

Automated Landmark Detection for Assessing Hip Conditions

A Cross-Modality Validation of MRI Versus X-Ray

Roberto Di Via, Vito Paolo Pastore, Francesca Odone, Siôn Glyn-Jones, Irina Voiculescu

MaLGa Center, University of Genoa | Nuffield Dept. of Orthopaedics, Oxford | Oxford Dept. of Computer Science



Clinical Motivation



20–25%

of the population affected by
Femoro-Acetabular Impingement (FAI)

#1 cause

of hip replacement in young adults

Hard

to diagnose at its early stages

Key Geometric Parameters

α -angle: quantifies femoral head asphericity
(cam morphology).

Pathological threshold: $\alpha > 65^\circ$

LCE angle: measures lateral acetabular
coverage (pincer morphology).

Pathological threshold: LCE $> 40^\circ$.

Imaging is essential for early detection and treatment planning



IEEE
ISBI 2026

Why MRI?



X-Ray Limitations

2D projection geometry

Tissue density projected into a single view plane

Up to 15° variability

Due to foot-to-hip alignment and pelvic tilt, sufficient to misclassify borderline cases

MRI Advantages

Volumetric 3D data

True 3D nature of the impingement area can be assessed

Alignment correction

Hip alignment can be corrected post-scan; angles measured in the correct plane

First step: validate that landmarks can be detected automatically in MRI slices

Contributions



1

Cross-Modality Validation

First validation of automated hip landmark detection between X-ray and MRI.
Equivalent performance on T1-weighted MRI (2.98 vs. 3.02 mm error, 87.5% diagnostic accuracy).

2

Clinical Workflow Integration

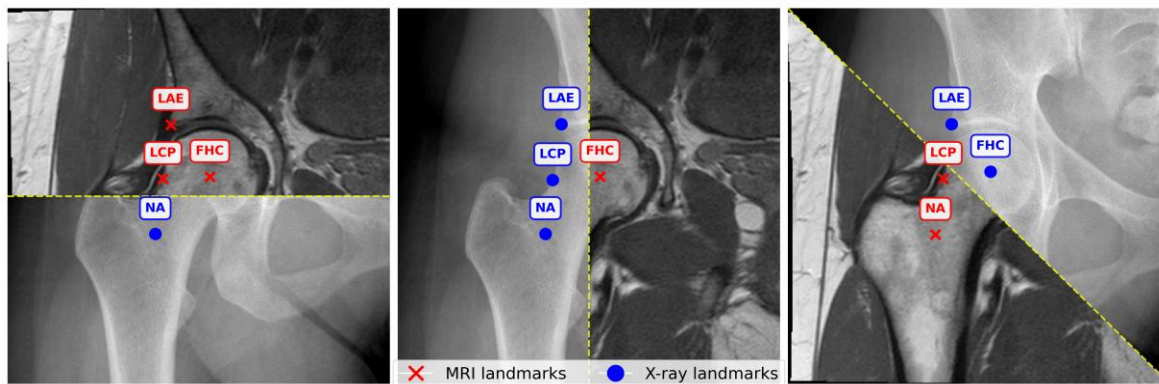
By extending measurements conventionally taken from X-rays to the corresponding slice in 3D MRI volumes, we enable seamless integration into clinical workflows.

3

Foundation for 3D Extensions

We establish a validated 2D foundation for future volumetric extensions towards full 3D geometric analysis and prediction.

Problem Formulation: Landmarks & Angles



Paired X-ray and T1-weighted MRI (slice 10) from the same subject, showing landmark correspondence across modalities.

The α -angle is notoriously difficult to calculate due to the LCP being hard to pinpoint, even by clinicians.

Four Anatomical Keypoints

FHC Femoral Head Centre

NA Neck-Axis Point

LAE Lateral Acetabular Edge

LCP Lateral Cam Point

α -angle: FHC→NA & FHC→LCP

LCE angle: Vertical & FHC→LAE

Method: Heatmap Regression



1

Ground-Truth Heatmaps

Gaussian heatmap for each landmark k , centred at annotated location (x_k, y_k) with $\sigma = 5$



2

Encoder-Decoder Network

UNet++ with ResNet18 encoder (ImageNet-initialised).
Trained to predict heatmaps by minimising NLL loss



3

Inference & Test-Time Aug.

