Imaging AI for Femoral Head Osteonecrosis: PRISMA-DTA Meta-analysis of Diagnostic Accuracy

Peyman Mirghaderi, MD, MPH[1]; Nasim Eshraghi, MD[1]; Hyun Su Kim, MD, PhD[1]; Chankue Park, MD, PhD[1]; Navid Faraji, MD[2]; Majid Chalian, MD[1]; Peter Thurlow, MD[1]

[1] University of Washington, Department of Radiology, Seattle, WA • [2] University Hospitals Cleveland Medical Center, Cleveland, OH

BACKGROUND

- Avascular necrosis of the femoral head (**AVNFH**) may progress to femoral head collapse and arthroplasty if diagnosis is delayed.
- Early radiographs are often insensitive; **MRI** is most sensitive, but interpretation and staging still vary.
- Imaging **AI** could support earlier detection, improve consistency, and aid triage in high-volume or low-resource settings.
- **Objective:** quantify diagnostic accuracy of imaging AI models for AVNFH and test whether performance differs by modality, stage focus, and evaluation design.

METHODS

- PRISMA-DTA systematic review/meta-analysis; search of **PubMed, Embase, Web of Science, and Scopus** through Aug 2025.
- Two reviewers independently screened studies, extracted data, and assessed bias with **QUADAS-2**.
- Included original imaging AI studies evaluating AVNFH detection/staging with extractable diagnostic performance.
- For quantitative synthesis, one best model per study was used; **external test sets were prioritized** over internal validation.
- Sen/Spe were pooled with a **bivariate random-effects model** with prespecified subgroup analyses.

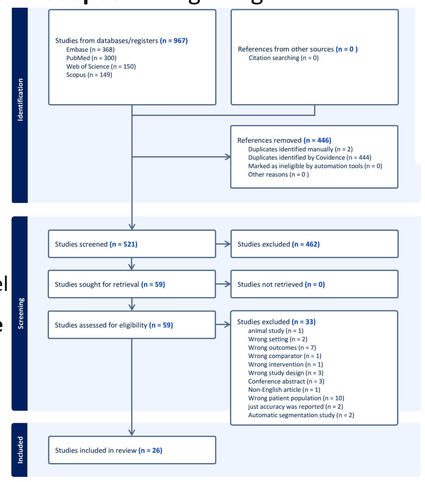


Figure 1. study selection flow diagram.

RESULTS

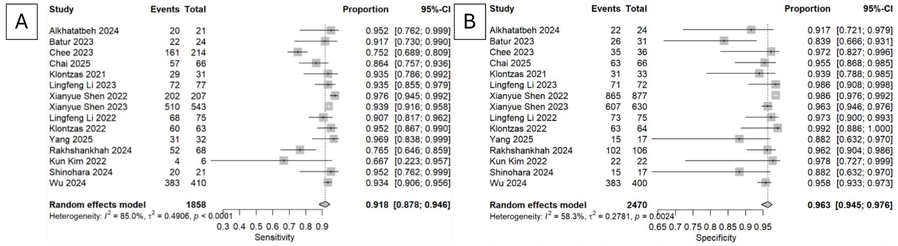
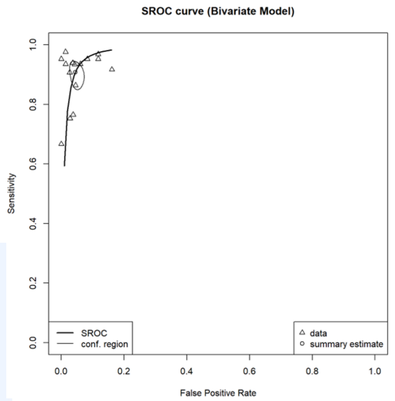


Figure 2. Overall pooled (A) sensitivity 0.92 (95% CI 0.88-0.95) and (B) specificity 0.96 (95% CI 0.95-0.98).



Bivariate summary point

Sensitivity: **0.91** Specificity: **0.96**

AUC: **0.98** DOR: **285**

Key subgroup findings

- MRI sensitivity exceeded radiography (0.95 vs 0.86; p=0.01).
- Early-stage-only vs any-stage and external vs internal evaluation showed no significant differences.

Figure 3. Bivariate HSROC with summary point (and 95% region) showed AUC 0.976.

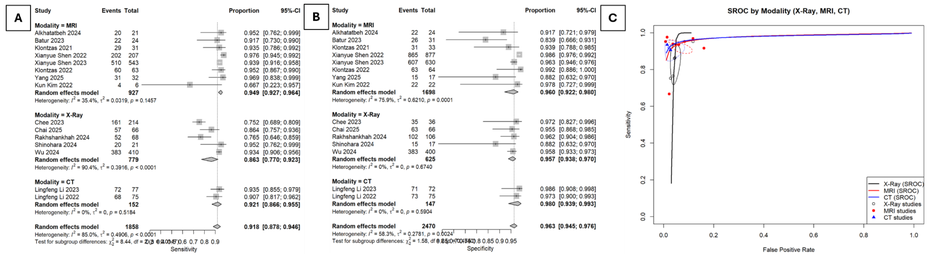


Figure 4. Modality subgroup: MRI was most (A) sensitive; (B) specificity was uniformly high across MRI, radiography, and CT. (C) SROC curves comparing Early-stage (red) versus Any-stage (black) studies. Summary performance was similar between subgroups (AUCs ≈ 0.97 vs. 0.98, p = 0.92).

DISCUSSION

- AI showed **high diagnostic accuracy** for AVNFH detection, with excellent specificity and AUC.
- MRI-based models were more sensitive than radiography-based, reflecting better conspicuity of early marrow changes on MRI.
- Early-stage-only studies and externally tested models performed similarly to broader-stage and internally tested models.
- **clinical implications:** MRI second-reader support, radiograph triage/referral support, and more reproducible staging.
- Limitations: many studies were retrospective, single-center, and heterogeneous.

CONCLUSIONS

- Imaging AI has strong potential for AVNFH detection.
- MRI-based AI appears more sensitive than radiography, while specificity remains high across modalities.
- Prospective multicenter validation and workflow studies are needed before routine clinical deployment.

REFERENCES

1. Page MJ et al. BMJ. 2021;372:n71.
2. McInnes MDF et al. JAMA. 2018;319:388-396.
3. Whiting PF et al. Ann Intern Med. 2011;155:529-536.
4. Reitsma JB et al. J Clin Epidemiol. 2005;58:982-990.
5. Petek D et al. EFORT Open Rev. 2019;4:85-97.