

A Study on the Application of Multimodal MRI Deep Learning in the Early Detection of Osteonecrosis of the Femoral Head

Wang Xingru

Inner Mongolia Medical University

Liu Kai

Inner Mongolia Medical University

Zhang Liwei

The Second Affiliated Hospital of Inner Mongolia Medical University

Ma Chao

The Second Affiliated Hospital of Inner Mongolia Medical University

Meng Lingting

Inner Mongolia Medical University

Liu Jianming

Inner Mongolia University of Technology

Xiao Hongxin

Inner Mongolia University of Technology

Liu Guangwen

Inner Mongolia University of Technology

Cao Yuju

Zhengzhou Traditional Chinese Medicine Orthopaedic Hospital Affiliated to Henan University of Chinese Medicine

Wang Jianzhong

20120328@immu.edu.cn

The Second Affiliated Hospital of Inner Mongolia Medical University

Research Article

Keywords: Osteonecrosis of the femoral head, deep learning, YOLO model

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11043385/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

A Study on the Application of Multimodal MRI Deep Learning in the Early Detection of Osteonecrosis of the Femoral Head

Wang Xingru^{1#}, Liu Kai^{1#}, Zhang Liwei^{2#}, Ma Chao², Meng Lingting¹, Liu Jianming³, Xiao Hongxin³, Liu Guangwen^{3*}, Cao Yuju^{4*}, Wang Jianzhong^{2*}

¹ Inner Mongolia Medical University, Hohhot, Inner Mongolia, 010059, China

² Department of Orthopaedics, The Second Affiliated Hospital of Inner Mongolia Medical University, Hohhot, Inner Mongolia, 010000, China

³ College of Intelligent Science and Technology, Inner Mongolia University of Technology, Hohhot 010080, China

⁴ Zhengzhou Traditional Chinese Medicine Orthopaedic Hospital Affiliated to Henan University of Chinese Medicine, Zhengzhou, Henan, 450000, China

Corresponding authors:

Wang Jianzhong (Lead corresponding author)

Department of Orthopedics, The Second Affiliated Hospital of Inner Mongolia Medical University, Hohhot, Inner Mongolia, 010000, China

Email: 20120328@immu.edu.cn

Liu Guangwen

College of Intelligent Science and Technology, Inner Mongolia University of Technology, Hohhot 010080, China

Email: liugw@imut.edu.cn

Cao Yuju

Zhengzhou Traditional Chinese Medicine Orthopedic Hospital Affiliated to Henan University of Chinese Medicine, Zhengzhou, Henan, 450000, China

Email: caoyuju@126.com

Abstract:

Objective: Osteonecrosis of the femoral head (ONFH) is a debilitating disorder that frequently leads to hip joint dysfunction. Early and accurate diagnosis is crucial to improving clinical outcomes. This study aimed to develop a magnetic resonance imaging (MRI)-based detection model using the You Only Look Once (YOLO) deep learning framework for automatic identification of early-stage ONFH, and to systematically evaluate how different MRI modalities affect detection performance and their potential for clinical decision support.

Methods: We retrospectively collected MRI data from 253 ONFH patients across two medical centers, including T1-weighted Imaging (T1WI) and T2-weighted Imaging (T2WI) sequences. After standardized preprocessing, the dataset was partitioned into training,

validation, and test sets (7:1:2). Two sequential experiments were conducted. Experiment 1 compared six YOLO generations (v3, v5, v8, v10, v11, v13) using single-center T1WI data, with precision (P), recall (R), mean average precision (mAP), and inference speed as primary metrics. Experiment 2 applied the optimal model to single-modality (T1WI or T2WI) and multimodal fused images from both centers, using the same metrics to quantify detection efficacy and validate multicenter robustness.

Results: In Experiment 1, YOLO v11 outperformed all other models, achieving a mAP@0.5 of 0.966, a recall of 0.941, and a per-frame inference time of 4.9 ms. In Experiment 2, the multimodal fusion input (T1WI+T2WI) yielded the highest performance across all modalities, with a mAP@0.5 of 0.9637, a recall of 0.9388, and the fastest inference time (3.0 ms), surpassing both T1WI (R=0.9159, 4.9 ms) and T2WI (R=0.9113, 6.0 ms) alone.

Conclusion: The proposed YOLO v11 model demonstrates excellent performance in MRI-based ONFH detection, and the multimodal fusion strategy further improves both diagnostic accuracy and computational efficiency. This approach addresses the dual clinical requirements for high precision and real-time processing, offering a practical and objective intelligent tool for early ONFH screening.

Novelty Statement: This study systematically benchmarks six YOLO generations for MRI-based ONFH detection and demonstrates that early T1WI/T2WI channel fusion improves both detection accuracy and computational efficiency. These findings provide a practical reference for selecting real-time deep learning models in clinical ONFH screening.

Keywords: Osteonecrosis of the femoral head; deep learning; YOLO model

Introduction

Osteonecrosis of the femoral head (ONFH) is a prevalent and challenging orthopaedic disorder that predominantly affects young and middle-aged adults, with a rising global incidence^[1-3]. Global epidemiological data indicate an annual incidence of 10 to 20 per 100,000 person-years, with the upward trend particularly pronounced in Asia^[4]. Early diagnosis is paramount for joint-salvage procedures, as delayed recognition often leads to femoral head collapse and the eventual need for total hip arthroplasty^[5-8]. Magnetic resonance imaging (MRI), with its superior sensitivity to bone marrow edema and early ischemic changes, has become the gold standard for early ONFH detection^[9-11]. However, conventional MRI interpretation remains subjective, time-consuming, and highly dependent on the radiologist's experience, resulting in considerable interobserver variability and a notable risk of missing subtle lesions^[3, 12]. These limitations underscore an urgent clinical need for automated, objective, and efficient diagnostic support tools.

Recent advances in deep learning-based object detection have opened new avenues for medical image analysis. Single-stage detectors, particularly the YOLO (You Only Look Once) series, have gained traction due to their end-to-end architecture and real-time inference capabilities, demonstrating strong performance in diverse tasks including orthopaedic abnormality detection, thyroid nodule identification, and pulmonary nodule screening^[13-16]. For instance, YOLO v10 has been applied to detect early-stage ONFH on radiographs, achieving area under the curve (AUC) values of 0.93–0.94^[17]; YOLO v8 has also been used to quantify empty lacunae in histological sections of ONFH^[18]. Despite these promising applications, no study has systematically benchmarked different YOLO versions for ONFH detection on MRI, nor has the potential benefit of multimodal MRI fusion been explored in this context. In clinical practice, ONFH evaluation typically requires combined assessment of T1-weighted Imaging (T1WI) and T2-weighted Imaging (T2WI) images, which provide complementary information—T1WI delineates anatomic structure and necrotic margins, while T2WI is highly sensitive to edema and inflammatory changes^[2, 19]. Multimodal fusion has proven advantageous in other oncological imaging settings^[20-22], but its utility in ONFH detection remains unknown.

Given the complementary nature of multimodal MRI and the powerful feature extraction capabilities of deep learning, integrating these technologies holds promise for enhancing diagnostic accuracy and efficiency. We therefore hypothesized that a YOLO-based model, when trained on fused T1WI and T2WI inputs, would achieve superior performance over

single-modality approaches for early ONFH detection. To test this, we conducted a systematic comparison of six YOLO generations and rigorously evaluated the impact of input modality on detection outcomes.

1 Materials and Methods

This retrospective study received ethical approval from the institutional review boards of Center 1 (The Second Affiliated Hospital of Inner Mongolia Medical University, No. EFY20260014) and Center 2 (Zhengzhou Hospital of Traditional Chinese Medicine for Orthopaedics and Traumatology, No. IECPJ2024120801). Informed consent was waived because all personally identifiable information was removed from the MRI data prior to analysis.

