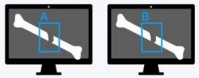
300 exams
50% with fractures
60 per body part

GROUND TRUTH

2 senior skeletal radiologists

A B

Labeling of all exams



IOU > 25 %

Yes

No +

Consensus with a third senior skeletal radiologist

READING

300 examens
x

3	Senior pediatric radiologists
5	Radiology residents

Without AI

With AI

ANALYSIS

- Patient-wise sensitivity
- Lesion-wise sensitivity
- Specificity
- Per patient false positive rate with fractures
- Per patient false positive rate without fractures

- By readers' level
- By body part
- By age
- By fracture type

Results

	Unaided	Aided	Absolute difference
SE_{PW}	73.17	82.67	9.5 [7.05, 11.95] (p<0.001)
SPE_{PW}	89.58	90.33	0.75 [-0.75, 2.25] (p=0.28)

Results

	Unaided	Aided	Absolute difference
SE_{PW}	73.17	82.67	9.5 [7.05, 11.95] (p<0.001)
SPE_{PW}	89.58	90.33	0.75 [-0.75, 2.25] (p=0.28)

Results

	Unaided	Aided	Absolute difference
SE_{PW}	73.17	82.67	9.5 [7.05, 11.95] (p<0.001)
SPE_{PW}	89.58	90.33	0.75 [-0.75, 2.25] (p=0.28)

Results

	Unaided	Aided	Absolute difference
residents ($N_{readers}$=5)	N = 300	N = 300	
SE_{PW}	71.73	82.0	10.27 [7.81, 12.72] (p< 0.001)
SPE_{PW}	86.27	87.6	1.33 [-1.28, 3.95] (p=0.23)
experts ($N_{readers}$=3)	N = 300	N = 300	
SE_{PW}	75.56	83.78	8.22 [-2.42, 18.87] (p=0.08)
SPE_{PW}	95.11	94.89	-0.22 [-1.18, 0.73] (p=0.42)

Results

	Unaided	Aided	Absolute difference
residents ($N_{readers}$=5)	N = 300	N = 300	
SE_{PW}	71.73	82.0	10.27 [7.81, 12.72] (p< 0.001)
SPE_{PW}	86.27	87.6	1.33 [-1.28, 3.95] (p=0.23)
experts ($N_{readers}$=3)	N = 300	N = 300	
SE_{PW}	75.56	83.78	8.22 [-2.42, 18.87] (p=0.08)
SPE_{PW}	95.11	94.89	-0.22 [-1.18, 0.73] (p=0.42)

Results

	Unaided	Aided	Absolute difference
residents ($N_{readers}$=5)	N = 300	N = 300	
SE_{PW}	71.73	82.0	10.27 [7.81, 12.72] (p< 0.001)
SPE_{PW}	86.27	87.6	1.33 [-1.28, 3.95] (p=0.23)
experts ($N_{readers}$=3)	N = 300	N = 300	
SE_{PW}	75.56	83.78	8.22 [-2.42, 18.87] (p=0.08)
SPE_{PW}	95.11	94.89	-0.22 [-1.18, 0.73] (p=0.42)

Results

	Unaided	Aided	Absolute difference
residents ($N_{readers}$=5)	N = 300	N = 300	
SE_{PW}	71.73	82.0	10.27 [7.81, 12.72] (p< 0.001)
SPE_{PW}	86.27	87.6	1.33 [-1.28, 3.95] (p=0.23)
experts ($N_{readers}$=3)	N = 300	N = 300	
SE_{PW}	75.56	83.78	8.22 [-2.42, 18.87] (p=0.08)
SPE_{PW}	95.11	94.89	-0.22 [-1.18, 0.73] (p=0.42)

Results

	Unaided	Aided	Absolute difference
hand/wrist	68.75	87.08	18.33 [10.74, 25.93] (p<0.001)
arm	83.33	89.58	6.25 [3.92, 8.58] (p<0.001)
shoulder	96.25	98.33	2.08 [-0.47 4.64] (p=0.09)
knee/leg	63.75	67.5	3.75 [0.99, 6.51] (p=0.01)
foot/ankle	53.75	70.83	17.8 [10.35, 23.82] (p<0.001)

