ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 9| September 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

FOUNDATION MODELS AND VISION-LANGUAGE MODELS IN RADIOLOGY: CLINICAL TRANSLATION, SAFETY, AND EVALUATION OF MULTIMODAL MEDICAL IMAGING AI

Sudha Y Soudi[1], Sacheena Kumbara[2], Prateek Yalawar[3]

[1]*Assistant Professor, Department of Medical Imaging Technology (AHS), Shree B.M. Patil Medical College, Research Centre and Hospital, Vijayapur – 58610, India.*

[2]*Assistant Professor, Department of Medical Imaging Technology, Shantha College of Allied Health Sciences, Bangalore – 562104, India.*

[3]*Assistant Professor, Department of Medical Imaging Technology, Srinivas Institute of Allied Health Sciences, Srinivas University, Mangalore – 574146, India.*

ABSTRACT

Background: Foundation models and vision-language models (VLMs) are reshaping medical artificial intelligence by enabling models to learn broad representations from heterogeneous imaging and textual data and subsequently adapt them to multiple clinical tasks. In radiology, these systems can integrate images, reports, clinical narratives and other contextual information, potentially extending artificial intelligence beyond narrowly defined classification or detection tasks.

Objective: This review examines the clinical translation of foundation models and VLMs in radiology, with emphasis on multimodal medical imaging, evaluation, patient safety, workflow integration, explainability, bias, privacy, and human oversight.

Methods: A narrative review was developed from contemporary peer-reviewed literature and authoritative technical and regulatory sources addressing medical foundation models, vision-language learning, multimodal radiology AI, clinical evaluation and responsible deployment. The literature was synthesized around model capabilities, clinical use cases, evaluation dimensions, safety risks and implementation requirements.

Results: Foundation models can support image interpretation, report generation, visual question answering, retrieval, summarization, clinical decision support and multimodal reasoning. However, strong benchmark performance does not establish clinical effectiveness. Translation requires external validation, calibration, subgroup analysis, robustness testing, prospective workflow evaluation, monitoring for hallucinations and clinically meaningful human-AI interaction. Particular risks include dataset shift, demographic and institutional bias, automation bias, privacy leakage, unsupported recommendations and opaque failure modes.

Conclusion: Radiology foundation models should be regarded as assistive clinical technologies rather than autonomous diagnostic replacements. A safe translation pathway requires clinically representative datasets, transparent evaluation, rigorous prospective testing, governance, cybersecurity, continuous monitoring and clearly defined accountability. Multimodal AI has substantial potential, but its clinical value will depend less on model scale alone and more on evidence quality, safety engineering and integration with radiologist-led care.

KEYWORDS: *Foundation Models; Vision-Language Models; Radiology; Multimodal AI; Medical Imaging; Generative AI; Clinical Translation; Patient Safety; Evaluation; Explainable AI.*

1. INTRODUCTION

Artificial intelligence in radiology has progressed from task-specific computer-aided detection systems toward large-scale foundation models capable of learning general-purpose representations from diverse datasets. Earlier radiology AI systems were typically developed for a narrowly defined output, such as classification of a finding, segmentation of an anatomical structure or detection of a lesion. Although such systems can achieve high performance under controlled conditions, their utility may be constrained by limited task scope, dependence on curated datasets and reduced portability across institutions and populations.

Foundation models represent a broader paradigm. They are trained on large and heterogeneous datasets and are intended to provide transferable representations that can be adapted to multiple downstream tasks. In medical imaging, this approach is particularly relevant because clinical information is inherently multimodal: a radiologist interprets images together with the indication, previous examinations, laboratory findings, clinical history and reporting context. Vision-language models extend this concept by jointly modelling visual information and language, allowing an AI system to connect image features with textual descriptions and clinical questions.

Recent developments include contrastive image-text learning, large language models connected to imaging encoders, multimodal generative systems and radiology-specific models. These systems have demonstrated capabilities such as report drafting, image-text retrieval, visual question answering and

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 9| September 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

clinical reasoning assistance. Yet the transition from research demonstrations to routine patient care is complex. A model can generate fluent language while producing an incorrect interpretation, and a system that performs well on a public benchmark may fail when exposed to different scanners, acquisition protocols, patient populations or institutional reporting practices.