1.1 Study Population

Participants were recruited from two centers. Center 1 contributed patients who underwent hip MRI between January 1, 2015, and October 1, 2024; Center 2 provided data from patients examined between November 1, 2018, and June 1, 2020. All MRI studies and corresponding radiology reports were independently reviewed by experienced musculoskeletal radiologists to confirm data accuracy.

1.2 Inclusion and Exclusion Criteria

1.2.1 Inclusion Criteria

- (1) MRI-based diagnosis of ONFH.
- (2) Intact femoral head contour on MRI (ARCO stage $\leq$ IIIA, i.e., pre-collapse).
- (3) Radiographic evidence of skeletal maturity.
- (4) Absence of other hip joint pathologies (e.g., femoroacetabular impingement, dysplasia) on imaging.
- (5) Complete imaging data.

1.2.2 Exclusion Criteria

- (1) Altered femoral head contour (ARCO stage $\geq$ IIIB).
- (2) Osteoarthritis (ARCO stage IV).
- (3) Prior hip surgery.
- (4) Poor image quality (severe artifacts, low signal-to-noise ratio) or missing sequences.

The participant selection workflow is illustrated in Figure 1.

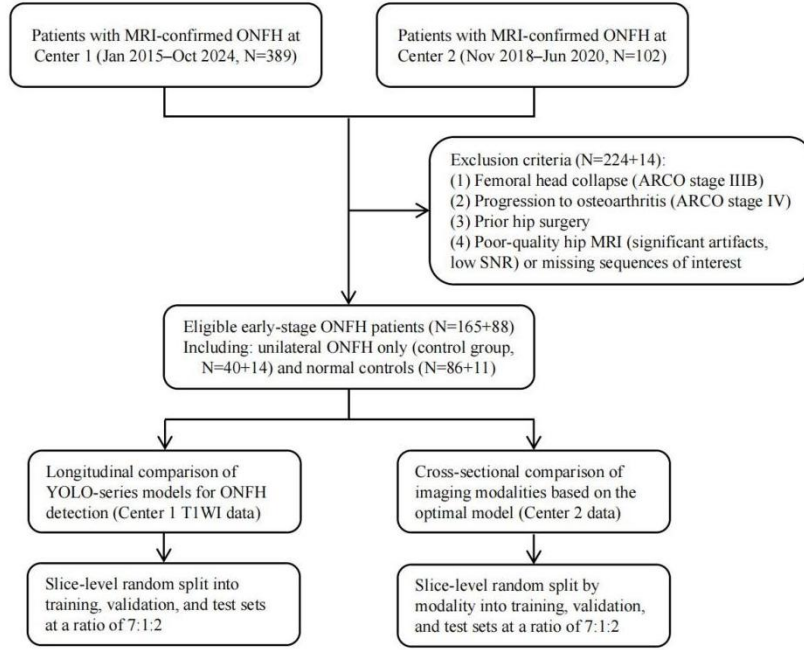


Figure 1 Flowchart of the selection and analysis process for this study

1.3 MRI Acquisition

Scanners varied across centers: Center 1 used a 3.0T Siemens Skyra, a 1.5T Philips Ingenia, and a 1.5T GE MR scanner; Center 2 used a 3.0T Siemens Skyra. Both T1WI and T2WI sequences were acquired for each participant, and all images were stored in DICOM format.

1.4 Annotation and Preprocessing

The femoral head was defined as the region of interest. Three radiologists manually annotated T1WI and T2WI images using Labelme software. After removing identifiers, images were converted to PNG format with a standardized naming scheme.

Two annotation classes were defined: Normal (regular morphology, uniform signal) and Abnormal (ONFH lesions, featuring geographic hypointense bands, bone marrow edema, or cystic changes consistent with ARCO stages I–II and pre-collapse IIIA).

Annotation proceeded slice-by-slice on T1WI, saved as JSON files; the same lesions were then manually segmented on corresponding T2WI slices. All annotations were reviewed and validated by two senior orthopaedic specialists (>5 years of experience in musculoskeletal imaging), with discrepancies resolved by consensus.

The validated JSON files were converted to YOLO-format TXT files, with class mapping: 0 = normal, 1 = abnormal. Using Python (labelme, cv2), polygon vertex coordinates were

extracted to compute the minimum bounding box (top-left x_1, y_1 ; bottom-right x_2, y_2). Normalized center coordinates (x_center , y_center) and dimensions (w , h) were calculated as:

$$x_{center} = \frac{\frac{x_1 + x_2}{2}}{W_{img}} \quad (1)$$

$$y_{center} = \frac{\frac{y_1 + y_2}{2}}{H_{img}} \quad (2)$$

$$w_{norm} = \frac{x_2 - x_1}{W_{img}} \quad (3)$$

$$h_{norm} = \frac{y_2 - y_1}{H_{img}} \quad (4)$$

where W_{img} and H_{img} are image width and height. The dataset was randomly split into training, validation, and test sets (7:1:2), preserving class proportions. A final verification step confirmed image-annotation matching, parameter validity, and class-ID consistency; ambiguous samples were excluded.

1.5 YOLO Model Selection Rationale

The YOLO family has evolved considerably since its inception. To provide a comprehensive benchmark, we selected six representative models (v3, v5, v8, v10, v11, v13) that span major architectural innovations. Table 1 summarizes their core features. Notably, YOLO v11 incorporates a lightweight CSPDarknet-27 backbone, C3K2 modules with dynamic convolution, and a C2PSA position-sensitive attention mechanism, which are particularly suited for detecting small lesions on edge devices—a key consideration for clinical deployment. The inclusion of v13, despite its higher computational cost, allowed us to assess the upper bound of localization precision (mAP@.5:.95).

Table 1. Core Architectural Features of the YOLO Models Evaluated in This Study

Model	Core Features	Architectural	Core Design Philosophy	Primary Advantages	Applicable Scenarios / Characteristics
YOLO v3	Darknet-53 backbone; multi-scale detection; residual connections for enhanced feature extraction.	backbone; detection; connections for feature	Single-stage detection framework; detection accuracy across varying object sizes via multi-scale fusion while	Excellent balance of speed and accuracy; significantly improved small-object detection over predecessors;	Suitable for general-purpose real-time object detection.

		maintaining real-time performance.	simple and deployable architecture.
YOLO v5	One-stage detection framework with an improved Backbone-Neck-Head architecture; supports end-to-end training.	Emphasizes engineering robustness and ease of use; balances lightweight design with high performance while improving training and deployment efficiency.	Mature training pipeline; fast inference; extensive ecosystem; supports dynamic anchor adjustment and adaptive image scaling; easy to transfer and deploy. Suitable for rapid construction of baseline detection models; applicable to scenarios requiring high real-time performance and stability.
YOLO v8	Improved CSPDarknet-53 backbone; PAN-FPN neck; anchor-free head; supports multi-tasking (detection, segmentation, pose estimation).	Pursues higher detection accuracy and speed; simplifies detection pipeline via anchor-free design while enhancing multi-task capabilities.	Outperforms YOLOv5 in both accuracy and speed; unified framework for multiple tasks; mature training and deployment ecosystem; compatible with diverse hardware platforms. Suitable for scenarios requiring high-precision real-time detection.
YOLO v10	Further optimized feature extraction and multi-scale fusion over YOLOv8; improved residual modules in backbone; enhanced feature fusion in neck; dynamic label assignment.	Focuses on feature representation stability and model convergence; improves localization robustness through architectural and training strategy optimization.	Maintains lightweight characteristics while improving localization accuracy and detection robustness; stable training process; superior performance in complex scenes. Suitable for tasks requiring a balance between model size and detection accuracy.
YOLO v11	Lightweight backbone (CSPDarknet-27); improved neck feature fusion; adaptive feature enhancement modules; supports dynamic	Focuses on extreme lightweighting and edge-device deployment; balances accuracy and speed through dynamic inference and feature	Smaller model size; faster inference; significantly improved deployment efficiency on edge devices while maintaining high
			Optimally suited for medical imaging scenarios (e.g., lesion detection on radiographs, fundus lesion recognition)

	inference mechanism.	enhancement.	detection accuracy.	and edge computing environments with limited computational resources.
			Potential advantages in	
		Prioritizes localization	high-precision	Applicable to
	Further enhanced feature	accuracy under stringent	localization; superior	applications with
YOLO	representation and	IoU conditions through	detection performance	stringent
v13	regression quality over	improved feature	for small and dense	requirements for
	previous versions.	expression and	objects; suitable for	object boundary
		regression capabilities.	fine-grained detection	localization.
			tasks.	

