Papers for group discussion

Name	URL
Fidelity of Medical Reasoning in Large Language Models	https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2837372
AI-integrated Screening to Replace Double Reading of Mammograms: A Population-wide Accuracy and Feasibility Study	https://pubs.rsna.org/doi/10.1148/ryai.230529
Developing, Purchasing, Implementing and Monitoring AI Tools in Radiology: Practical Considerations. A Multi-Society Statement from the ACR, CAR, ESR, RANZCR and RSNA	https://pubs.rsna.org/doi/10.1148/ryai.230513
Collaborative integration of AI and human expertise to improve detection of chest radiograph abnormalities	https://pubs.rsna.org/doi/10.1148/ryai.240277
MammosighTR: Nationwide Breast Cancer Screening Mammogram Dataset with BI-RADS Annotations for Artificial Intelligence Applications	https://pubs.rsna.org/doi/10.1148/ryai.240841
Ethical Obligations to Inform Patients About Use of AI Tools	https://jamanetwork.com/journals/jama/article-abstract/2836687?utm_source=linkedin_company&utm_medium=social_jama&utm_term=17834243609&utm_campaign=article_alert&l
Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study	https://www.thelancet.com/journals/langas/article/PIIS2468-1253(25)00133-5/abstract
Validation of artificial intelligence software for automatic calcium scoring in cardiac and chest computed tomography	https://www.sciencedirect.com/science/article/pii/S0720048X25004097
Socio-Demographic Modifiers Shape Large Language Models' Ethical Decisions	https://link.springer.com/article/10.1007/s41666-025-00211-x?utm_source=rct_congratemailt&utm_medium=email&utm_campaign=nonoa_20250812&utm_content=10.1007%2Fs4

Name	URL
Collaboration between clinicians and vision–language models in radiology report generation	https://www.nature.com/articles/s41591-024-03302-1
DeepISLES: a clinically validated ischemic stroke segmentation model from the ISLES'22 challenge	https://www.nature.com/articles/s41467-025-62373-x
FactEHR: A Dataset for Evaluating Factuality in Clinical Notes Using LLMs	https://arxiv.org/abs/2412.12422v2
One shot at trust: building credible evidence for medical artificial intelligence	https://www.thelancet.com/journals/landig/article/PIIS2589-7500(25)00065-2/fulltext
Influence of AI Decision Support on Radiologists' Performance and Visual Search in Screening Mammography	https://pubs.rsna.org/doi/10.1148/radiol.243688
Cancer Detection in Breast MRI Screening via Explainable AI Anomaly Detection	https://pubs.rsna.org/doi/10.1148/radiol.241629
New Mammography Tools — The Need for Clinically Meaningful Assessment Standards	https://www.nejm.org/doi/full/10.1056/NEJMp2500274?query=TOC
Perspective: AI productivity will not benefit employed radiologists	https://www.sciencedirect.com/science/article/pii/S3050577125000313
Retrospective evaluation of interval breast cancer screening mammograms by radiologists and AI	https://link.springer.com/article/10.1007/s00330-025-11833-5
Medical Hallucination in Foundation Models and Their Impact on Healthcare	https://www.medrxiv.org/content/10.1101/2025.02.28.25323115v1.full
Current Pathology Foundation Models are unrobust to Medical Center Differences	https://arxiv.org/pdf/2501.18055
In-context learning enables multimodal large language models	https://www.nature.com/articles/s41467-024-51465-9?fromPaywallRec=false

Name	URL
to classify cancer pathology images	
MedCLIP-SAMv2: Towards universal text-driven medical image segmentation	https://www.sciencedirect.com/science/article/abs/pii/S1361841525002968?via%3Dihub
Redefining Bias Audits for Generative AI in Health Care	https://ai.nejm.org/doi/10.1056/AIp2500015
Evaluating Hospital Course Summarization by an Electronic Health Record–Based Large Language Model	https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2837483
MedAgentBench: A Virtual EHR Environment to Benchmark Medical LLM Agents	https://ai.nejm.org/doi/full/10.1056/AIdbp2500144
Asking the Right Questions: Benchmarking Large Language Models in the Development of Clinical Consultation Templates	https://www.arxiv.org/abs/2508.01159
CytoFM: The first cytology foundation model	https://openaccess.thecvf.com/content/CVPR2025W/CVMI/papers/Ivezic_CytoFM_The_first_cytology_foundation_r
Transformer patient embedding using electronic health records enables patient stratification and progression analysis	https://www.nature.com/articles/s41746-025-01872-z
Economic analysis of an AI-enabled ECG alert system: impact on mortality outcomes from a pragmatic randomized trial	https://www.nature.com/articles/s41746-025-01735-7
An evaluation framework for ambient digital scribing tools in clinical applications	https://www.nature.com/articles/s41746-025-01622-1
Generalist medical foundation model improves prostate cancer segmentation from multimodal MRI images	https://www.nature.com/articles/s41746-025-01756-2
Empirically derived evaluation requirements for responsible	https://www.nature.com/articles/s41746-025-01784-y

