Iraqi Journal of Intelligent Computing and Informatics (IJICI)
Vol. 4, 2. December 2025, pp. 194~211
ISSN:2791-2868,DOI: 10.52940/ijici.v4i2.147

Vision and Multimodal Foundation Models in Medical Imaging: A Comprehensive Review of Architectures, Clinical Trends, and Future Directions

Hikmat Z .Neima[1], Rana M. Ghadban[2], Ghaihab H. Adday[1]
[1] Department of Computer Science, College of CSIT, Basra University, Basra, Iraq
[2] Department of Intelligent Medical Systems, College of CSIT, Basra University, Basra, Iraq

Article Info

Article history:

Received Nov., 24, 2025
Revised Dec., 18, 2025
Accepted Dec., 29, 2025

Keywords:

Foundation Models ,
Medical Imaging ,
Multimodal Learning,
Medicine, Promotable Segmentation , Precision Vision Transformers.

ABSTRACT

Foundation models (FMs) are revolutionizing medical imaging by transitioning from task-specific algorithms to large-scale , generalizable systems that can learn from a broad range of multimodal data. Recent advances in these fields—transformer-based visual encoders , promptable segmentation architectures , vision–language models , and parameter-efficient fine-tuning—have resulted in improved performance among segmentation , detection , classification and report generation techniques in a variety of modalities such as MRI , CT , ultrasound , X-ray , endoscopy , and digital pathology. Domain specific FMs (including prostate MRI, brain MRI , retinal , ultrasound and pathology models) have proved to be effective in providing high label efficiency and competitive or better performance with the mainstream deep learning models , in particular under low-annotation conditions. Trends in the research emphasize such techniques as large-scale pretraining, multimodal integration , cross-task generalization , data-efficient learning , and the development of universal feature encoders. Simultaneously , extensive benchmarking and external validation indicate performance variability , motivating the continued development of standardized evaluation protocols. Adoption by clinical practice has been restricted because of interpretability , bias, workflow integration, computational requirements , and regulatory uncertainty. New options such as personalizable AI , continual learning , federated model adaptation , and imaging–genomics integration , stand out to make FMs key for the future of precision medicine. This article consolidates architectural , pioneering foundation models , clinical evaluation , and translational advancements , drawing upon the current context and future direction of foundation-model medical imaging.

Corresponding Author:
Hikmat Z. Neima
Department of Computer Science, College of CSIT, Basra University, Basra, Iraq
Email: hikmat.taher@uobasrah.edu.iq

1. INTRODUCTION

Foundation models (FMs) have been a disruptive paradigm of AI , where focus moves away from task-driven pipelines to large pre-trained and generalizable to a wide variety of clinical tasks , modalities, and data types. This shift is particularly powerful in medical imaging. Historically , medical image analysis has been performed using models trained on a constrained set of problems – organ segmentation , lesion diagnosis and disease identification , for example – as these require large datasets of annotated images with heavy dependence on the expert's knowledge. The advance of the FMs , however , has shifted the focus of medical image analysis towards unified , multimodal and multitask models that learn from large volumes of highly heterogeneous medical and non-medical data and rapidly learn new tasks with low supervised workload [1] , [2] , [3].

Recent architectural advances (e.g. , transformer-based visual encoders , vision–language models , promptable segmentation , SAM) have increased the representational power of medical AI by allowing for cross-modality reasoning, robust feature extraction, and improved generalization over institutions and populations. There is a growing emphasis on domain-specific FMs , with systems for prostate MRI , brain MRI , retinal images , chest radiographs, ultrasound, endoscopy , pathology and ECG interpretation outperforming traditional approaches especially in regions lacking labels [4] , [5] , [6].

Simultaneously , multimodal and multitask learning is becoming extensively used in the research community , as FMs combine imaging with clinical notes , laboratory data, genomics , and physiological information to help more complete diagnosis and prognosis modelling [7] , [8]. Self-supervised learning , few-shot learning , and parameter-efficient fine-tuning are enabling unprecedented label efficiency , while federated learning and privacy-preserving frameworks offer pathways for large-scale , multi-institutional model development without compromising patient confidentiality [9] , [10] , [11].

Notwithstanding such progress , clinical translation remains limited. Obstacles include interpretability , bias , regulatory uncertainty , computational constraints , and challenges in integrating FMs into real-world workflows such as PACS, RIS, and EHR systems [12] , [13]. Initial clinical evaluations show significant zero-shot and few-shot performance yet display variation across populations and modalities, highlighting the importance of rigorous benchmarking and external validation [14].

At the frontier , emerging opportunities such as personalized AI , continual learning , federated FM adaptation , and imaging–genomics integration are positioning FMs as major enablers of next-generationprecision medicine. These directions move beyond static , one-size-fits-all systems toward models that evolve with clinical practice, adapt to individual patient trajectories , and provide real-time , context-aware decision support [15] , [16].

This review synthesizes recent architectural advances , leading foundation model systems , major research trends , clinical evaluation practices , and emerging opportunities for personalized and continual-learning medical AI. By integrating insights across modalities , specialties , and methodological innovations , it provides a comprehensive perspective on the evolving landscape of foundation models in medical imaging and highlights key challenges and research directions necessary to achieve safe , equitable , and clinically effective deployment. This review focuses on vision-based , vision–language , and multimodal foundation models for medical imaging and clinical imaging workflows , and does not aim to comprehensively cover non-imaging large language model applications such as drug discovery or purely textual clinical decision support.

This review makes four key contributions. (1) It provides a consolidated taxonomy of foundation models in medical imaging , spanning general-purpose vision–language models and specialized medical-domain FMs across radiology , pathology , ophthalmology , ultrasound , ECG , and EHR applications. (2) It synthesizes recent architectural advances—including transformer backbones , multimodal and self-supervised pretraining , promptable segmentation , and parameter-efficient adaptation—and links them to major research trends. (3) It critically evaluates current clinical adoption , benchmarking practices , and translational barriers such as interpretability , bias , workflow integration , and regulatory challenges. (4) It outlines emerging directions for personalized AI , continual learning , federated collaboration , and multimodal clinical reasoning , offering a forward-looking agenda for safe and effective FM deployment in precision medicine.

2. LITERATURE SEARCH AND REVIEW METHODOLOGY

This comprehensive review synthesizes recent research on vision and multimodal foundation models (FMs) in medical imaging. The methodological approach follows **structured narrative review practices** commonly adopted in contemporary AI-in-medicine surveys , emphasizing breadth of coverage , conceptual synthesis , and translational relevance rather than exhaustive quantitative aggregation [17] , [13].

2.1 Search Strategy and Sources

A structured literature search was conducted across **Scopus, PubMed , IEEE Xplore , and Web of Science** , covering publications from **2019 to 2026**. The search combined keywords and phrases including foundation model, vision-language model , medical imaging foundation models , multimodal AI , zero-shot learning ,pretraining , and clinical adoption. This strategy was designed to capture both methodological advances and clinically oriented studies , and is consistent with search protocols used in recent surveys of multimodal and vision–language foundation models [18]. Additional relevant articles were identified through citation snowballing and cross-referencing of key review papers , following established study selection practices [19].