1.6 Experimental Design

1.6.1 Experiment 1: Benchmarking YOLO Generations

Dataset: Center 1 T1WI data only. After screening, 251 participants (165 ONFH, 86 normal) yielded 1,889 images. Stratified split: training (1,322), validation (188), test (379) slices.

Procedure: Six YOLO models were trained with identical hyperparameters and preprocessing on the same dataset, and evaluated on the common test set.

1.6.2 Experiment 2: Modality Comparison Using Optimal Model

Dataset: Both centers, T1WI and T2WI. After screening, 350 participants (253 ONFH, 97 normal) and 5,754 images (Center 1: 251; Center 2: 99) were included. Stratified split into training/validation/test (7:1:2). Per-modality counts are shown in Table 2.

Fusion Strategy: We performed early (input-level) fusion. T1WI and T2WI images were resized to 512×512 pixels, then combined via RGB channel stacking (T1WI and T2WI as separate channels) to produce a single color composite. YOLO labels were scaled accordingly. Single-modality inputs used grayscale single-channel images.

Procedure: Using the optimal model from Experiment 1, three separate runs were conducted: (1) T1WI alone, (2) T2WI alone, (3) fused T1WI+T2WI. All runs used identical training settings and evaluation metrics.

Table 2: Dataset Division

Modality	Training Set	Validation Set	Test Set
Modality 1 (T1WI)	2104	300	604
Modality 2 (T2WI)	1924	273	549
Multimodal Fusion	1838	262	527

The three datasets are consistent in terms of image types, lesion severity, and other aspects, effectively preventing data bias from affecting the experimental results.

It should be noted that the three modality datasets originated from the same pool of 5,754 images; the slight variation in sample sizes across modalities reflects the availability of paired T1WI and T2WI sequences for each image.

1.7 Implementation and Training

All experiments ran on a single machine with NVIDIA RTX 2060 GPU (6GB VRAM, CUDA 12.8), Windows OS, PyTorch 2.7.0, and TorchVision 0.22.0. Data processing used Python 3.9.15 and OpenCV 4.9.0. The Ultralytics YOLO framework handled training and evaluation; visualization used Matplotlib and Seaborn; statistical analysis used NumPy and Pandas.

To ensure reproducibility, all models were trained with uniform hyperparameters: 200 epochs, batch size 8, AdamW optimizer (momentum 0.937, weight decay 5e-4), initial learning rate 1e-3, reduced to 1e-4 during fine-tuning. All evaluations used the same test set, and core metrics were automatically computed by the Ultralytics framework and verified via Python scripts.

1.8 Evaluation Metrics

We assessed models across four dimensions: detection accuracy, inference speed, localization precision, and classification correctness.

1.8.1 Accuracy Metrics

From the confusion matrix, we derived:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Precision-Recall (PR) curve and Average Precision (AP) (area under PR curve) were calculated using the standard VOC2012 interpolation. Mean Average Precision (mAP) was reported at two IoU thresholds:

$$mAP@0.5 = AP_{IoU=0.5} \quad (9)$$

$$mAP@.5:.95 = \frac{1}{10} \sum_{t=0.5}^{0.95, \Delta t=0.05} AP_{IoU=t} \quad (10)$$

1.8.2 Speed Metric

Per-frame total processing time (ms) was measured from image input to output, reflecting end-to-end inference speed.

2 Results

2.1 Performance of YOLO Generations (Experiment 1)

Table 3 presents the comparative results. YOLO v11 achieved the best trade-off, with a recall of 0.941, mAP@0.5 of 0.966, and the fastest inference (4.9 ms) among the high-performing models. YOLO v13 showed superior mAP@.5:.95 (0.954) but was 2.3× slower (11.3 ms), rendering it less practical for real-time use. Thus, YOLO v11 was selected for Experiment 2. Figure 2 visualizes these comparisons.

Table 3 Comparison of Experimental Results for Different YOLO Models on

Model	P	R	mAP@0.5	mAP@.5:.95	Total Processing Time per Frame (ms)
YOLO v3	0.911	0.919	0.956	0.825	77.3
YOLO v5	0.903	0.887	0.947	0.817	7.6
YOLO v8	0.914	0.911	0.959	0.843	5.6
YOLO v10	0.911	0.909	0.966	0.843	5.1
YOLO v11	0.899	0.941	0.966	0.845	4.9
YOLO v13	0.910	0.929	0.959	0.954	11.3

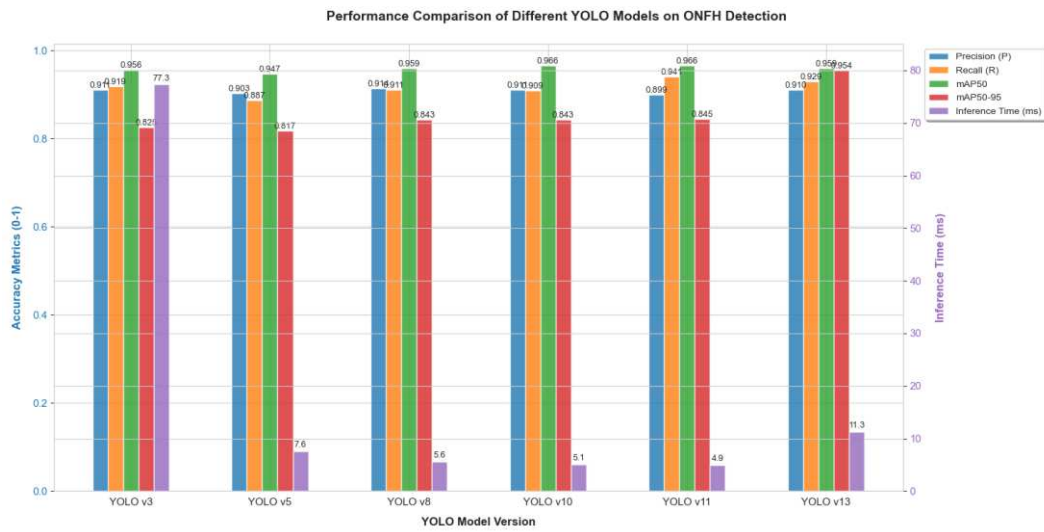


Figure 2 Performance Comparison of Different YOLO Models in the ONFH Detection Task