Name	URL
deployments of AI in safety-critical settings	
Artificial intelligence-derived retinal age gap as a marker for reproductive aging in women	https://www.nature.com/articles/s41746-025-01699-8
A scoping review and evidence gap analysis of clinical AI fairness	https://www.nature.com/articles/s41746-025-01667-2
Uncovering ethical biases in publicly available fetal ultrasound datasets	https://www.nature.com/articles/s41746-025-01739-3
A multimodal visual–language foundation model for computational ophthalmology	https://www.nature.com/articles/s41746-025-01772-2
Prospective pragmatic trial of automated retinal photography and AI glaucoma screening in Australian primary care	https://www.nature.com/articles/s41746-025-01768-y
How AI is used in FDA-authorized medical devices: a taxonomy across 1,016 authorizations	https://www.nature.com/articles/s41746-025-01800-1
Deep learning on routine full-breast mammograms enhances lymph node metastasis prediction in early breast cancer	https://www.nature.com/articles/s41746-025-01831-8
Transformer optimization with meta learning on pathology images for breast cancer lymph node micrometastasis	https://www.nature.com/articles/s41746-025-01833-6
Population-scale cross-sectional observational study for AI-powered TB screening on one million CXRs	https://www.nature.com/articles/s41746-025-01832-7
Framework for bias evaluation in large language models in healthcare settings	https://www.nature.com/articles/s41746-025-01786-w
Characteristics, licensing, and ethical considerations of openly accessible	https://www.nature.com/articles/s41746-025-01818-5

Name	URL
oral-maxillofacial imaging datasets: a systematic review	
Retrieval-augmented generation elevates local LLM quality in radiology contrast media consultation	https://www.nature.com/articles/s41746-025-01802-z
Benchmarking vision-language models for diagnostics in emergency and critical care settings	https://www.nature.com/articles/s41746-025-01837-2
Negative control-calibrated difference-in-difference analyses: addressing unmeasured confounding in RWD with application to racial/ethnic differences	https://www.nature.com/articles/s41746-025-01821-w
X-ray2CTPA: leveraging diffusion models to enhance pulmonary embolism classification	https://www.nature.com/articles/s41746-025-01857-y
Vision-language model for report generation and outcome prediction in CT pulmonary angiogram	https://www.nature.com/articles/s41746-025-01807-8
Automating liver biopsy segmentation with a robust, open-source tool for pathology research: the HOTSPoT model	https://www.nature.com/articles/s41746-025-01870-1
Systematic review on the technology's role in supporting lung cancer patients in the treatment journey	https://www.nature.com/articles/s41746-025-01913-7
Comparative analysis of AI support levels in clinical interpretation of traumatic pelvic radiographs	https://www.nature.com/articles/s41746-025-01923-5
A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare	https://www.nature.com/articles/s41746-025-01900-y
BlurryScope enables compact, cost-	https://www.nature.com/articles/s41746-025-01882-x

Name	URL
effective scanning microscopy for HER2 scoring using deep learning on blurry images	
Keyword-based AI assistance in the generation of radiology reports: A pilot study	https://www.nature.com/articles/s41746-025-01889-4
Machine learning based CAGIB score predicts in-hospital mortality of cirrhotic patients with acute gastrointestinal bleeding	https://www.nature.com/articles/s41746-025-01883-w
Evaluation of performance of generative large language models for stroke care	https://www.nature.com/articles/s41746-025-01830-9
Physics-informed machine learning digital twin for reconstructing prostate cancer tumor growth via PSA tests	https://www.nature.com/articles/s41746-025-01890-x
A consensus statement on the use of digital twins in medicine	https://www.nature.com/articles/s41746-025-01897-4
The role of face regions in remote photoplethysmography for contactless heart rate monitoring	https://www.nature.com/articles/s41746-025-01814-9
AI-driven preclinical disease risk assessment using imaging in UK biobank	https://www.nature.com/articles/s41746-025-01771-3
Predicting outcomes following endovascular aortoiliac revascularization using machine learning	https://www.nature.com/articles/s41746-025-01865-y
Re-identification of patients from imaging features extracted by foundation models	https://www.nature.com/articles/s41746-025-01801-0
Osteoarthritis progression pattern based on patient specific characteristics using machine learning	https://www.nature.com/articles/s41746-025-01878-7

Name	URL
Enhancing EHR-based pancreatic cancer prediction with LLM-derived embeddings	https://www.nature.com/articles/s41746-025-01869-8
Clinical and economic impact of a large language model in perioperative medicine: a randomized crossover trial	https://www.nature.com/articles/s41746-025-01858-x
Demographic inaccuracies and biases in the depiction of patients by artificial intelligence text-to-image generators	https://www.nature.com/articles/s41746-025-01817-6