Landmark coordinates extracted as spatial argmax.
TTA averages predictions across augmented views



IEEE
ISBI 2026



Dataset & Experimental Setup

89

patients

Paired AP X-ray + MRI

T1

coronal MRI

Annotation on slice 10

512×512

resolution

Min-max normalised

FAIT Trial Cohort

Collected at Oxford University. Pathological cohort with clinical and imaging evidence of FAI but without significant osteoarthritis or hip dysplasia. Data acquired using the same protocol, across multiple UK sites, with some patients contributing multiple time points.

Data Split: 65 : 10 : 25 (Train : Val : Test) → 105 : 17 : 40 images (57 : 8 : 24 patients)

Balanced across α -angle distributions given by Kolmogorov–Smirnov testing.

Results: Landmark Localisation



Landmark	Mean RE (mm) X-ray	Mean RE (mm) MRI	Median RE (mm) X-ray	Median RE (mm) MRI
FHC	1.07±0.08	1.78±0.13	1.01±0.11	0.84±0.14
NA	2.35±1.02	1.25±0.30	1.61±0.17	1.13±0.36
LAE	3.18±0.64	3.17±0.30	2.81±0.74	2.78±0.09
LCP	5.50±1.01	5.72±0.76	1.76±0.66	2.55±0.64
Overall	3.02±0.10	2.98±0.23	1.49±0.16	1.62±0.29

SDR@2/3/4 mm (%) X-ray: 62.5 / 79.2 / 85.4 MRI: 58.3 / 73.5 / 83.1

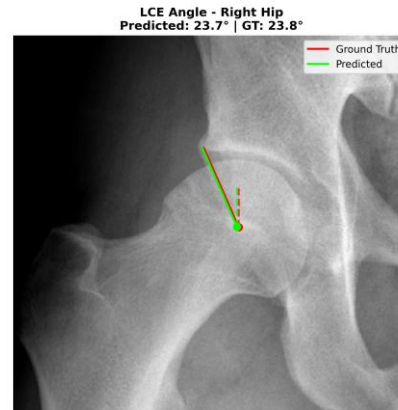
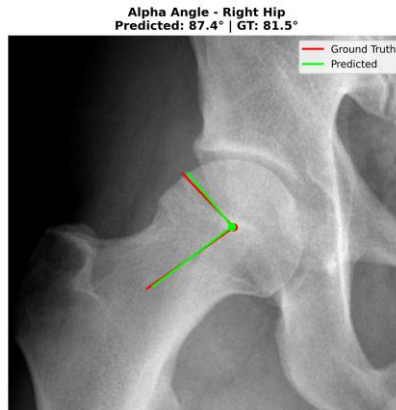
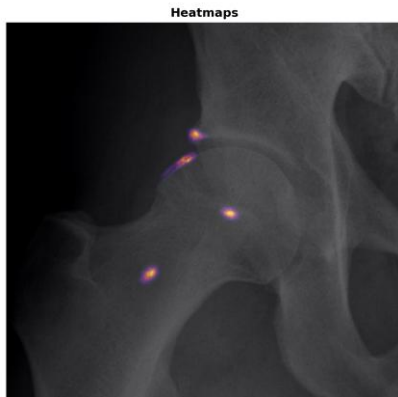
Key:

- MRI achieves 2.98±0.23 mm overall MRE, closely matching X-ray (3.02±0.10 mm).
- Complementary strengths: X-ray excels at FHC (bone-air contrast), MRI at NA (soft-tissue delineation).
- LCP hardest for both, reflecting inherent clinical variability.