2.2 Detailed Evaluation of YOLO v11 (Figures 3–9)

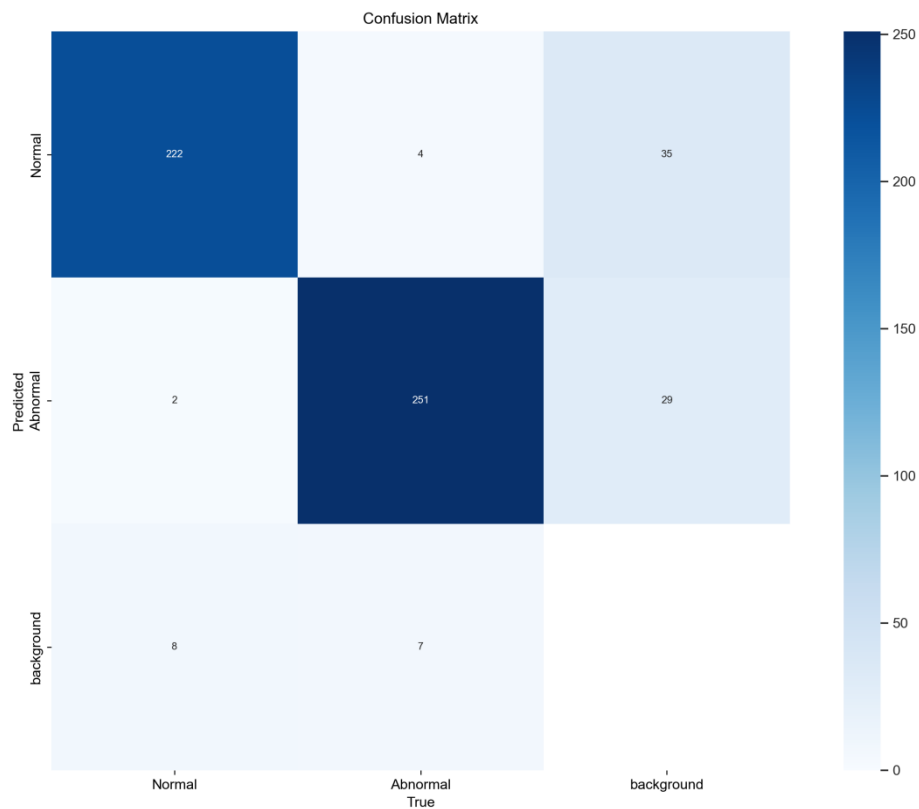


Figure 3 Confusion Matrix

Figure 3 shows the confusion matrices for the three-class object detection task, illustrating the match between the YOLO v11 model's predictions and the ground truth labels. The categories include Normal, Abnormal, and Background. The values in the figure represent the number of predicted and ground truth samples, respectively, while the color bars

on the right indicate the magnitude of these values; darker colors represent larger values.

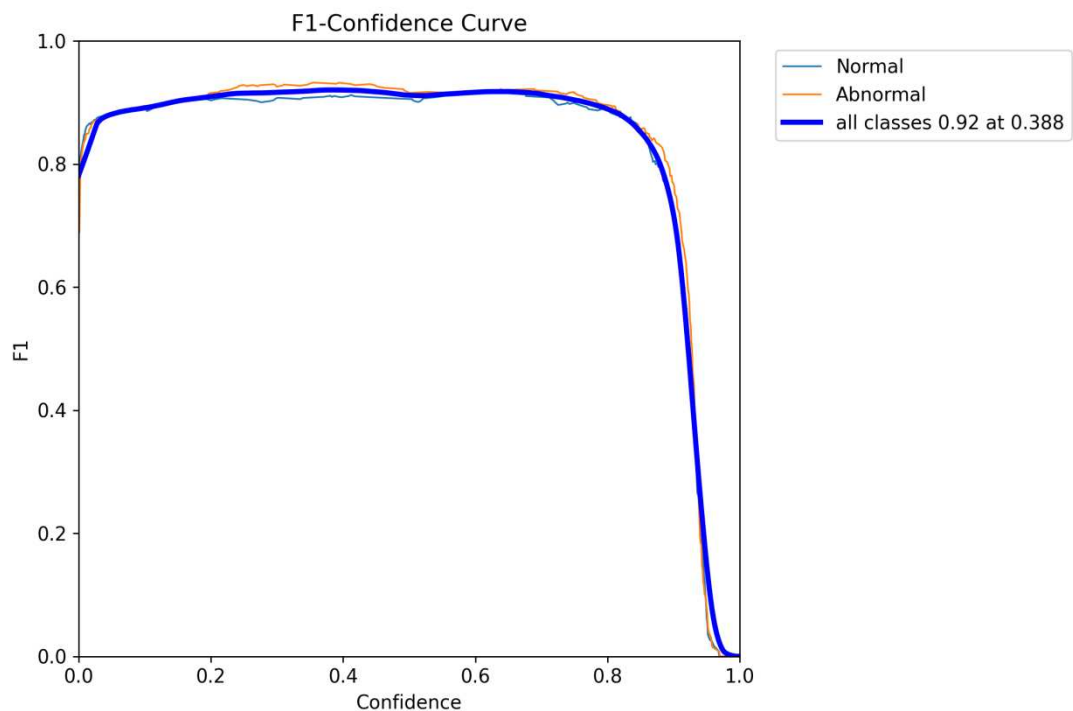


Figure 4 F1 Confidence Curve

Figure 4 shows the F1 confidence curve, which visually illustrates the overall classification performance of the YOLO v11 model at different confidence thresholds. The highest F1 score in the figure is 0.92, corresponding to a confidence threshold of 0.388. Within the 0–0.8 confidence range, the F1 score remains above 0.8; when confidence exceeds 0.8, the F1 score rapidly drops to 0.

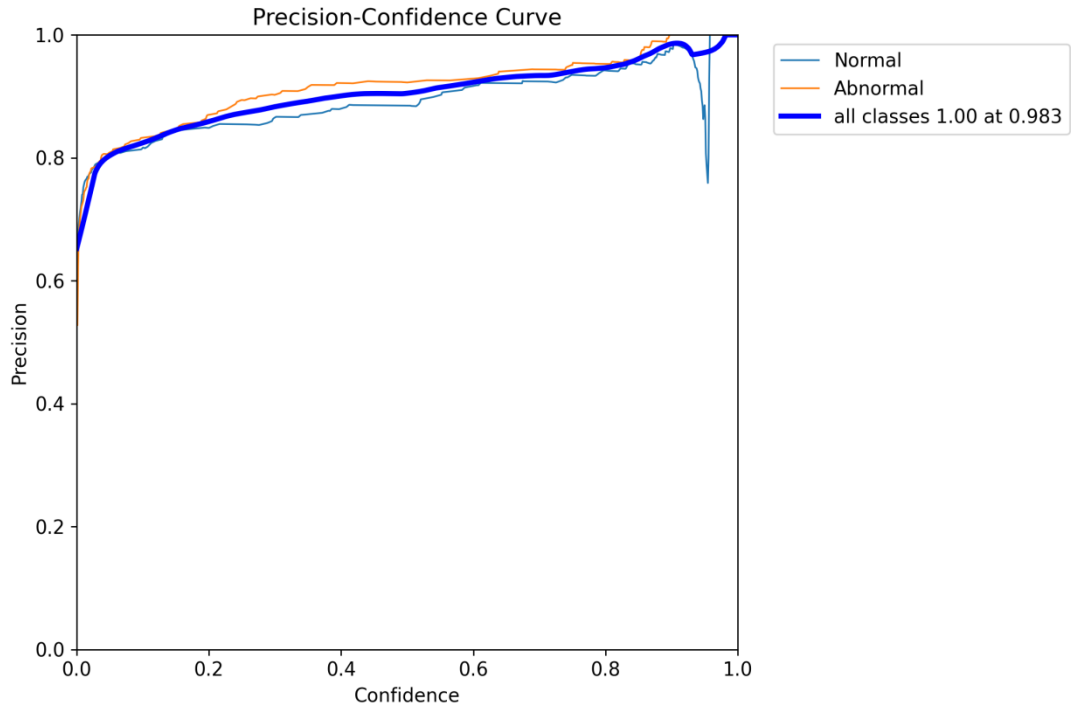


Figure 5: Precision Confidence Curve

Figure 5 shows the precision performance of the YOLO v11 model at different confidence thresholds. The highest value of the overall precision in the figure is 1.00, corresponding to a confidence threshold of 0.983. When confidence ranges from 0 to 0.8, precision gradually increases from approximately 0.8 to nearly 0.95; when confidence ranges from 0.8 to 1.0, precision rises further, reaching 1.00 at a confidence level of 0.983. However, as confidence approaches the threshold of 1.0, the curve for normal classes experiences a brief decline.