2.2 Inclusion Criteria

Studies were included if they:

1. Investigated foundation models , large vision models , or vision–language models applied to medical imaging.
2. Reported contributions related to model architecture , large-scale pretraining , adaptation strategies , multimodal reasoning , evaluation , or clinical translation.
3. Provided quantitative results or qualitative analysis relevant to medical imaging tasks or clinical workflows.

These criteria are aligned with inclusion practices adopted in recent foundation-model-focused medical imaging reviews [20].

2.3 Exclusion Criteria

Studies were excluded if they:

- Focused exclusively on non-medical applications of foundation models ,
- Addressed traditional deep learning approaches without foundation-model characteristics ,
- Lacked accessible full text , or
- Were not published in English.

2.4 Study Selection Approach

Study selection was conducted in two stages. First , titles and abstracts were screened to remove clearly irrelevant publications. Second , full-text assessment was performed to evaluate relevance with respect to foundation-model architecture , multimodal integration , evaluation methodology , and translational potential. This staged screening process follows established practices for structured narrative reviews in foundation-model and healthcare AI research [13].

2.5 Data Extraction and Synthesis

From each included study , information was extracted regarding:

- Model category (vision-only , vision–language , or multimodal foundation model) ,
- Architectural design (e.g. , transformer-based , SAM-derived , self-supervised , or parameter-efficient fine-tuning approaches) ,
- Pretraining scale , modality coverage , and data diversity ,
- Targeted downstream tasks (e.g. , segmentation , detection , report generation , or clinical reasoning) , and
- Clinical relevance , adoption barriers , and translational readiness.

The extracted findings were synthesized qualitatively to identify recurring architectural patterns , emerging research trends , evaluation practices , and open challenges , following synthesis strategies commonly used in recent foundation-model surveys in medical imaging [17].

2.6 Methodological Limitations

This review adopts a structured narrative approach rather than a fully systematic or meta-analytic methodology , which may introduce selection bias and limits the ability to quantitatively compare performance across studies. Although multiple major bibliographic databases were searched and citation snowballing was employed , it is possible that relevant studies—particularly preprints , non-English publications , or rapidly evolving industrial reports—were not captured. In addition , reported performance metrics across foundation model studies vary substantially in terms of datasets , evaluation protocols , and clinical contexts , which constrains direct cross-study comparability. Finally , many included works focus on proof-of-concept evaluations or retrospective analyses , and therefore their findings may not fully reflect real-world clinical deployment conditions. These limitations highlight the need for standardized benchmarks , transparent reporting practices , and prospective clinical validation in future foundation model research.

3. FOUNDATIONS OF VISION AND MULTIMODAL FOUNDATION MODELS

Foundation models (FMs) represent a major conceptual shift in artificial intelligence. Unlike traditional deep learning systems trained for narrow , task-specific purposes , FMs are large-scale architectures pre-trained on extensive datasets that enable strong generalization across diverse downstream applications. Their scalability , robust representation learning , and adaptability to multimodality have positioned them as central technologies in modern medical AI , supporting vision-based tasks , clinical text processing , and integrated healthcare analytics [21] , [22] , [23].

Two broad categories of FMs are now prominent in healthcare: general-purpose foundation models , originally developed for natural images or language; and medical-specific foundation models , pre-trained directly on clinical data across modalities such as MRI , CT , ultrasound , ECG , EHR , or whole-slide pathology images. Figure 1 shows general-purpose and medical-specific models across imaging modalities.

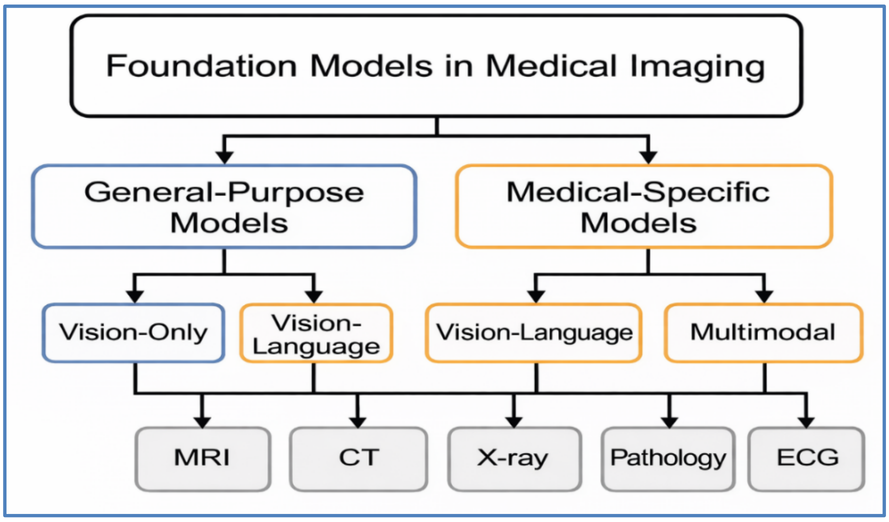

Figure 1: Taxonomy of Vision , Vision–Language , and Multimodal Foundation Models in Medical Imaging.

3.1 General-Purpose Foundation Models in Medical Imaging

General-purpose models , although not designed for clinical use , have been adapted for medical imaging and multimodal healthcare applications.

- CLIP and SigLIP.

CLIP and SigLIP are large-scale image–text alignment models that demonstrate strong performance in natural image understanding. However , studies have shown that these models generalize poorly to medical domains such as skin imaging , endoscopy , and oral diagnostics without additional domain-specific pretraining , due to the substantial visual distribution shift between natural and clinical imagery [24].

- Large Language Models (LLMs): GPT , BERT , PaLM , LLaMA.

Transformer-based LLMs have been widely used to summarize clinical notes , assist tumor board discussions , and support decision-making workflows. Their ability to encode medical knowledge contextually makes them powerful tools for downstream tasks in EHR analysis , clinical reasoning , and multimodal report generation [25].

- Segment Anything Model (SAM).

SAM introduced prompt-based universal segmentation capabilities. Although powerful in natural images , SAM underperforms in medical contexts—particularly for multimodal MRI or CT segmentation—unless enhanced through medical-domain adaptation mechanisms such as volumetric adapters or medical-specific fine-tuning [26] , [27].

Overall , general-purpose models provide strong priors but require substantial adaptation to achieve clinically reliable performance.

3.2 Medical-Specific Foundation Models

Medical-specific foundation models (FMs) address many of the limitations of general-purpose models by explicitly incorporating domain knowledge , clinical priors , and modality-specific structural characteristics during pretraining. Unlike natural-image-based FMs , these models are trained directly on large-scale medical datasets , enabling them to capture anatomical regularities , disease-specific patterns , and imaging physics that are essential for reliable clinical performance.