Results

	Unaided	Aided	Absolute difference
hand/wrist	68.75	87.08	18.33 [10.74, 25.93] (p<0.001)
arm	83.33	89.58	6.25 [3.92, 8.58] (p<0.001)
shoulder	96.25	98.33	2.08 [-0.47 4.64] (p=0.09)
knee/leg	63.75	67.5	3.75 [0.99, 6.51] (p=0.01)
foot/ankle	53.75	70.83	17.8 [10.35, 23.82] (p<0.001)

Results

	Unaided	Aided	Absolute difference
hand/wrist	68.75	87.08	18.33 [10.74, 25.93] (p<0.001)
arm	83.33	89.58	6.25 [3.92, 8.58] (p<0.001)
shoulder	96.25	98.33	2.08 [-0.47 4.64] (p=0.09)
knee/leg	63.75	67.5	3.75 [0.99, 6.51] (p=0.01)
foot/ankle	53.75	70.83	17.8 [10.35, 23.82] (p<0.001)

Results

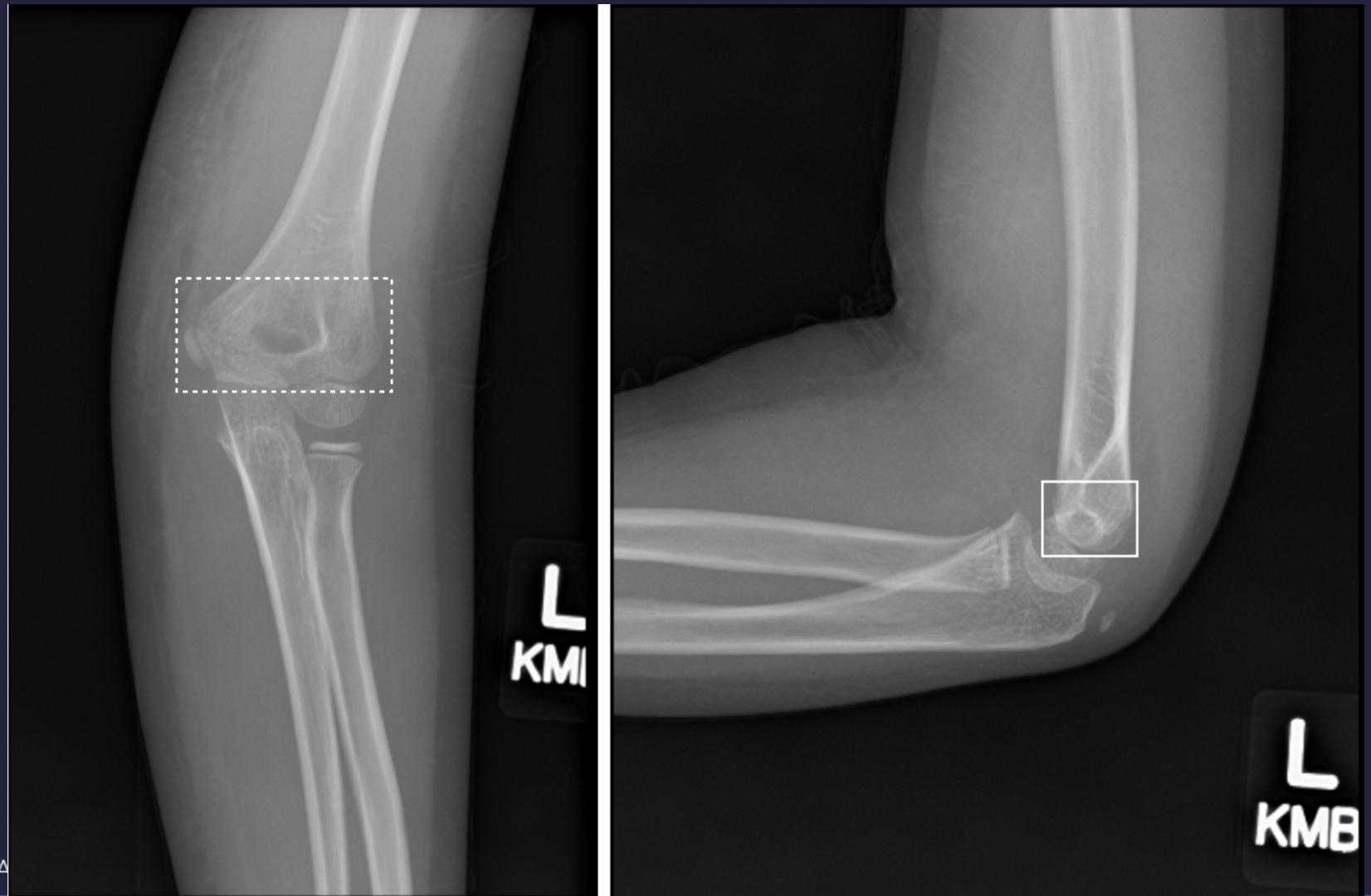
	Unaided	Aided	Absolute difference
Obvious	N_{fract} = 125	N_{fract} = 125	
SE_{FW}	83.8	90.9	7.1 [4.61, 9.59] (p<0.001)
Non-obvious	N_{fract} = 48	N_{fract} = 48	
SE_{FW}	52.34	66.67	14.32 [10.11, 18.53] (p<0.001)