Therefore, the central question is no longer whether large multimodal models can process medical images, but whether they can do so reliably, safely and transparently within real clinical workflows. This review focuses on that translational gap.

2. OBJECTIVES

- To describe the role of foundation models and vision-language models in radiology and multimodal medical imaging.
- To review emerging clinical applications and potential workflow benefits.
- To examine methodological and clinical approaches for evaluating multimodal medical imaging AI.
- To identify major safety, ethical, privacy, bias and governance challenges associated with clinical deployment.
- To propose a practical framework for responsible clinical translation and continuous monitoring.

3. METHODOLOGY

This manuscript is a narrative review. Relevant literature was considered across the domains of medical foundation models, vision-language learning, multimodal radiology AI, generative artificial intelligence, clinical validation, safety and responsible AI. Priority was given to peer-reviewed publications, major methodological papers and authoritative guidance relevant to the evaluation and deployment of AI in healthcare. The evidence was thematically synthesized rather than statistically pooled because of substantial heterogeneity in model architecture, datasets, clinical tasks and outcome measures.

4. FOUNDATION MODELS AND VISION-LANGUAGE MODELS IN RADIOLOGY

4.1 Foundation models

A foundation model is a broadly pretrained model that can be adapted to multiple downstream applications. In radiology, foundation-model development may involve large collections of medical images, radiology reports, clinical text or combinations of these data. The objective is to learn representations that capture anatomical, pathological and semantic relationships that can later be transferred to specific tasks.

4.2 Vision-language models

VLMs combine visual encoders with language representations. In radiology, the visual component may process radiographs, CT, MRI, ultrasound or other imaging data, while the language component processes reports, prompts or clinical questions. Depending on architecture and training strategy, a VLM can align images and text, retrieve relevant reports, answer questions about images or generate textual descriptions.

4.3 Multimodal medical AI

Clinical reasoning rarely depends on an image alone. Multimodal systems can potentially combine imaging with structured clinical information, prior reports, demographics, laboratory values and other contextual information. This creates opportunities for more context-aware assistance but also introduces additional sources of error. Incorrect, outdated or irrelevant clinical context can influence model outputs, making provenance and input validation essential.

Table 1. Potential capabilities of radiology foundation models	Clinical relevance
Image–text alignment	Links imaging findings with radiology language and enables retrieval or semantic search.
Report generation and drafting	May assist preliminary reporting, structured reporting and communication of findings.
Visual question answering	Allows clinicians to query an image or imaging study using natural language.
Multimodal summarization	Can combine images with reports and selected clinical context.
Transfer learning	Supports adaptation to downstream imaging tasks with less task-specific training.
Decision-support assistance	May surface relevant findings, differentials or contextual information for clinician review.
Cross-modal retrieval	Can retrieve images from textual descriptions or text associated with similar imaging.

5. CLINICAL APPLICATIONS

Radiology foundation models have potential applications across the imaging pathway. In chest imaging, multimodal systems may assist with detection and description of common findings, report drafting and longitudinal comparison. In neuroimaging, they may support the integration of imaging observations with clinical descriptions. In abdominal, musculoskeletal and oncological imaging, multimodal systems may assist image-text retrieval, structured reporting and information synthesis. In emergency radiology, rapid summarization and prioritization may be particularly valuable when large numbers of studies compete for attention.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 9| September 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Report generation is one of the most visible applications. A VLM can generate a preliminary report from an image, potentially reducing repetitive documentation. However, linguistic fluency must not be equated with factual accuracy. Hallucinated findings, omission of subtle abnormalities, incorrect laterality and inappropriate certainty can have direct clinical consequences.

Another important application is retrieval and knowledge support. Multimodal models can potentially connect a current study with prior reports or similar cases. This may facilitate longitudinal assessment and educational support. However, retrieval systems must preserve patient identity controls, ensure correct study matching and avoid presenting irrelevant or outdated information as clinically current.