Modality- and Organ-Specific Foundation Models

A growing body of work has focused on developing FMs tailored to specific imaging modalities or anatomical systems:

- **PCaSAM (Prostate MRI):** A multimodal MRI foundation model designed for prostate cancer analysis , demonstrating improved segmentation accuracy and enhanced PI-RADS scoring across external validation cohorts [6].
- **SAM-Brain3D (Brain MRI):** A brain-specific FM trained on more than 66 ,000 labeled MRI volumes , achieving state-of-the-art performance across multiple brain sub-modalities and disease-related segmentation tasks [27].
- **CT-based Foundation Models:** Scalable CT foundation models have been proposed to support multi-organ analysis and cross-task generalization , enabling effective transfer across diverse anatomical regions and clinical tasks [4].
- **Retinal and Endoscopic Foundation Models:** Medical-domain foundation models have also been developed for retinal imaging and gastrointestinal endoscopy , supporting disease detection , segmentation , and classification with improved label efficiency and robustness compared with task-specific baselines [4].
- **USFM (Ultrasound):** A universal ultrasound foundation model trained on more than two million images across organs and acquisition protocols , demonstrating strong label efficiency and cross-organ generalization [4].
- **KED (ECG Foundation Model):** A large-scale ECG FM capable of zero-shot interpretation across cardiac rhythms and abnormalities , achieving cardiologist-level diagnostic performance and providing interpretable explanations [28].
- **Ark (Chest X-ray):** A radiography foundation model trained on large-scale public datasets , outperforming proprietary chest X-ray models in classification , localization , and segmentation tasks while maintaining strong generalization across datasets [29].

Pathology Foundation Models

Histopathology has emerged as one of the fastest-growing application domains for medical foundation models , driven by the availability of large whole-slide image datasets and the need for robust generalization across staining protocols and scanners:

- **CONCH , UNI , and Virchow2:** These pathology foundation models achieve high performance in colorectal cancer microsatellite instability prediction and demonstrate strong robustness to stain variation and resolution differences. CONCH , in particular , reported balanced accuracies exceeding 0.77 on external validation datasets [30].
- **Clinical Histopathology Imaging Evaluation Foundation Models:** Large-scale pathology FMs enable consistent and generalizable cancer characterization across multiple tumor types and datasets , supporting both diagnostic and prognostic workflows [31].

Graph-Based and Multimodal Foundation Models

Beyond image-centric architectures , foundation models are increasingly extended to structured and multimodal clinical data:

- **Foundation Model for Functional Connectivity:** A graph transformer-based FM designed for fMRI analysis , outperforming multiple competing approaches in cognitive and psychiatric prediction tasks by modeling complex brain connectivity patterns [32].
- **EHR Foundation Model (FM-SM):** A foundation model trained on large-scale electronic health record data , demonstrating strong hospital-to-hospital transferability and improved performance in low-label settings with minimal fine-tuning [23].

Collectively , these medical-specific foundation models illustrate how domain-aware pretraining enables the learning of anatomical structures , physiological relationships , and modality-dependent invariances that are difficult—or impossible—to acquire from natural-image or generic multimodal data alone. Their success underscores the importance of medical-domain data , task alignment , and modality-specific design in achieving clinically reliable and generalizable foundation-model-based AI systems.

3.3 Evidence of Effectiveness in Real Clinical Studies

Medical foundation models have demonstrated robust generalization , label efficiency , and clinical reliability:

- PCaSAM improved PI-RADS scoring by 8.3–8.9% and achieved DSC > 0.70 on external datasets [6].
- SAM-Brain3D outperformed state-of-the-art baselines across 14 MRI sub-modalities [27].
- USFM maintained strong multi-organ performance even with only 20% labeled data [4].
- KED matched experienced cardiologists in zero-shot clinical ECG interpretation [28].
- Ark surpassed Google’s CXR-FM in chest X-ray classification , localization , and segmentation [29].
- CONCH achieved pathologist-validated interpretability scores of up to 92.4% [30].
- FM-SM provided transferable EHR modeling with minimal data requirements [23].

These results confirm the superiority of medical-specific FMs for many real-world clinical applications.

3.4 Comparative Insights and Limitations

- General-purpose models (CLIP , SigLIP , SAM) offer broad generalization but show limited performance in medical imaging without domain adaptation [24] , [26].
- Medical-specific models typically achieve higher diagnostic accuracy , segmentation quality , and generalization due to domain-tailored pretraining [4].
- Parameter-efficient fine-tuning improves adaptation for low-resource clinical settings [33] , [34].
- Many models remain limited to their primary modality and face challenges when generalizing to cross-domain or cross-task settings.

Table 1 shows a summary of several foundation models.

Table 1. Selected Foundation Models and Their Effectiveness.

Model (Reference)	**Domain / Modality**	**Key Findings**
CLIP, SigLIP [24]	General-purpose	Limited performance on clinical images
PCaSAM [6]	Prostate MRI	Higher DSC, improved PI-RADS score
SAM-Brain3D [27]	Brain MRI	Multi-submodality state-of-the-art segmentation
USFM [4]	Ultrasound	High robustness, strong label efficiency
KED [28]	ECG	Zero-shot, cardiologist-level performance
Ark [29]	Chest X-ray	Outperforms Google CXR-FM
CONCH / UNI / Virchow2 [30]	Pathology	High accuracy, validated interpretability
FM-SM [23]	EHR	Strong transferability in low-label settings

4. ARCHITECTURAL ADVANCES IN MEDICAL FOUNDATION MODELS

Recent architectural developments in FMs have significantly reshaped the landscape of medical image analysis. Modern FMs are pre-trained on large-scale heterogeneous datasets and subsequently adapted to a wide spectrum of downstream clinical tasks with minimal labeled data. This shift has enabled stronger cross-domain generalization , improved robustness , and enhanced multimodal integration across segmentation , classification , detection , and localization tasks [35] , [36]. Evolution of medical foundation model architectures from CNN-based designs to transformers , promptable segmentation , and multimodal foundation models , highlighting key adaptation mechanisms is illustrated in Figure 2.

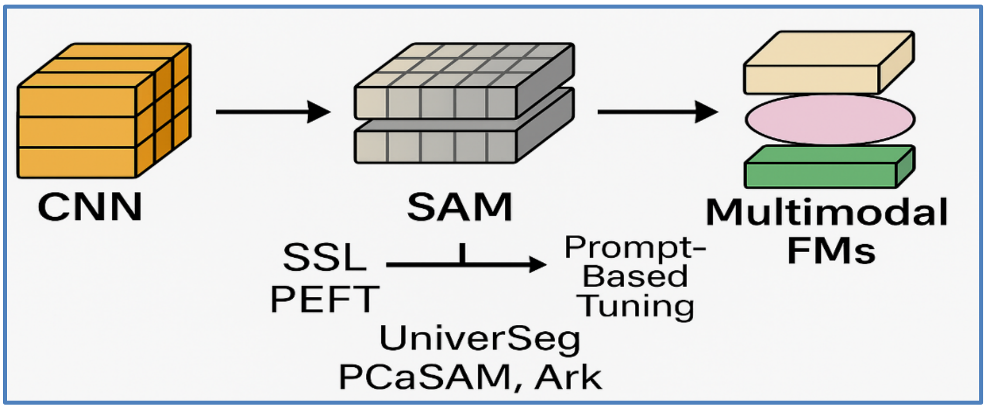

Figure 2: Evolution of Medical Foundation Model Architectures.