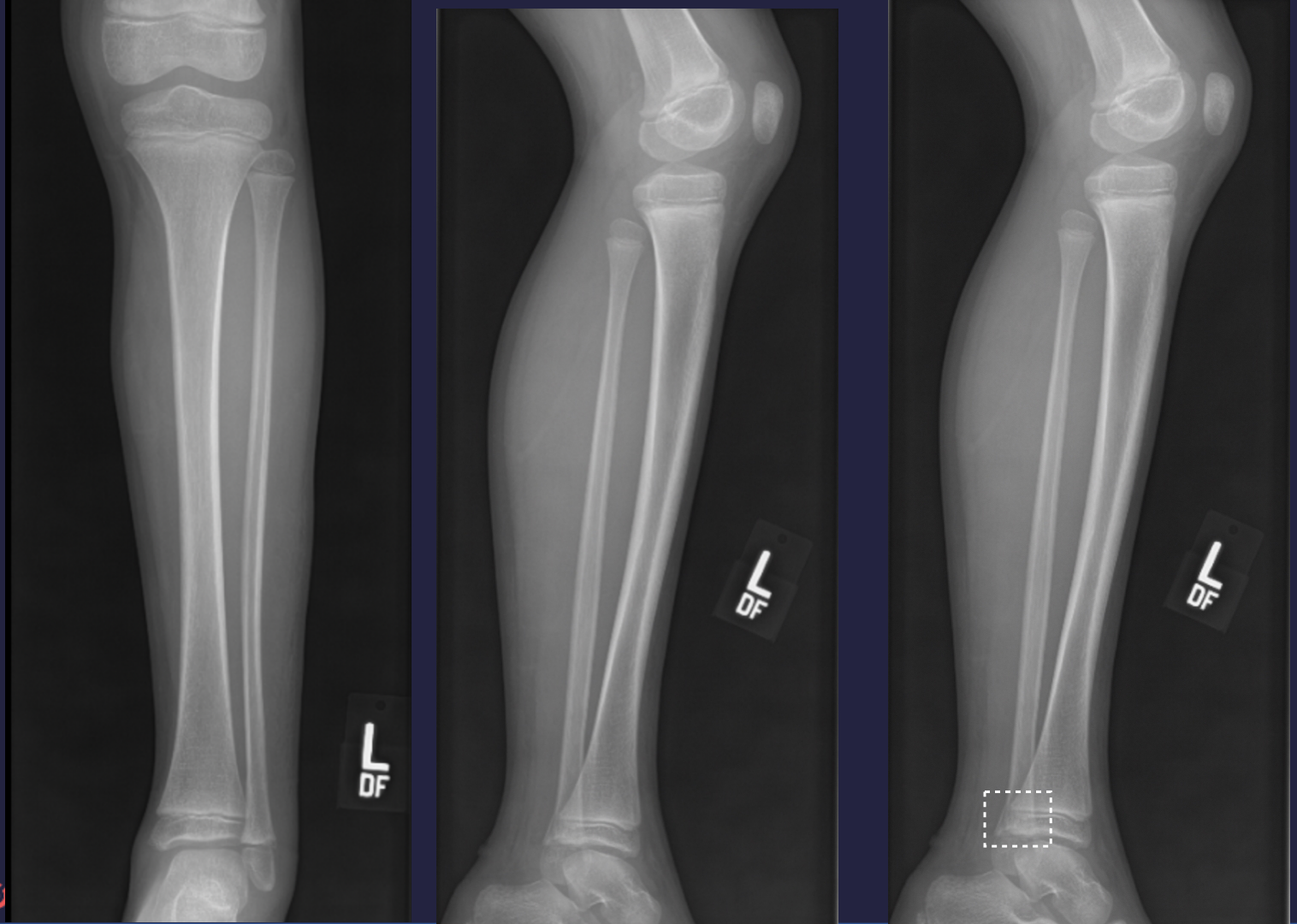
Discussion

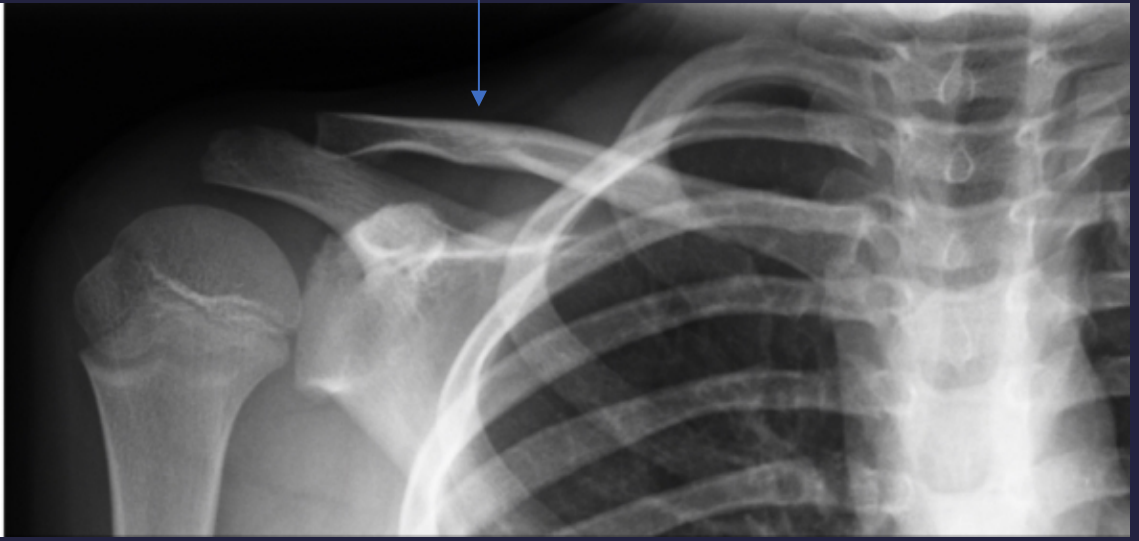
- Significant increase in sensitivity for all readers (73% to 83%, p < 0,001)
  - All anatomical locations
  - All types of factures, including pediatric-specific fractures
  - Greater in cases of non-obvious fractures
- No decrease in specificity