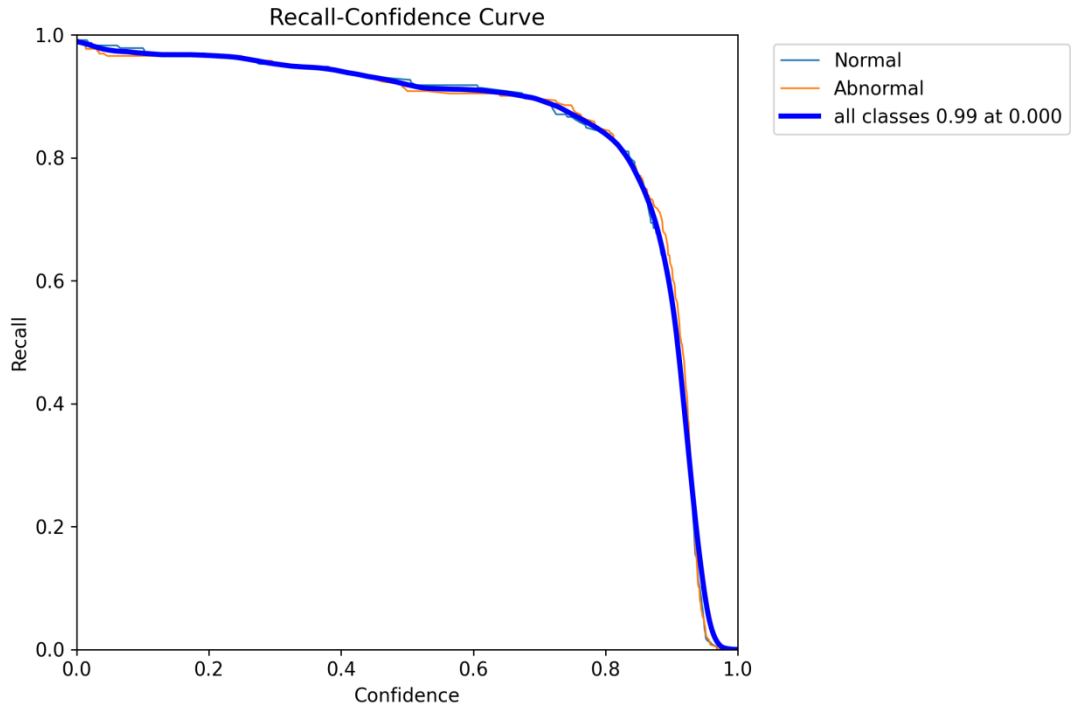


Figure 6: Recall Confidence Curve

Figure 6 shows the recall performance of the YOLO v11 model at different confidence thresholds. The labeled overall recall in the figure is 0.99, corresponding to a confidence threshold of 0.000. When confidence ranges from 0 to 0.8, recall remains above 0.85, and approaches 0.99 when confidence is 0; as the confidence threshold increases, recall gradually decreases; when confidence ranges from 0.8 to 1.0, recall drops sharply to 0.

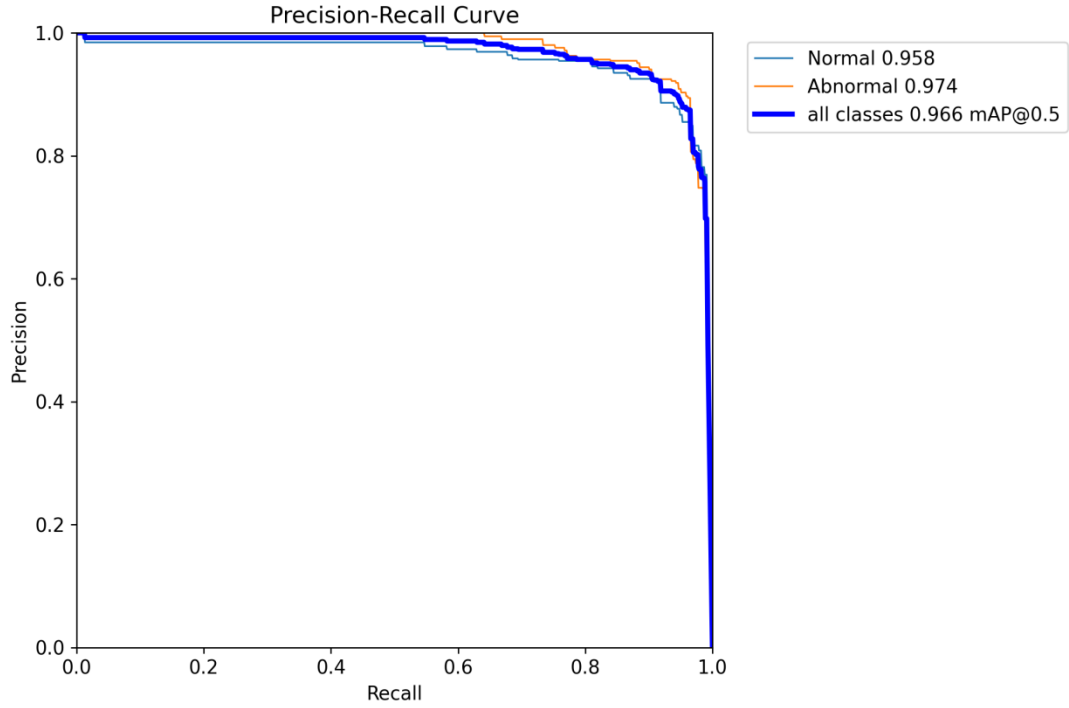


Figure 7: Precision-Recall Curve

Figure 7 Precision-Recall curve illustrating the precision performance of the YOLO v11 model across varying recall levels. The AP values for each class are annotated: 0.958 for the normal class and 0.974 for the abnormal class, with an overall mAP@0.5 of 0.966. Precision remained above 0.95 when recall ranged from 0 to 0.8, and dropped sharply to 0 as recall approached 1.0.