4.1 Transformer-Based and Hierarchical Vision Models

Vision Transformers (ViT) and hierarchical models like Swin Transformers are the preliminary works in the direction of general medical FMs. They have the ability to model long-range global dependencies and thus to model the anatomy structure and pathological variation more effectively. To the above end , Swin Transformers (with hierarchical shifted windows) can further increase efficiency and scalability for high resolution medical images [37] , [38]. These architectures are also robust for multimodal learning , assisting in incorporating imaging with complementary data , including clinical metadata.

4.2 Prompt-Based Segmentation Models: Adaptation of SAM for Medicine

SAM adaptation for the context of medicine. The Segment Anything Model (SAM) has brought prompt-based segmentation to FM design. Although SAM is able to generalize well to natural images , its application to medical images is limited by the appearance characteristic of the modality specific to these images. To fill in the existing gaps , the recent works like 3DSAM-adapter adopt volumetric computation , spatial adapters and parameter-efficient tuning methods to expand the possibilities of SAM in the case of CT or MRI segmentation tasks [39]. The adaptations show how promptable segmentation paradigms can effectively be scaled and transferred to clinical imaging.

4.3 Domain-Specialized Medical Foundation Models

In addition to generalized models , domain-specific FMs have also developed to address the specific requirements of medical imaging. UniverSeg presents a flexible , label-efficient segmentation scheme that is robust for many kinds of organs and modalities with few or no labels. Likewise , PCaSAM demonstrates competitive or superior performance on both internal and external validation cohorts compared with generalist FMs. These specific models demonstrate that the inclusion of medical-specific priors can be successful in FM architectures.

4.4 Multimodal and Cross-Modal Architectures

One of the hallmarks of the new generation of medical FMs is their utilization of non-image modalities , which allows a more interpretative clinical picture. Multimodal foundation models in medical imaging combine medical imaging with patient history , ECG signals, and radiology reports to optimize diagnostic reasoning [8] , [36]. Novel approaches like MedLAM , a self-supervised 3D anatomical localization model and MedLSAM

, a combined model of MedLAM and SAM , illustrate how multimodal synergy can alleviate some annotations and help make a more effective global approach in cases of complex volumetric data [5].

4.5 Self-Supervised Learning and Parameter-Efficient Adaptation

Self-supervised learning (SSL) has been developed for FM pretraining and provides an open platform for training on unlabeled inputs to learn semantically enriched representations. RadioDINO demonstrates that contrastive and self-distillation methods also yield robust features that outperform classical supervised methods in the classification and segmentation problems [40].

Parameter-efficient fine-tuning (PEFT) enables efficient next stage adaptation. Methods like Few-Shot Efficient Fine-Tuning (FSEFT) and Embedded Prompt Tuning (EPT) create little collections of learnable parameters — adapter or prompt tokens — to reposition FM representations for particular tasks to reduce computation to a minimum while keeping a high level of accuracy [33] , [34].

4.6 Generative and Hybrid Architectures

Generative and hybrid architectures are more and more used in FM pipelines. Autoencoder–GAN–transformer hybrid models enhance synthetic image plausibility , enhance high-level image enhancement and text-based interpretation , contributing to the transparency in diagnostic reporting [41]. Diffusion based architectures are also becoming effective for denoising , reconstruction and data augmentation. In the context of medical object detection , hybrid architectures that integrate YOLO , vision transformers and advanced attention mechanisms (e.g. , CSP , SPP , BiFPN) have been shown to have enhanced localization performance and efficiency in lesion detection tasks [42].

4.7 Challenges and Open Technical Gaps

While there have been significant advancements , there are still a number of technical issues to overcome. The general-purpose vision FMs suffer from potential underperformance on medical databases because of the gaps in their domain and the lack of medical priors , and thus may need to use domain adapted training or structural modification [35]. Scarcity , inconsistent validation protocols and a lack of large multimodal medical data sets also limit scalability and cross-institutional generalization [37]. It is important that such gaps be addressed as clinically reliable FM deployment is a goal. Key architectural advances in medical foundation models are represented in Table 2.

Table 2. Key Architectural Advances in Medical Foundation Models.

Model / Ref.	Domain / Application	Core Architectural Contribution	Remarks
ViT, Swin Transformer [37], [38]	Segmentation, classification	Global and hierarchical attention	Strong generalization; multimodal compatibility
SAM → 3D Adapters [39]	Organ / tumor segmentation	Prompt-based segmentation; volumetric tuning	Efficient adaptation; domain gap constraints
UniverSeg [5]	Multi-organ tasks	Domain-specialized segmentation	Outperforms task-specific baselines
PCaSAM [6]	Prostate cancer	Multimodal MRI segmentation	Strong internal/external generalization
MedLAM / MedLSAM [5]	Volumetric imaging	3D localization + SAM integration	Reduces annotation burden; competitive performance
RadioDINO (SSL) [40]	Classification, segmentation	Radiomics-aware self-supervision	Greater robustness vs supervised learning
FSEFT [33]	Few-shot segmentation	Parameter-efficient fine-tuning	Resource-efficient; high accuracy
EPT [34]	Few-shot segmentation	Embedded prompt tuning	Improves calibration and performance

GAN / AE / Diffusion Hybrids [41]	Synthetic data, enhancement	Generative modeling + attention	Better realism; interpretable outputs
YOLO + Hybrid Attention [42]	Lesion / object detection	CSP, SPP, BiFPN + ViT attention	High precision; fast inference

5. MAJOR RESEARCH TRENDS IN MEDICAL IMAGING FOUNDATION MODELS

Foundation models have rapidly become a core research direction in medical imaging , enabling improvements in diagnostic accuracy , workflow automation , and clinical decision-making across MRI , CT , X-ray , ultrasound , endoscopy , retinal imaging , dermoscopy , and pathology [1] , [2] , [43] , [44]. Current research trends center on pretraining scale , multimodal capabilities , data efficiency , robustness , and adaptation strategies , with growing emphasis on clinical translation.

5.1 Large-Scale Pretraining and Cross-Modality Generalization

A dominant trend is pretraining FMs on extremely large , heterogeneous imaging datasets to develop rich visual representations that transfer effectively to downstream tasks with minimal labeled data [7] , [45]. This enables broad generalization across modalities and pathologies.

Alongside generalist models , there is growing momentum toward **modality-specific FMs** that address the physics , artifacts , and diagnostic requirements of each modality—for example , ultrasound [45] , CT [5] , endoscopy [4] , and retinal imaging [24].

Research also explores adapting general-purpose vision FMs—such as SAM , ViT , and CLIP—for medical tasks through architectural modification or domain-specific pretraining [46] , [47] , balancing the benefits of large natural-image pretraining against fully medical-native models.

