# Characterizing Hip Joint Morphology Using a Multitask Deep Learning Model

Bardia Khosravi, Lainey Bukowiec, John Patrick Mickley, Jacob Francis Oeding, Pouria Rouzrokh, Bradley James Erickson[1], Emmanouil Grigoriou, Michael J Taunton[1], Cody Wyles[1]
[1]Mayo Clinic

INTRODUCTION: Abnormal hip morphology is an important risk factor for hip pain, functional decline, and the development of osteoarthritis (OA). Plain radiographs remain the gold standard for initial workup; however, expert classification of morphological abnormalities seen in FAI and DDH has shown poor inter- and intra-rater reliability. Thus, this study aimed to evaluate the efficacy of deep learning (DL) models in classifying basic hip morphological abnormalities compared to human expert evaluation. A limited number of these technologies have been applied for computer vision analysis of radiographs displaying FAI and DDH. The authors hypothesized that DL-augmented image interpretation would show moderate performance given the difficulty in establishing a consensus human-driven ground truth in a dataset.

METHODS:

*Manual annotation*

Two trained research assistants annotated 400 AP radiographs, identifying: (a) ischial spine sign; (b) femoral head cam deformity; (c) hip dysplasia; and (d) any abnormalities including the previous three categories in combination with others. Ground truth for a hold-out test set of 100 images was determined by a fellowship trained hip surgeon surgeon.

*Joint localization*

A previously validated object detection model was utilized to localize regions of interest (ROIs) within AP pelvic radiographs. Each image was annotated with bounding boxes around the joint area, considering critical anatomical landmarks. Joint ROIs that contained any hardware were removed from the dataset.

*Morphology characterization*

A convolution-based deep learning model, ConvNeXt-Tiny was used to analyze hip joint morphology. This model predicts four specific joint characteristics (ischial spine sign, dysplasia, cam deformity, other abnormalities) and patient sex. Incorporating the auxiliary prediction of patient sex served to stabilize the model's training.

RESULTS:

Radiographs from 500 patients (mean age of 47 years, 49% female) were analyzed. The ensemble pipeline processes each radiograph in 832 ms.

Performance metrics are shown in table 1. The model achieved high accuracy and AUROC scores. The model demonstrated high specificity for ischial spine sign (96.0%) and all abnormalities (85.3%). Sensitivity was high for cam deformity (80.0%) and dysplasia (75.0%). The positive predictive value (PPV) was high for all abnormalities (93.5%) and ischial spine sign (87.5%).

Gwet's AC1, indicating inter-rater agreement, was substantial for dysplasia (0.83) and all abnormalities (0.88), and moderate for ischial spine sign (0.75) and cam deformity (0.61).

DISCUSSION AND CONCLUSION:

A novel approach utilizing DL techniques for the automated detection of morphological hip pathologies was developed. The results demonstrate the efficacy of a ConvNeXt-Tiny model trained on a dataset of AP pelvic radiographs, achieving good, but nowhere near perfect, accuracy in predicting various hip joint characteristics. Human expert inter-rater reliability was also good, but not perfect. Taken together, these results show the promise and current limitations of both using plain radiographs for simple morphological hip classifications, and the downstream impact this has on developing reliable DL models.

The current study’s model contributes to the growing body of evidence supporting the application of DL in musculoskeletal radiographic analysis to the diagnosis of FAI and DDH. The model examined solely the presence of two signs of FAI (ischial spine sign, cam deformity) or dysplasia, rather than conducting a complex workup involving measurement of continuous metrics such as Tonnis or center-edge angles. Clinicians often use MRI and CT scans to evaluate morphological hip abnormalities; however, the current paper deliberately examined the most basic form of hip pain workup – radiographs without accounting for angles – to evaluate whether DL could effectively classify simple measures with performance comparable to human interpretation.

The current study highlights the potential and inherent limitations of DL methods in augmenting diagnostic capabilities for hip pathologies, particularly FAI and DDH, through automated analysis of morphology present on pelvic radiographs. The findings suggest a promising avenue for leveraging AI-driven technologies to enhance musculoskeletal radiographic interpretation and improve patient care, but also introduce a strong note of caution in training models based on metrics upon which even experts have poor consensus, which will ultimately influence model capability.

*Figure 1. ROC curves for the DL movel in detecting hip joint abnormalities*

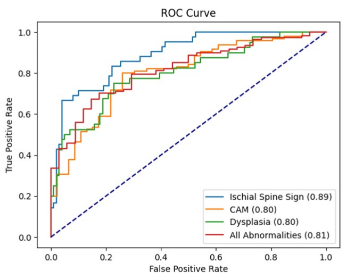


*Figure 2. Grad-CAM visualizations highlighting the regions of interest that the DL model focuses. The visualizations provide insights into the model's decision-making process and ability to identify relevant anatomical landmarks.*

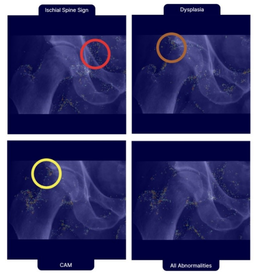


*Table 1. Performance metrics of the deep learning model for detecting hip joint abnormalities.*

| Abnormality | Accuracy | Sensitivity | Specificity | PPV | NPV | F1-score | AUROC |
|---|---|---|---|---|---|---|---|
| **Ischial Spine Sign** | 87.2% | 66.7% | 96.0% | 87.5% | 87.2% | 0.757 | 0.89 |
| **CAM** | 78.0% | 80.0% | 73.9% | 86.4% | 64.2% | 0.831 | 0.80 |
| **Dysplasia** | 76.6% | 75.0% | 77.2% | 56.6% | 88.6% | 0.645 | 0.80 |
| **All Abnormalities** | 71.6% | 67.3% | 85.3% | 93.5% | 45.3% | 0.783 | 0.81 |