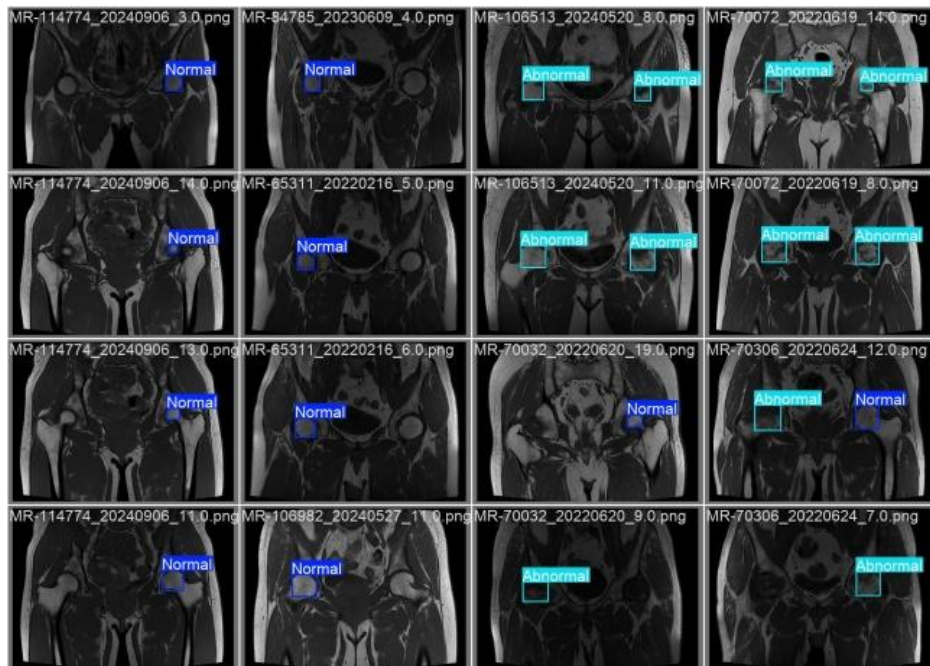


Figure 8: Schematic of original image labels

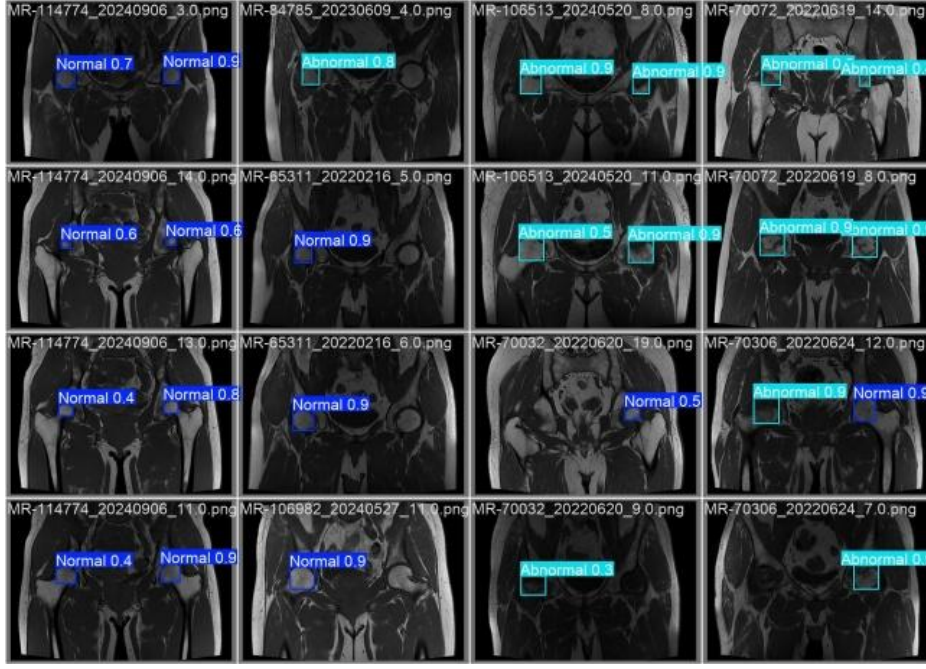


Figure 9: Schematic of predicted labels for original images

Figures 8 and 9 clearly illustrate the YOLO v11 model’s detection of ONFH lesions in hip MRI images. Normal samples are labeled as “Normal,” with confidence scores mostly ranging from 0.7 to 0.9; abnormal lesions are labeled as “Abnormal,” with confidence scores mostly at 0.9.

2.3 Modality Comparison (Experiment 2)

Table 4 shows that the multimodal fusion input consistently outperformed single modalities across all key metrics. Notably, fusion achieved the highest recall (0.9388), mAP@0.5 (0.9637), mAP@.5:.95 (0.8650), and the lowest per-frame time (3.0 ms)—even faster than T1WI alone (4.9 ms). This suggests that early fusion not only improves accuracy but also enhances computational efficiency, likely through better GPU utilization.

Table 4 Comparison of Basic Performance of the YOLO v11 Model Across Different Modalities

Modality	P	R	mAP@0.5	mAP@.5:.95	Total Time per Frame (ms)
Mode 1 (T1WI)	0.8927	0.9159	0.9570	0.8558	4.9
Mode 2 (T2WI)	0.8869	0.9113	0.9562	0.7910	6.0
Multimodal Fusion	0.8925	0.9388	0.9637	0.8650	3.0

3 Discussion

In this study, we systematically benchmarked six generations of YOLO models (v3

through v13) for MRI-based detection of ONFH and identified YOLO v11 as the optimal choice, offering the most favorable trade-off among sensitivity, precision, and inference speed. Building on these results, we further validated that multimodal fusion of T1WI and T2WI significantly improves both detection performance and computational efficiency relative to single-modality inputs. Collectively, these findings provide a practical framework for integrating deep learning-based assistive tools into clinical workflows for early ONFH diagnosis^[10, 23].

Architectural advantages of YOLO v11 underpin its superior performance. The C3K2 module, equipped with dynamic convolution kernels, adaptively modulates the receptive field according to lesion size, thereby enhancing the detectability of subtle foci. The C2PSA module, which integrates channel separation with position-sensitive attention, globally emphasizes pathological regions while suppressing background interference. The decoupled head design—employing CIOU loss for bounding-box regression and Focal loss for classification—effectively addresses the class imbalance inherent to our dataset (55.6% positive samples)^[24]. By contrast, YOLO v13 achieved higher localization precision (mAP@.5:.95: 0.954 vs. 0.845) but at a 2.3-fold increase in inference time (11.3 ms vs. 4.9 ms). Such precision gains, while statistically notable, do not justify the substantial computational cost in clinical settings, reinforcing the necessity of prioritizing deployment efficiency over isolated technical metrics.

Threshold selection has direct clinical implications. The maximum F1 score (0.92) was observed at a confidence threshold of 0.388; the F1 score remained above 0.8 for thresholds below 0.8, then declined precipitously. Precision peaked at 1.00 at a threshold of 0.983, rising from 0.8 to 0.95 in the low-confidence range but deteriorating sharply at higher thresholds. Recall reached 0.99 at a threshold of 0, stayed above 0.85 in the low-confidence range, and dropped to zero at high thresholds. These patterns indicate that low thresholds favor high recall at the expense of precision, whereas high thresholds improve precision but substantially compromise recall. In emergency screening, where minimizing missed diagnoses is paramount, a threshold near 0.388 is recommended. When false positives are of greater concern, a higher threshold may be considered, but should not exceed 0.8 to avoid a precipitous drop in recall. The precision-recall curve further substantiates the model's robustness, with AP values of 0.958 for normal and 0.974 for abnormal femoral heads, and an overall mAP@0.5 of 0.966; precision consistently exceeded 0.95 across the 0–0.8 recall range. The confusion matrix revealed high true-positive rates for lesions, with minor

misclassifications between background and normal samples. Qualitative visualization demonstrated confidence scores of 0.7–0.9 for normal specimens and ≥ 0.9 for typical lesions, while lesions with ill-defined margins were assigned lower confidence—a property that facilitates prioritization of high-confidence findings and aids in early lesion identification.

Superiority of multimodal fusion derives from the complementary information furnished by T1WI and T2WI. T1WI provides high-resolution anatomical detail and clearly delineates necrotic margins, thereby furnishing precise spatial cues for localization; T2WI is exquisitely sensitive to bone marrow edema and inflammatory responses, which constitute early hallmarks of active pathology^[25, 26]. This complementarity enables the model to concurrently optimize localization accuracy and lesion discriminability^[27]. By employing early fusion via channel concatenation, the model processes both sequences simultaneously at the input layer, fully exploiting their spatial correspondence while circumventing information loss that typically accompanies late-fusion strategies. Notably, the per-frame inference time for multimodal input was shorter (3.0 ms) than that for T1WI alone (4.9 ms), possibly reflecting more efficient utilization of GPU parallel computing resources with dual-channel input—a finding that further supports the clinical feasibility of fusion in time-constrained environments.