5.2 Expanding Applications and Model Capabilities

FMs now achieve expert-level or near-expert performance in core clinical tasks including organ delineation , tumor detection , and multidisease classification [43] , [46]. Models such as SAM-Med2D , MedSAM , and EyeCLIP demonstrate broad applicability across imaging types.

Emerging multimodal FMs integrate images , text , and clinical metadata , supporting cross-task generalization and unified pipelines for detection , segmentation , diagnosis , and report generation [7] , [48]. FMs are increasingly used as **universal feature encoders** , supplying robust high-level representations that benefit tasks ranging from tumor classification to disease progression modeling [47].

5.3 Data Efficiency and Robustness to Variability

Through self-supervision , zero-shot , and few-shot learning , FMs consistently reduce reliance on large annotated datasets. FMs also show increased resilience to real-world variation , including equipment differences , imaging protocols , annotation variability , and patient diversity. This robustness is critical for clinical deployment , particularly across multi-center settings. Self-supervised , semi-supervised , and few-shot learning remain major research directions aimed at mitigating data annotation challenges and improving generalization under data scarcity [2] , [17] , [43].

5.4 Model Adaptation , Evaluation , and Open Science

A key research question is whether medical FMs should be adapted from large natural-image models or built entirely from medical data. Evidence supports both approaches , with some studies demonstrating benefits from general-purpose pretraining for medical segmentation [39].

Given the diversity of tasks and modalities , the field is calling for unified evaluation frameworks and benchmarks to systematically assess model robustness and cross-domain generalization [2] , [17]. Large-scale public datasets such as GastroNet-5M and DIAS are accelerating model development and enabling reproducibility [17] , [45] , reflecting a strong trend toward open science.

5.5 Challenges and the Road Ahead

Clinical adoption requires interpretable and trustworthy AI. Explainability , fairness , and ethical deployment remain key research challenges [43]. Integration of imaging with genomics , proteomics , and other molecular data represents an emerging frontier aimed at uncovering biological mechanisms and enabling precision medicine [7] , [47]. High-variability modalities such as ultrasound , photographic imaging , and endoscopy highlight the need for specialized architectures to overcome noise , artifacts , and domain gaps [49] , [50].

Efficiency , robustness to degraded image quality , edge deployment , and workflow integration remain practical hurdles for real-world deployment. Table 3 illustrates the trends in medical imaging foundation models.

Table 3. Key Trends in Medical Imaging Foundation Models.

Trend / Theme (Ref.)	Core Description	Implications
Large-scale pretraining [7], [45]	Massive datasets enabling broad generalization	Improved cross-modality transfer and scalability
Modality-specific FMs [4], [24], [45]	Tailored models for CT, ultrasound, retinal imaging, endoscopy	Better robustness to modality-specific artifacts
Transfer learning and adaptation [39], [46], [47]	Adapting SAM, ViT, and CLIP for medical tasks	Balances general-purpose pretraining with medical specificity
Data efficiency and robustness [2], [17], [43]	Few-shot learning and resilience to domain shifts	Reduced annotation cost; improved multi-center generalization
Benchmarking and open science [17], [45]	Public datasets and standardized evaluation protocols	Improved reproducibility and fair comparison
Interpretability and ethics [43]	Explainability, fairness, and responsible deployment	Increased clinical trust and adoption
Multimodal and genomic integration [7], [47]	Imaging–genomics fusion for precision medicine	Deeper biological insight and personalized care
Real-world deployment [49], [50]	Workflow integration, robustness, and edge deployment	Practical clinical feasibility

Overall , research in medical imaging foundation models is advancing rapidly toward scalable , multimodal , data-efficient , and clinically aligned systems. Core trends include large-scale pretraining , modality-specific design , improved robustness , and ethical deployment. The field is steadily moving toward integrated , explainable models capable of supporting real-world clinical workflows at scale.

6. CLINICAL ADOPTION AND EVALUATION OF MEDICAL FOUNDATION MODELS

Although FMs are still at an early stage of clinical deployment , interest in their translational potential has accelerated across radiology , pathology , ophthalmology , neurology , and oncology. Their multimodal and multitask capabilities—spanning text , images , and structured data—support applications in diagnosis , treatment planning , report generation , and automated image analysis [3] , [12] , [51]. Figure 3 demonstrates clinical translation pipeline of foundation models from pretraining to deployment , highlighting key barriers and mitigation strategies for real-world adoption.

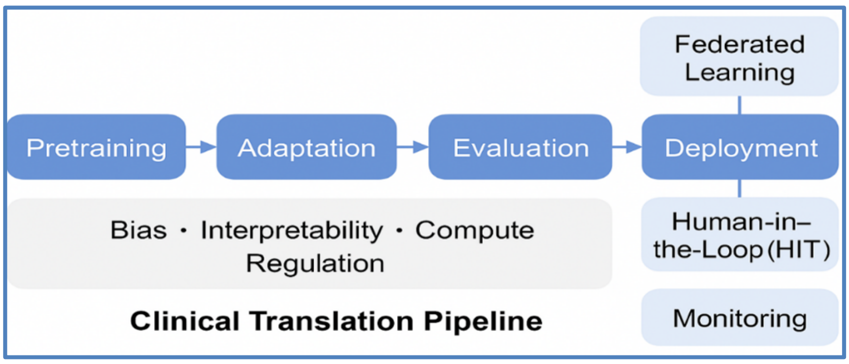

Figure 3: Clinical Translation Pipeline of Foundation Models.

6.1 Early Clinical Applications and Demonstrated Potential

FMs are increasingly evaluated as generalist backbones for a variety of clinical tasks:

- In digital pathology and oncology , FM-based systems have achieved high diagnostic accuracy and , in some settings , matched or exceeded pathologist performance after limited fine-tuning [3] , [51] , [52].
- In radiology , several foundation-model-based chest X-ray interpreters and multimodal diagnostic assistants have demonstrated promising performance in image interpretation and report generation , although routine clinical integration remains limited [12].
- Across multiple imaging domains , FMs leverage transfer learning to deliver strong performance even where labeled data are scarce , highlighting their potential for data-limited specialties and rare diseases [6] , [52].

These results support the notion that FMs can serve as versatile , reusable components in clinical AI pipelines—provided their behavior is rigorously evaluated and contextually constrained.

6.2 Evaluation Practices and Benchmarking Trends

A key translational trend is the shift from purely algorithmic metrics toward systematic , comparative evaluation:

- Studies increasingly compare FMs against strong task-specific baselines , particularly in segmentation and histopathology , to determine whether FM generality translates into real clinical benefit [14] , [52] , [53].
- Zero-shot and few-shot evaluation protocols are commonly employed to probe generalization to unseen datasets , institutions , and tasks [17] , [52]. These analyses often reveal that while FMs are highly flexible , performance can be uneven across populations and tasks.
- Some evaluations have shown that non-foundational , domain-specific models outperform general-purpose FMs in challenging histopathology tasks , emphasizing the importance of high-quality , multimodal medical data and thorough expert validation [14].

Overall , the community is moving toward rigorous benchmarking and reproducibility , recognizing that FM performance must be assessed under realistic clinical conditions , not just curated research datasets.