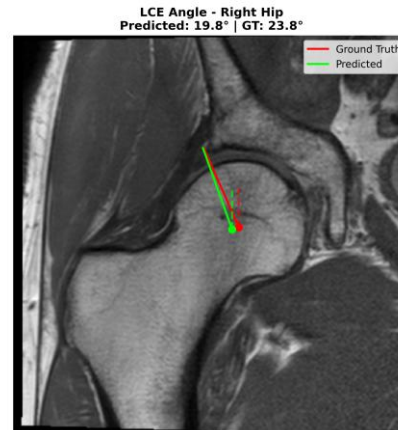
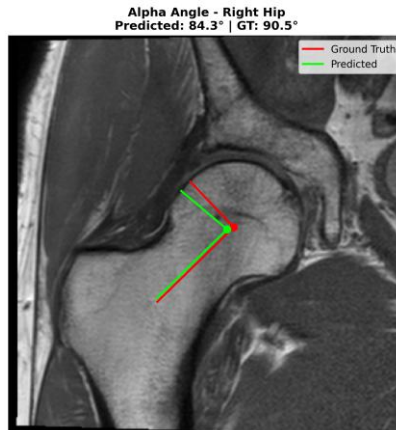
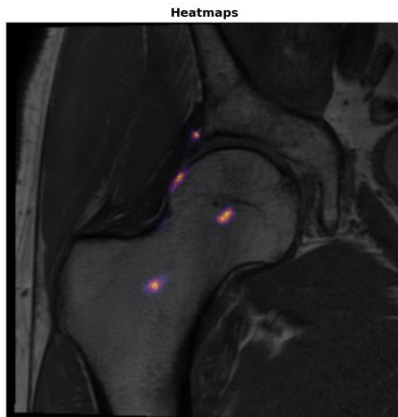
Qualitative Results



X-Ray



MRI



Left:
predicted heatmaps.

Centre: α -angle
(predicted in green,
ground truth in red).

Right: LCE angle.
(predicted in green,
ground truth in red).



IEEE
ISBI 2026



Angle Agreement & Diagnostic Accuracy

Metric	X-ray	MRI
LCE-angle agreement		
MAE (°)	2.97±1.14	2.96±0.52
Median (°)	2.22±0.42	1.28±0.20
ICC (2,1)	0.82±0.20	0.73±0.14
α -angle agreement		
MAE (°)	12.18±1.63	13.64±1.77
Median (°)	5.61±1.41	5.45±0.77
ICC (2,1)	0.52±0.06	0.41±0.15

Cam Screening ($\alpha > 65^\circ$)

87.5%

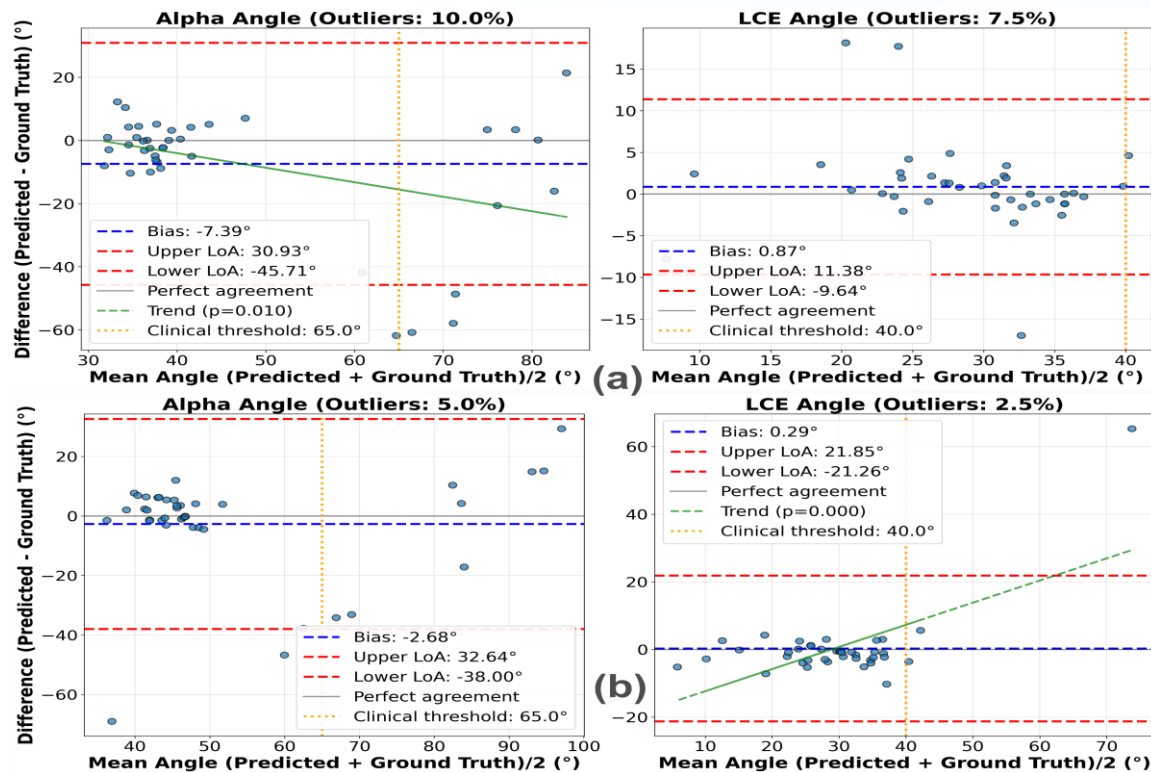
Diagnostic Accuracy
(identical for both modalities)