Comparison with prior work contextualizes our contribution within the evolving landscape of deep learning in ONFH imaging. Most previous investigations have centered on segmentation or classification paradigms. For segmentation, the U-Net architecture introduced by Ronneberger et al.^[28] has served as a foundational framework. Peixu Wang et al.^[29] augmented U-Net with PointRend for lesion detection on early-stage ONFH MRI, attaining an AUC of 0.97. Liang C et al.^[30] further refined U-Net by incorporating geometric and pixel-level supervision, achieving 97.73% accuracy in femoral head segmentation. Li et al.^[31] proposed a multiscale geometric embedding convolutional neural network that leverages geometric supervisory signals—including length and area losses—to enhance the Dice coefficient to 93.34%. Cao et al.^[32] extended the YOLO framework to YOLOv12 with a Mamba-like linear attention module for hip MRI region-of-interest detection, achieving efficient localization of the acetabulum and necrotic regions. Wang et al.^[33] developed a GAM-YOLO segmentation model for multi-sequence MRI (T1WI, FS-T2WI, and Cor STIR) diagnosis of early ONFH, achieving a diagnostic accuracy of 97.30%. Regarding staging, Xianyu Shen et al.^[34, 35] developed models based on ResNeSt and DenseNet for early diagnosis and JIC classification, achieving an AUC of 0.98 and an overall multicenter accuracy of 83.8%, respectively. Kim et al.^[36] combined five MRI sequences via

self-supervised learning and an attention-fusion network to determine ARCO staging, achieving an AUC of 92.19%. While these studies adopted segmentation, classification or region-of-interest detection architectures, our work adopts an object-detection paradigm to systematically evaluate the YOLO family in ONFH detection, highlighting the comprehensive advantages of YOLOv11 in both accuracy and efficiency.

Clinical deployment strategies can be tailored to specific practice settings. For high-throughput emergency departments, YOLO v11 with multimodal fusion offers the optimal combination of speed (4.9 ms) and sensitivity (recall 0.941), effectively capturing subtle lesions while minimizing misses, with mAP@0.5 of 0.966 ensuring reliable localization for minute-level diagnosis of post-traumatic ONFH. For outpatient clinics, YOLO v10 provides comparable mAP@0.5 (0.966), a recall of 0.909, and a per-frame time of 5.1 ms, with a lightweight architecture that runs stably on standard GPUs or high-end CPUs. For primary care facilities with constrained hardware, YOLO v8 represents an ideal compromise (mAP@0.5=0.959, precision=0.914), with moderate computational demands facilitating deployment on mid-range devices. For research applications where real-time performance is not mandatory, YOLO v13 may be justified by its superior localization precision (mAP@.5:.95=0.954), provided that high-end GPU resources are available. These tiered recommendations enable balanced decision-making between performance and deployment costs across diverse healthcare contexts.

Although this study preliminarily validates the advantages of YOLO v11 combined with multimodal MRI in ONFH detection, several limitations must be acknowledged. First, the retrospective design and modest sample size constrain generalizability; prospective multicenter studies with larger, standardized cohorts are warranted. Second, although early fusion proved efficacious, we did not explore more sophisticated feature-level or decision-level fusion strategies, which might further harness the complementary strengths of multimodal data; hybrid Transformer-CNN architectures represent a promising avenue for enhanced integration^[37]. Third, the current model performs only object detection (bounding-box localization) without pixel-level segmentation, precluding volumetric quantification of necrotic tissue—a parameter of clinical relevance for surgical planning. Future iterations should extend to segmentation, volumetric assessment, and collapse risk prediction. Fourth, prospective validation in real-world clinical environments is lacking; rigorously designed human–machine comparison studies are essential to ascertain the model’s impact on radiologists’ diagnostic accuracy, reading time, and confidence. Finally, although the model

runs efficiently on mid-range GPUs, further lightweighting via pruning and quantization could facilitate deployment on portable or low-cost devices, broadening accessibility in resource-limited settings. Additionally, integration of clinical variables—including age, etiology, ARCO stage, and laboratory parameters—into imaging-based datasets could enable multidimensional decision-making, thereby further enhancing diagnostic accuracy and interpretability.

4 Conclusions

This study systematically compared six generations of YOLO models for MRI-based ONFH detection and identified YOLO v11 as offering the optimal balance of accuracy and speed. We further demonstrated that multimodal fusion of T1WI and T2WI significantly enhances both detection performance and computational efficiency, surpassing single-modality approaches. Our findings provide a robust benchmark for model selection in orthopaedic imaging and establish a technical foundation for developing intelligent, real-time assistive diagnostic systems for early ONFH, with clear potential for clinical translation.

List of abbreviations

Abbreviation	Full Term
ONFH	osteonecrosis of the femoral head
MRI	magnetic resonance imaging
YOLO	You Only Look Once
T1WI	T1-weighted Imaging
T2WI	T2-weighted Imaging
P	precision
R	recall
mAP	mean average precision
AUC	area under the curve
PR	Precision-Recall
AP	Average Precision

Declarations

Ethics approval and consent to participate

This retrospective study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Ethical approval was obtained separately from two independent institutional ethics

committees: (1) the Ethics Committee of The Second Affiliated Hospital of Inner Mongolia Medical University (Approval No. EFY20260014; Approval Date: February 10, 2026); and (2) the Ethics Committee of Zhengzhou Traditional Chinese Medicine Orthopedic Hospital Affiliated to Henan University of Chinese Medicine (Approval No. IECPI2024120801; Approval Date: December 30, 2024). The requirement for informed consent was waived because of the retrospective design and because all patient data were anonymized prior to analysis.

Consent for publication

Not applicable.

Availability of data and materials

The raw MRI data that support the findings of this study are not publicly available due to patient privacy and ethical restrictions. De-identified data, annotation files, and preprocessed datasets generated during the current study are available from the corresponding author upon reasonable request, subject to approval by the ethics committees of both participating centers and a data use agreement. The source code used for YOLO model training and evaluation is available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Key Project of the Inner Mongolia Natural Science Foundation (2025ZD013), the “Talent Revitalization of Inner Mongolia” Project Team of the Inner Mongolia Autonomous Region (2025TYL12), the Inner Mongolia Autonomous Region Science and Technology Project for the Construction of High-Level Clinical Specialties in Public Hospitals in the Capital Area (2024SGGZ102), the Inner Mongolia Autonomous Region Science and Technology Project for the Development of High-Level Clinical Specialties in Public Hospitals (2023SGGZ143), the Inner Mongolia Medical University Key Project (YKD2024ZD005)

Author Contributions Statement

Wang Xingru, Liu Kai, and Zhang Liwei contributed equally to this work as co-first authors. Liu Guangwen, Cao Yuju, and Wang Jianzhong are co-corresponding authors.

Wang Xingru: Data collection, image annotation, experimental implementation, and writing – original

draft.

Liu Kai: Image annotation, image review, and manuscript revision and editing.

Zhang Liwei: Image annotation, image review, manuscript revision and editing, and funding acquisition.

Ma Chao: Data collection, patient screening, and clinical information curation.

Meng Lingting: Data preprocessing, image format standardization, and clinical information verification.

Liu Jianming: Technical support, algorithm implementation, and figure preparation.

Xiao Hongxin: Technical support and data visualization.

Liu Guangwen: Co-corresponding author, technical supervision, and manuscript review.

Cao Yuju: Co-corresponding author, data collection, patient screening, and manuscript review.

Wang Jianzhong: Lead corresponding author, overall project coordination, overall supervision, funding acquisition, and final manuscript approval.