6.3 Barriers to Clinical Integration

Despite promising technical results , widespread clinical adoption remains limited due to multi-layered challenges:

- **Interpretability and trust.** Limited transparency and difficulty explaining FM decisions continue to erode clinician trust and impede regulatory acceptance [3] , [13].
- **Bias and generalizability.** FMs are sensitive to biases in training data and may not generalize across underrepresented populations , institutions , or imaging protocols [13] , [54].
- **Workflow and infrastructure integration.** Practical integration with PACS/RIS and EHR platforms remains non-trivial. Many FM prototypes operate as stand-alone research tools rather than seamlessly embedded components [12] , [55].
- **Computational and resource constraints.** Training and deploying large FMs require considerable computational resources and extensive curated datasets , which can exacerbate inequities between well-funded centers and resource-limited institutions [54] , [55]. **Regulatory and safety concerns.** Regulatory frameworks have yet to fully address FMs. Challenges include lack of formal approvals , limited external validation , and safety risks such as hallucinations and automation bias [12] , [53].
- **Ethical and societal issues.** Persistent concerns include privacy , fairness , and dependence on proprietary FMs controlled by a small number of commercial entities [54] , [56].

6.4 Emerging Solutions and Directions for Clinical Translation

To overcome these barriers , several strategies are being actively explored:

- **Human-in-the-loop and hybrid decision-making.** Human-in-the-loop frameworks aim to keep clinicians central in the decision process , using FMs as assistive tools rather than autonomous decision-makers [12].
- **Federated and privacy-preserving learning.** Federated learning and secure model-sharing paradigms are considered for multi-institutional FM training without centralizing sensitive data [12].
- **Continuous monitoring and national registries.** Proposals for ongoing performance monitoring , equity-focused design , and national registries seek to track real-world FM behavior and identify disparities over time [12] , [53].
- **Model compression and lightweight deployment.** Distillation and compression strategies are being developed to reduce computational demands. Integration into widely used platforms—such as QuPath for digital pathology—facilitates pragmatic adoption in routine workflows [55].
- **New reporting and evaluation standards.** Recent work calls for standardized reporting guidelines , more robust external validation , and closer collaboration between AI researchers , clinicians , and regulators to ensure safe , equitable deployment [53] , [56].

Table 4 demonstrates several trends and challenges in clinical adoption of foundation models.

Table 4. Trends and Challenges in Clinical Adoption of Foundation Models.

Clinical Theme (Reference)	Core Observations	Implications for Adoption
High diagnostic accuracy [3], [51], [52]	FMs can match or exceed clinician performance after fine-tuning	Strong translational potential, but requires validation
Need for rigorous evaluation [14], [17], [52], [53]	Benchmarking vs strong baselines; zero- and few-shot testing; external validation	Ensures reliable and reproducible clinical performance
Integration barriers [12], [13], [55]	Workflow incompatibility; limited regulatory approval; low clinician trust	Slows routine deployment in clinical environments
Data and compute demands [54], [55]	Large, diverse datasets and high computational requirements; risk of inequity	Limits adoption in resource-constrained settings
Ethical and societal concerns [54], [56]	Privacy, bias, explainability, dependence on proprietary models	Necessitates governance and ethical oversight
Emerging mitigation strategies [12], [53], [55], [56]	Human-in-the-loop, federated learning, lightweight models, registries, standards	Enables safer and more scalable clinical translation

7. FUTURE DIRECTIONS: PERSONALIZED , CONTINUAL , AND FEDERATED FM-BASED AI

7.1 Generalization as a Basis for Personalization

By leveraging diverse multimodal datasets and self-supervised objectives , FMs can generalize across a wide range of imaging modalities—including SPECT , PET , MRI , and CT—without requiring full retraining for every new task [16] , [20] , [57]. Their ability to support zero-shot and few-shot learning makes them particularly valuable in data-scarce settings such as rare diseases and novel theranostic applications , where annotated cohorts are inherently limited [57].

This broad generalization capability provides a strong foundation for patient-specific adaptation , where models can be tuned or conditioned on individual characteristics while retaining global clinical knowledge.

7.2 Emerging Opportunities for Personalized AI

A key future direction is **personalized medical AI**—models that integrate patient-level imaging features with other clinical and molecular signals to deliver individualized risk estimates , treatment recommendations , and prognostic trajectories. Multimodal FMs can fuse imaging data with EHRs , laboratory results , genomics , and physiological signals , supporting richer patient representations than imaging alone [8] , [15] , [58].

For example , integrating echocardiography with ECG signals , or MRI with longitudinal clinical history , can refine personalized cardiac diagnosis and therapy selection [8] , [58] . Earlier work on patient-specific modeling in cardiovascular imaging already demonstrated the clinical potential of individualized parameter estimation [59] , and FMs are poised to scale such approaches across modalities and diseases. Large language models and vision–language models further enable personalized report generation and knowledge-based interpretation , tailoring narrative explanations and decision support to the patient's context rather than producing generic outputs [20] , [60].

7.3 Continual Learning for Evolving Clinical Practice

In real clinical environments , data distributions evolve over time due to changing protocols , emerging diseases , and new imaging technologies. Continual learning techniques offer a pathway for FMs to adapt incrementally without catastrophic forgetting. Regularization-based , memory-based , and generative replay strategies allow models to incorporate new tasks or cohorts while preserving performance on previously learned tasks [61] , [62]. A continual-learning–based COVID-19 screening system , for instance , was able to progressively incorporate new chest X-ray data and classification categories while maintaining high accuracy , demonstrating the feasibility of on-the-fly adaptation in deployed systems [61]. Expansion-based and generative replay methods further support long-term deployment by selectively retaining task-relevant knowledge and mitigating data retention issues [62].

Combined with FMs , continual learning could support lifelong clinical AI systems that evolve alongside medical practice instead of remaining static after initial deployment.

7.4 Federated and Collaborative Learning at Scale

Federated learning complements FMs by enabling large-scale , multi-institutional collaboration without centralized data pooling. In this paradigm , local model updates are shared instead of raw patient data , addressing privacy constraints while leveraging diverse cohorts [11] , [60] , [63].

When combined with FMs , federated learning can:

- Improve robustness by exposing models to heterogeneous imaging protocols and populations.
- Support FM pretraining or adaptation across distributed centers.
- Enable privacy-preserving continual updates as hospitals accumulate new cases [29] , [63].

This suggests a future in which regional or global medical FMs are collaboratively trained and refined in a federated manner , with personalization layers added at each institution or even at the patient level.

7.5 Multimodal , Multi-Task , and Generative Learning

FMs naturally support multi-task learning , where tasks such as reconstruction , segmentation , detection , and quantification are jointly optimized. This has the potential to streamline workflows , reduce the number of separate models clinicians must maintain , and exploit task synergies [17] , [64].

Their multimodal capabilities , coupled with generative AI , also enable:

- Synthetic data generation to alleviate annotation bottlenecks.
- Modality translation (e.g. , PET from MRI) to support virtual imaging.
- Scenario simulation for rare conditions and edge cases [57] , [60].