Sensitivity: 54.55% **Specificity:** 100%

PPV: 100% **NPV:** 85.29%

Key: LCE-angle shows excellent clinical agreement ($ICC \geq 0.73$). α -angle shows moderate agreement (ICC 0.41–0.52), comparable to inter-observer variability among clinicians (≈ 0.45 –0.56). Both modalities achieve the same diagnostic accuracy with perfect specificity.

Bland-Altman Agreement Analysis



Key Findings

LCE Angle

MRI: minimal bias (0.87°), narrow LoA ($\pm 10.5^\circ$), no proportional bias ($p=0.304$)

X-ray: wider LoA ($\pm 21.6^\circ$), proportional bias ($p<0.001$) at higher coverage

α Angle

Both underestimate clinician values (X-ray: -2.68° , MRI: -7.39°)

Broad LoA for both (± 35 – 38°)

MRI shows proportional bias ($p=0.010$), reflecting slice-selection sensitivity

Despite these biases, both achieve reliable cam-type classification.

Ablation Studies



Model	TTA	MRE (mm) ↓	SDR@2mm (%) ↑	Cam Acc. (%) ↑
UNet++ (ResNet18)	X	3.02±0.28	56.46±5.87	82.33±3.79
UNet++ (ResNet18)	✓	2.98±0.23	58.44±4.86	84.75±2.77
UNet++ (VGG16)	X	3.33±0.03	56.25±0.89	80.00±7.07
UNet++ (VGG16)	✓	3.29±0.36	57.09±1.93	80.17±4.41
UNet (MiT-B1)	✓	3.98±0.54	57.33±0.91	69.83±6.85
DPT (ResNet18)	✓	4.44±0.19	50.21±3.02	70.17±6.56
DPT (MaxViT-Base)	✓	3.63±0.99	55.94±1.33	77.50±10.61

Findings: UNet++ with ResNet18 achieves optimal performance. Lighter encoders outperform heavier alternatives, suggesting moderate-capacity networks with multi-scale skip connections are well-suited for this task. TTA consistently improves all architectures, providing 1.4–3.6 percentage point gains in cam accuracy while reducing variance.

Conclusion & Future Work



Main Conclusions

Automated landmark detection on MRI achieves accuracy equivalent to X-rays

MRE: 2.98 vs. 3.02 mm | Diagnostic accuracy: 87.5% both modalities

Geometric measures of FAI can be extracted reliably from routine MRI scans

MRI offers opportunities for refined development of screening angles, particularly for asymmetries or pelvic rotations

Current Limitations

- Single MRI slice analysis
- Moderate α -angle reliability (cam-point ambiguity)
- Modest cohort size

Future Work

- Volumetric metrics from 3D MRI
- Adjacent-slice training for data augmentation
- Uncertainty quantification

Code: <https://malga-vision.github.io/Landmarks-Hip-Conditions-AcademicPage/>

Thank You



Questions?

Correspondence to: irina@cs.ox.ac.uk

Code: github.com/Malga-Vision/Landmarks-Hip-Conditions