Discussion

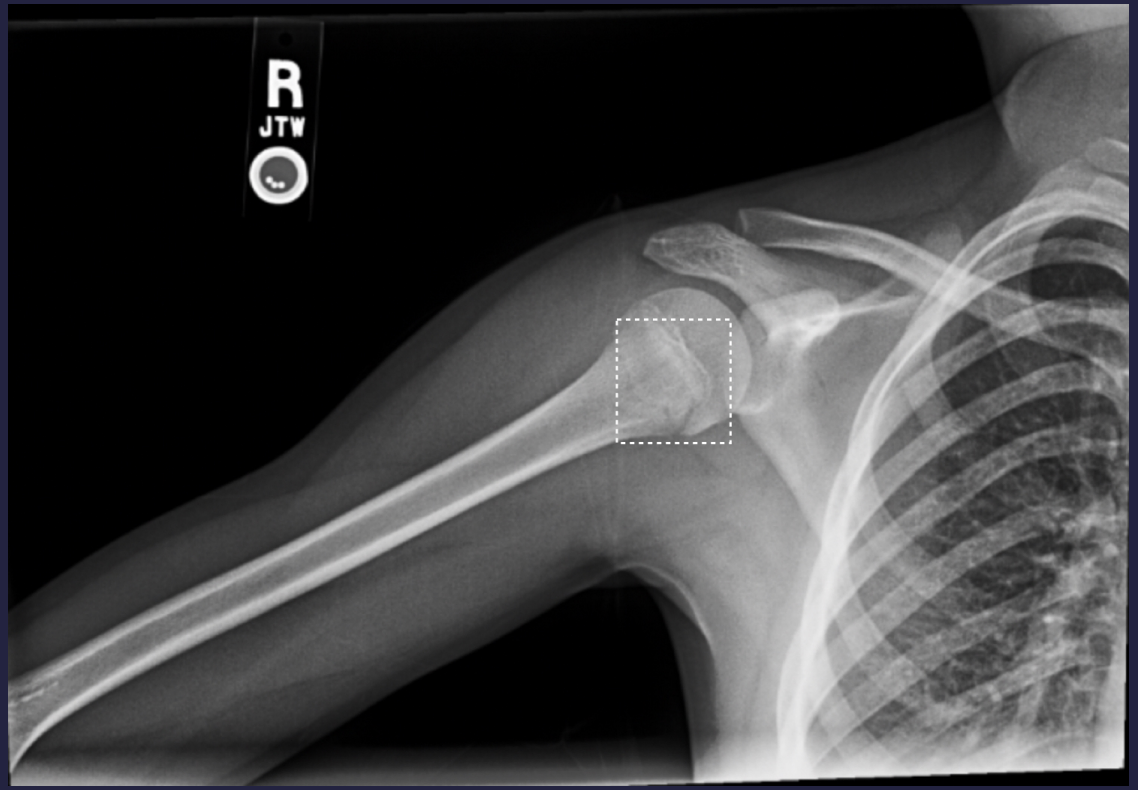
- First study about the peformance of the AI help for radiologists in pediatric population
- In pediatric population (England et al, AJR, 2018) : standalone study, similar results
- Duron et al, Radiology, 2021 : adult population, similar results. Improvement of sensitivity from 70.7% to 79.4%


L
KMI
L
KMB


L
DF
L
DF
L
DF




RT
JKJ


R
JTW


L
A0

Limitations

- Retrospective study
- Dataset artificially constituted
- Artificial reading conditions : cognitive bias
- No clinical informations : context bias

Conclusion

- Average increase in sensitivity of 9.5 %
- Greater increase among junior radiologists
- Greater increase in non-obvious fractures
- No decrease in specificity

- Needs confirmation by a larger scale study

ESPR
European Society of Paediatric Radiology

Assessment of an AI aid in detection of pediatric appendicular skeletal fractures by senior and junior radiologists

T. Nguyen, R. Maarek, A.-L. Hermann, H. Ducou Le Pointe
Trousseau Hospital, Paris

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

GLEAMER