These directions enhance both model robustness and the capacity for personalized scenario exploration.

7.6 Challenges and Research Priorities

Despite their promise , FM-based personalized and continual-learning systems face several unresolved challenges:

- **Interpretability and trust.** Many FM architectures remain black boxes , complicating clinical acceptance and regulatory approval [57] , [60].
- **Bias and fairness.** Personalized AI must avoid reinforcing existing disparities , especially when trained on skewed datasets [17] , [64].
- **Privacy and governance.** Even with federated learning , model updates may leak sensitive information if not properly secured [11] , [65].

- **Regulatory and computational constraints.** High training costs , energy consumption , and evolving regulatory frameworks remain major barriers to widespread adoption [60].

Future research should prioritize adaptive and personalized FM frameworks that combine transformer architectures , graph neural networks , and hybrid AI designs , while embedding explicit mechanisms for interpretability , uncertainty quantification , and data governance [8] , [60] , [64]. Emerging opportunities for personalized and continual foundation models-based AI are reported in Table 5.

Table 5. Emerging Opportunities for Personalized and Continual FM-Based AI.

Future Opportunity (Reference)	Role of Foundation Models	Enabling Techniques
Personalized AI [8], [15], [20], [57], [58], [60]	Patient-specific analysis and tailored recommendations	Multimodal integration; VLM/LLM-based interpretation
Continual learning [61], [62]	Ongoing adaptation to new data and tasks without forgetting	Regularization, memory-based methods, generative replay
Federated learning [11], [29], [60], [63], [65]	Privacy-preserving, multi-institutional FM training and adaptation	Distributed optimization; secure aggregation
Multi-task and multimodal learning [17], [57], [60], [64]	Joint reconstruction, segmentation, quantification, and data generation	Self-supervised and generative modeling
Clinical workflow enhancement [12], [55], [60]	Automated reporting, case triage, and decision support	Prompt engineering; VLMs; LLM-driven report generation

8. CONCLUSION AND FUTURE OUTLOOK

Foundation models have rapidly emerged as a transformative force in medical imaging , shifting the field from siloed , task-specific algorithms toward unified , multimodal , and adaptable AI systems. Across radiology , pathology , ophthalmology , cardiology , neurology , and oncology , recent advances demonstrate that foundation models can achieve—or even surpass—state-of-the-art performance in segmentation , detection , classification , and clinical reasoning , often with dramatically reduced annotation requirements. Their ability to leverage large-scale pretraining , multimodal alignment , and parameter-efficient adaptation positions them as foundational infrastructure for the next generation of clinical AI. Despite this progress , the path to widespread clinical adoption remains incomplete. Current foundation models still face significant challenges , including domain shift , limited interpretability , dataset biases , high computational demands , and the difficulty of integrating large models into complex clinical ecosystems such as PACS , RIS , and EHR systems. Benchmark studies highlight that foundation models can be extremely high quality but have variance across populations, modalities, and clinical settings. These results highlight the need for robust external validation, standardized evaluation guidelines, and for ongoing monitoring within a deployment to ensure reliability, equity, and safety. Simultaneously, emerging technology directions—including personalized AI, continual learning, federated foundation-model training, and imaging–genomics integration—provide promising avenues forward to address these limitations.

Personalized models can personalize predictions by patient clinical trajectory, while continual learning would enable the evolution of FMs with new data, emerging diseases, and changing clinical practices. Federated and privacy-preserving strategies are crucial for supporting integrated multi-institutional collaborations without introducing sensitive patient data, and multimodal reasoning holds promise for richer diagnostic insights through integration of imaging with biological, textual, and temporal information. These advances all suggest a future where foundation models take their place as integral parts of precision medicine: customizable, explainable, clinically informed AI systems that can help with diagnosis, prognosis, therapeutic planning, and long-term, individualized treatment. It will take progress not only on model architecture, but also on data governance, regulatory frameworks, the pathways to clinical integration, and human-centered design to achieve this vision. Foundation models represent a paradigm shift in medical imaging and hold great potential for scalable, generalizable, and patient-centric AI. In this review, we summarize, inter alia, the architectural advancements, popular models, research patterns, and challenges of translational implementation with future prospects, and a guidance to help researchers, clinicians and developers to enable the safe, equitable, and clinically relevant application of foundation-model based medical imaging to make good use of this rapidly evolving paradigm.

In the future, the next generation of foundation medical imaging models will likely transition away from static general-purpose architectures in the form of static models in general to dynamic systems which are context-aware and with learning from data. However, as health care environments constantly change—new imaging technologies, population shifts, and newly emergent diseases—foundation models must develop a mechanism for safe, real-time updating. Ongoing improvements in continual learning, test-time adaptation, and uncertainty-aware refinement of model will be pivotal to ensure deployed systems endure in the clinical future. Harmonization of training on the global datasets in future research will be key to enhance the diversity of populations while dealing with the issues of fairness and bias on underrepresented groups of people. Another large frontier in biointegration deals with multimodal biological data. The alignment between imaging, genomics, proteomics, metabolomics, and longitudinal physiological data allows for a new window into disease mechanisms, and personalised treatment response from patients. Vision–language–omics foundation models may constitute the core of next Gen precision medicine platforms, and for advanced phenotyping and predictive models that map the complete biological complexity of each patient. Last but not least, sustainable clinical utilization will demand new development efforts in computational efficiency, governance, and regulatory compliance. Lightweight, distilled versions of foundation models will be required to be developed for implementation under resource constraints and on edge devices, such as in portable ultrasound scanners. As a policy matter, transparent reporting standards, robust safety testing, and responsible data sharing mechanisms will determine the long-term acceptance and impact of these technologies. Given such coordinated steps across model design, infrastructure, and clinical collaboration, foundation models have the potential to be essential, trusted elements of global healthcare systems.

REFERENCES

[1] R. E. and B. M , “Foundation models in medical image analysis: Overview and prospects ,” in *Proceedings 2024 IEEE international conference on future machine learning and data science fmlds 2024* , 2024. doi: 10.1109/FMLDS63805.2024.00011.

[2] M. S. *et al.* , “A systematic review of generalization research in medical image classification ,” *Computers in Biology and Medicine* , vol. 183 , 2024 , doi: 10.1016/j.compbiomed.2024.109256.

[3] J. K. *et al.* , “A systematic review of vision and vision-language foundation models in ophthalmology ,” *Advances in Ophthalmology Practice and Research* , 2025 , doi: 10.1016/j.aopr.2025.10.004.

[4] J. J. *et al.* , “USFM: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis ,” *Medical Image Analysis* , vol. 96 , 2024 , doi: 10.1016/j.media.2024.103202.

[5] L. W. , X. W. , L. K. , Z. X. , and Z. S , “MedLSAM: Localize and segment anything model for 3D CT images ,” *Medical Image Analysis* , vol. 99 , 2025 , doi: 10.1016/j.media.2024.103370.