All authors have read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

References

- [1] Cui L, Zhuang Q, Lin J, et al. Multicentric epidemiologic study on six thousand three hundred and ninety five cases of femoral head osteonecrosis in China [J]. *International orthopaedics*, 2016, 40(2): 267-276.
- [2] Zhao D, Zhang F, Wang B, et al. Guidelines for clinical diagnosis and treatment of osteonecrosis of the femoral head in adults (2019 version) [J]. *Journal of orthopaedic translation*, 2020, 21: 100-110.
- [3] Li W L, Tan B, Jia Z X, et al. Exploring the Risk Factors for the Misdiagnosis of Osteonecrosis of Femoral Head: A Case-Control Study [J]. *Orthopaedic surgery*, 2020, 12(6): 1792-1798.
- [4] Kubo T, Ueshima K, Saito M, et al. Clinical and basic research on steroid-induced osteonecrosis of the femoral head in Japan [J]. *Journal of orthopaedic science : official journal of the Japanese Orthopaedic Association*, 2016, 21(4): 407-413.
- [5] Chen Y, Miao Y, Liu K, et al. Evolutionary course of the femoral head osteonecrosis: Histopathological - radiologic characteristics and clinical staging systems [J]. *Journal of orthopaedic translation*, 2022, 32: 28-40.
- [6] Hines J T, Jo W L, Cui Q, et al. Osteonecrosis of the Femoral Head: an Updated Review of ARCO on Pathogenesis, Staging and Treatment [J]. *Journal of Korean medical science*, 2021, 36(24): e177.
- [7] Mont M A, Salem H S, Piuze N S, et al. Nontraumatic Osteonecrosis of the Femoral Head: Where Do We Stand Today?: A 5-Year Update [J]. *J Bone Joint Surg Am*, 2020, 102(12): 1084-1099.

- [8] Dubé M D, Emara A K, Deren M E, et al. Techniques of core decompression in the treatment of idiopathic avascular necrosis of the femoral head [J]. Archives of orthopaedic and trauma surgery, 2024, 145(1): 82.
- [9] Pierce T P, Jauregui J J, Cherian J J, et al. Imaging evaluation of patients with osteonecrosis of the femoral head [J]. Current reviews in musculoskeletal medicine, 2015, 8(3): 221-227.
- [10] Choi H R, Steinberg M E, E Y C. Osteonecrosis of the femoral head: diagnosis and classification systems [J]. Current reviews in musculoskeletal medicine, 2015, 8(3): 210-220.
- [11] Mankin H J. Nontraumatic necrosis of bone (osteonecrosis) [J]. The New England journal of medicine, 1992, 326(22): 1473-1479.
- [12] Alkhatatbeh T, Alkhatatbeh A, Li X, et al. A single sequence MRI-based deep learning radiomics model in the diagnosis of early osteonecrosis of femoral head [J]. Frontiers in bioengineering and biotechnology, 2024, 12: 1471692.
- [13] Yan Z, Zhang B, Wang D. An FPGA-Based YOLOv5 Accelerator for Real-Time Industrial Vision Applications [J]. Micromachines, 2024, 15(9).
- [14] Ragab M G, Abdulkadir S J, Muneer A, et al. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023) [J]. IEEE Access, 2024, 12: 57815-57836.
- [15] Yeerjiang A, Wang Z, Huang X, et al. YOLOv1 to YOLOv10: A Comprehensive Review of YOLO Variants and Their Application in Medical Image Detection [J]. Journal of Artificial Intelligence Practice, 2024, 7(3).
- [16] Debsarkar S S, Aronow B, Prasath V B S. Advancements in automated nuclei segmentation for histopathology using you only look once-driven approaches: A systematic review [J]. Computers in biology and medicine, 2025, 190: 110072.
- [17] Chai R, Tian N, Wan G, et al. Automated detection of early-stage osteonecrosis of the femoral head in adult using YOLOv10: Multi-institutional validation [J]. European journal of radiology, 2025, 184: 111983.
- [18] Shinohara I, Inui A, Murayama M, et al. Quantification of Empty Lacunae in Tissue Sections of Osteonecrosis of the Femoral Head Using YOLOv8 Artificial Intelligence Model [J]. Journal of Biomedical Materials Research Part B: Applied Biomaterials, 2024, 112(12): e35512.
- [19] Li Y, El Habib Daho M, Conze P H, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification [J]. Computers in biology and medicine, 2024, 177: 108635.
- [20] Emmett L, Papa N, Counter W, et al. Reproducibility and Accuracy of the PRIMARY Score on PSMA PET and of PI-RADS on Multiparametric MRI for Prostate Cancer Diagnosis Within a Real-World Database [J]. Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 2024, 65(1): 94-99.
- [21] Wang X, Wang X, Zhang Y, et al. Development of the prediction model based on clinical-imaging omics: molecular typing and sentinel lymph node metastasis of breast cancer [J]. Annals of translational medicine, 2022, 10(13): 749.

- [22] Abdullah K A, Marziali S, Nanaa M, et al. Deep learning-based breast cancer diagnosis in breast MRI: systematic review and meta-analysis [J]. *European radiology*, 2025, 35(8): 4474-4489.
- [23] Wang J, Zheng Y, Luo J, et al. Applications of Deep Learning Models on the Medical Images of Osteonecrosis of the Femoral Head (ONFH): A Comprehensive Review [J]. *IEEE Access*, 2024, 12: 57613-57632.
- [24] Khanam R, Hussain M. YOLOv11: An Overview of the Key Architectural Enhancements [J/OL] 2024, arXiv:2410.17725[<https://ui.adsabs.harvard.edu/abs/2024arXiv241017725K>]. 10.48550/arXiv.2410.17725
- [25] Gao S, Zhu H, Wen M, et al. Prediction of femoral head collapse in osteonecrosis using deep learning segmentation and radiomics texture analysis of MRI [J]. *BMC medical informatics and decision making*, 2024, 24(1): 320.
- [26] Zubair M, Hussain M, Albashrawi M A, et al. A comprehensive review of techniques, algorithms, advancements, challenges, and clinical applications of multi-modal medical image fusion for improved diagnosis [J]. *Computer methods and programs in biomedicine*, 2025, 272: 109014.
- [27] Zhang R, Sheng J, Zhang Q, et al. A review of multimodal fusion-based deep learning for Alzheimer's disease [J]. *Neuroscience*, 2025, 576: 80-95.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. *CoRR*, 2015, abs/1505.04597.
- [29] Wang P, Liu X, Xu J, et al. Deep learning for diagnosing osteonecrosis of the femoral head based on magnetic resonance imaging [J]. *Computer methods and programs in biomedicine*, 2021, 208: 106229.
- [30] Liang C, Ma Y, Li X, et al. Aiding Diagnosis and Classifying of Early Stage Osteonecrosis of the Femoral Head with Convolutional Neural Network Based on Magnetic Resonance Imaging [J]. *Indian journal of orthopaedics*, 2025, 59(1): 121-127.
- [31] Li X, Lv S, Tong C, et al. MsgeCNN: Multiscale geometric embedded convolutional neural network for ONFH segmentation and grading [J]. *Medical physics*, 2023, 50(6): 3788-3800.
- [32] Cao L, Li R, Zhou D, et al. Deep Learning-Based Diagnosis and Classification of Femoral Head Necrosis; proceedings of the 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA), F 28-30 March 2025, 2025 [C].
- [33] Wang Y, Wu G, Wen D, et al. Global Attention Mechanism-Enhanced Deep Learning for Multi-Sequence MRI Diagnosis of Early Femoral Head Osteonecrosis [J]. *Annals of the New York Academy of Sciences*, 2026, 1562(1): e70354.
- [34] Shen X, Luo J, Tang X, et al. Deep Learning Approach for Diagnosing Early Osteonecrosis of the Femoral Head Based on Magnetic Resonance Imaging [J]. *The Journal of arthroplasty*, 2023, 38(10): 2044-2050.
- [35] Shen X, He Z, Shi Y, et al. Development and Validation of an Automated Classification System for Osteonecrosis of the Femoral Head Using Deep Learning Approach: A Multicenter Study [J]. *The Journal of arthroplasty*, 2024, 39(2): 379-386.e372.

- [36] Kim B, Lee G Y, Park S H. Attention fusion network with self-supervised learning for staging of osteonecrosis of the femoral head (ONFH) using multiple MR protocols [J]. Medical physics, 2023, 50(9): 5528-5540.
- [37] Hussain S S, Shah P M, Dawood H, et al. A swin transformer and CNN fusion framework for accurate Parkinson disease classification in MRI [J]. Scientific reports, 2025, 15(1): 15117.