Foundation models may also enable personalized workflow tools. For example, a system could summarize relevant clinical information before image interpretation or generate a structured preliminary impression for radiologist verification. Such applications should be designed as human-in-the-loop systems in which the radiologist retains authority over the final clinical interpretation.

6. EVALUATION OF MULTIMODAL RADIOLOGY AI

Evaluation should extend beyond accuracy on a single test dataset. A clinically useful model must demonstrate technical validity, generalizability, calibration, robustness, safety and workflow value.

Table 2. Multidimensional evaluation framework	Key questions
Technical performance	Does the model perform accurately on representative imaging data?
External validation	Does performance remain acceptable at independent institutions?
Subgroup performance	Are errors distributed differently across sex, age, ethnicity, disease prevalence or other relevant groups?
Calibration	Do predicted probabilities or confidence estimates correspond to observed outcomes?
Robustness	Does performance remain stable with different scanners, protocols, image quality and acquisition conditions?
Clinical utility	Does the system improve diagnostic or workflow outcomes rather than only benchmark scores?
Human-AI interaction	Does the interface support appropriate reliance and reduce automation bias?
Safety	How often does the system hallucinate, omit clinically important findings or generate unsafe recommendations?
Monitoring	Can performance drift and emerging failure modes be detected after deployment?

7. SAFETY AND CLINICAL TRANSLATION

The clinical translation of foundation models requires explicit safety engineering. Generative models can produce plausible but unsupported statements, a phenomenon commonly described as hallucination. In radiology, hallucinations may involve fabricated findings, incorrect measurements, wrong anatomical localization or unjustified differential diagnoses. Because the language may appear highly professional, users may underestimate the probability of error.

Dataset shift is another major concern. Imaging equipment, protocols, patient demographics, disease prevalence and reporting conventions vary between institutions. A model trained predominantly on one population may perform differently elsewhere. External validation should therefore be performed using data that represent the intended deployment environment.

Automation bias can occur when clinicians over-rely on an AI recommendation. Effective interface design should make uncertainty and limitations visible and should support independent verification. AI output should not be presented with a level of certainty that exceeds the evidence supporting the prediction.

Privacy and cybersecurity are equally important. Medical imaging contains protected health information, and multimodal models may process large volumes of patient-linked text and images. Secure data governance, controlled access, auditability, appropriate de-identification and clear data-retention policies are necessary components of deployment.

Model updates create an additional governance issue. A system may change following retraining, fine-tuning or vendor updates. Version control, change documentation, revalidation and post-deployment surveillance are therefore necessary to maintain clinical safety.

8. BIAS, EQUITY AND GENERALIZABILITY

Foundation models inherit characteristics of their training data. Under-representation of particular populations, institutions or disease presentations can lead to uneven performance. In radiology, this may be amplified by differences in image acquisition and disease prevalence. Evaluation should therefore report subgroup performance wherever relevant and should examine clinically meaningful error patterns rather than only overall averages.

Equity also concerns access. Large models may require substantial computational infrastructure, specialized expertise and reliable digital connectivity. If deployment is concentrated in high-resource settings, benefits may not be distributed equally.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 9| September 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Efficient model architectures, validated adaptation strategies and responsible implementation planning may help reduce this gap.

9. EXPLAINABILITY AND HUMAN OVERSIGHT

Explainability in multimodal AI should be approached pragmatically. A model-generated explanation is not necessarily a faithful representation of the internal reasoning process. Useful clinical transparency may instead include evidence-linked outputs, highlighted image regions, cited source information, uncertainty estimates, model version, input provenance and clear indication of whether content was generated or retrieved.

Human oversight remains central. The appropriate role of AI is to augment radiologists by reducing repetitive work, supporting information synthesis and highlighting potentially relevant findings. The final diagnostic decision should remain with a qualified clinician operating within established institutional responsibilities and escalation pathways.

10. IMPLEMENTATION FRAMEWORK FOR RADIOLOGY DEPARTMENTS

- Define the clinical problem and intended user before selecting a model.
- Specify the model's input, output, limitations, intended population and failure conditions.
- Perform local and external validation using clinically representative data.
- Assess subgroup performance and identify clinically important disparities.
- Evaluate calibration, robustness and uncertainty rather than relying only on accuracy.
- Conduct silent or retrospective evaluation before active clinical deployment where feasible.
- Introduce human-in-the-loop workflows with clear responsibility for final decisions.
- Establish audit logs, incident reporting, model version control and post-deployment monitoring.
- Reassess the system after major software, model or workflow changes.
- Ensure governance covering privacy, cybersecurity, procurement, accountability and informed use.

11. SUGGESTIONS

Radiology departments should establish multidisciplinary AI governance teams involving radiologists, medical physicists, radiographers or technologists, clinical informaticians, data scientists, information-security personnel and institutional administrators. Evaluation protocols should be agreed before deployment, including predefined safety thresholds and clinically meaningful outcomes. Model outputs should be incorporated into workflows only after determining how they affect reporting time, diagnostic accuracy, workload and patient safety.

Future studies should increasingly use prospective, multicentre designs and real-world endpoints. Reporting should include dataset provenance, patient demographics, imaging protocols, reference standards, missing-data handling, subgroup analyses and failure cases. Reproducibility and transparent reporting are essential for meaningful comparison between multimodal systems.

12. CONCLUSION

Foundation models and vision-language models represent a major evolution in radiology AI because they can connect medical images with language and broader clinical context. Their potential applications include report assistance, multimodal summarization, image-text retrieval, visual question answering and decision-support augmentation. Nevertheless, clinical translation requires more than impressive demonstrations or benchmark performance. The principal requirements are external validation, robust evaluation, subgroup analysis, uncertainty awareness, privacy protection, cybersecurity, human oversight and continuous post-deployment monitoring.

The future of multimodal radiology AI should therefore be defined by responsible augmentation rather than autonomous replacement. Models that are clinically validated, transparent about limitations and embedded within safe workflows may provide meaningful benefits while preserving radiologist accountability and patient-centred care.

13. AREA FOR FURTHER RESEARCH

Future research should investigate prospective multicentre evaluation of foundation models, standardized multimodal benchmarks, clinically meaningful measures of report quality, hallucination detection, calibration, fairness across populations, federated or privacy-preserving training, model drift, radiologist-AI interaction and the effect of AI on patient outcomes. Comparative studies should examine whether multimodal foundation models provide incremental clinical value over conventional task-specific AI systems and established radiologist workflows.

REFERENCES

1. *Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616:259–265.*
2. *Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. arXiv. 2021.*
3. *Moor M, Huang Q, Wu S, et al. Med-Flamingo: a multimodal medical few-shot learner. arXiv. 2023.*
4. *Zhang S, Xu Y, Usuyama N, et al. BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv. 2023.*
5. *Zhou Y, Chia YK, Adelani DI, et al. A foundation model for generalizable disease detection from chest X-rays. Nature. 2024.*
6. *Yang Z, Mitra A, Xu J, et al. Polymath: multimodal large language models for medicine. arXiv. 2023.*
7. *Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. 2023;620:172–180.*
8. *Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. Nature Medicine. 2025;31:943–950.*
9. *World Health Organization. Ethics and governance of artificial intelligence for health. WHO; 2021.*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 9| September 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

10. *World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. WHO; 2024.*
11. *U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. FDA; 2021.*
12. *Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet. 2019;393:1577–1579.*
13. *Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trials evaluating artificial intelligence interventions: the CONSORT-AI extension. Nature Medicine. 2020;26:1364–1374.*
14. *Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: SPIRIT-AI extension. Nature Medicine. 2020;26:1351–1363.*
15. *Sounderajah V, Ashrafian H, Golub RM, et al. Developing a consensus-based reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open. 2021;11:e047709.*
16. *Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine. 2019;17:195.*

AUTHOR DECLARATION

The authors declare that this manuscript is an original review prepared for academic publication and has not been simultaneously submitted to another journal. All authors should review and approve the final manuscript and author order before submission. Any institutional affiliation, correspondence details, funding statement, conflict-of-interest statement and acknowledgements should be completed and verified by the authors before submission.