[6] Z. Y. *et al.* , “Generalist medical foundation model improves prostate cancer segmentation from multimodal MRI images ,” *npj Digital Medicine* , vol. 8 , 2025 , doi: 10.1038/s41746-025-01756-2.

[7] C. X. , Z. J. , T. Y. , Z. Y. , L. Y. , and Y. H , “Redefining imaging genomics for the next decade ,” *Science Bulletin* , vol. 70 , 2025 , doi: 10.1016/j.scib.2025.04.058.

[8] B. A. , A. N. , B. S. , and D. N , “Exploring advanced deep learning approaches in cardiac image analysis: A comprehensive review ,” *Computers in Biology and Medicine* , vol. 196 , 2025 , doi: 10.1016/j.compbiomed.2025.110708.

[9] J. M. *et al.* , “GastroNet-5M: a multicenter dataset for developing foundation models in gastrointestinal endoscopy.” 2025. doi: 10.1053/j.gastro.2025.07.030.

[10] L. Z. and M. N , “Multi-region ultrasound-based deep learning for post-neoadjuvant therapy axillary decision support in breast cancer.” 2025. doi: 10.1016/j.ebiom.2025.105916.

[11] E. S. , D. S. , S. L. , A. F. , N. E. , and T. S , “Artificial intelligence in cardiovascular imaging and intervention ,” *Herz* , vol. 49 , 2024 , doi: 10.1007/s00059-024-05264-z.

[12] U. P. and Y. C , “The emergence of foundation models in U.S. radiology: a narrative review of clinical utility , safety , and evaluation ,” *Academic Radiology* , 2025 , doi: 10.1016/j.acra.2025.10.063.

[13] T. M. , B. S. , R. M. , H. R. , J. C. , and C. E , “Harmonizing foundation models in healthcare: A comprehensive survey of their roles , relationships , and impact in artificial intelligence’s advancing terrain ,” *Computers in Biology and Medicine* , vol. 189 , 2025 , doi: 10.1016/j.compbiomed.2025.109925.

[14] A. S. *et al.* , “Foundation models for histopathology—fanfare or flair ,” *Mayo Clinic Proceedings Digital Health* , vol. 2 , 2024 , doi: 10.1016/j.mcpdig.2024.02.003.

[15] Y. F. , L. B. , Z. Z. , S. T. , B. V. , and D. J , “AI in SPECT imaging: Opportunities and challenges ,” *Seminars in Nuclear Medicine* , vol. 55 , 2025 , doi: 10.1053/j.semnuclmed.2025.03.005.

[16] B. Y. , L. J. , Y. C. , J. X. , and Y. Q , “Artificial intelligence in medical imaging: From task-specific models to large-scale foundation models ,” *Chinese Medical Journal* , vol. 138 , 2025 , doi: 10.1097/CM9.0000000000003489.

[17] N. S. and L. B , “A narrative review of foundation models for medical image segmentation: zero-shot performance evaluation on diverse modalities ,” *Quantitative Imaging in Medicine and Surgery* , vol. 15 , 2025 , doi: 10.21037/qims-2024-2826.

[18] Y. Lu and A. Wang , “Integrating language into medical visual recognition and reasoning: A survey ,” *Medical Image Analysis* , vol. 102 , p. 103514 , 2025 , doi: 10.1016/j.media.2025.103514.

[19] S. Waffenschmidt , E. Hausner , W. Sieben , T. Jaschinski , M. Knelangen , and I. Overesch , “Effective study selection using text mining or a single-screening approach: a study protocol ,” *Syst Rev* , vol. 7 , no. 1 , p. 166 , Dec. 2018 , doi: 10.1186/s13643-018-0839-x.

[20] R. J. , K. H. , C. Y. , and Y. S , “Vision-language foundation models for medical imaging: a review of current practices and innovations ,” *Biomedical Engineering Letters* , vol. 15 , 2025 , doi: 10.1007/s13534-025-00484-6.

[21] Z. S. and M. D , “On the challenges and perspectives of foundation models for medical image analysis ,” *Medical Image Analysis* , vol. 91 , 2024 , doi: 10.1016/j.media.2023.102996.

[22] A. M. , Z. A. , K. S. , K. K. , and L. S , “From task-specific to foundation models: A paradigm shift in medical vision-language analysis ,” *Computer Science Review* , vol. 59 , 2026 , doi: 10.1016/j.cosrev.2025.100831.

[23] G. L. *et al.* , “A multi-center study on the adaptability of a shared foundation model for electronic health records ,” *npj Digital Medicine* , vol. 7 , 2024 , doi: 10.1038/s41746-024-01166-w.

[24] Y. R. *et al.* , “Real-world readiness: Evaluating AI model performance in degraded photographic imaging for cancer detection ,” *Expert Systems with Applications* , vol. 303 , 2026 , doi: 10.1016/j.eswa.2025.130569.

[25] T. M. *et al.* , “The data scientist as a mainstay of the tumor board: global implications and opportunities for the global south ,” *Frontiers in Digital Health* , vol. 7 , 2025 , doi: 10.3389/fdgth.2025.1535018.

[26] Z. Y. , S. Z. , and J. R , “Segment anything model for medical image segmentation: Current applications and future directions ,” *Computers in Biology and Medicine* , vol. 171 , 2024 , doi: 10.1016/j.compbiomed.2024.108238.

[27] D. Z. *et al.* , “Brain foundation models with hypergraph dynamic adapter for brain disease analysis ,” *Pattern Recognition* , vol. 172 , 2026 , doi: 10.1016/j.patcog.2025.112595.

[28] T. Y. *et al.* , “Foundation model of ECG diagnosis: Diagnostics and explanations of any form and rhythm on ECG ,” *Cell Reports Medicine* , vol. 5 , 2024 , doi: 10.1016/j.xcrm.2024.101875.

[29] M. D. , P. J. , S. V. S. , G. M. , and L. J , “Ark + : Supervised training a single high-performance AI foundation model from many differently labeled datasets—no label consolidation required ,” *Medical Image Analysis* , vol. 108 , 2026 , doi: 10.1016/j.media.2025.103828.

[30] B. M. *et al.* , “Benchmarking pathology foundation models for predicting microsatellite instability in colorectal cancer histopathology ,” *Computerized Medical Imaging and Graphics* , vol. 127 , 2026 , doi: 10.1016/j.compmedimag.2025.102680.

[31] D. C. *et al.* , “Deep learning–powered whole slide image analysis in cancer pathology ,” *Laboratory Investigation* , vol. 105 , 2025 , doi: 10.1016/j.labinv.2025.104186.

[32] W. Y. , C. V. , P. G. , K. P. , E. T. , and Y , “A graph transformer-based foundation model for brain functional connectivity network ,” *Pattern Recognition* , vol. 169 , 2026 , doi: 10.1016/j.patcog.2025.111988.

[33] S.-R. J. , D. J. , and B. A. I , “Towards foundation models and few-shot parameter-efficient fine-tuning for volumetric organ segmentation ,” *Medical Image Analysis* , vol. 103 , 2025 , doi: 10.1016/j.media.2025.103596.

[34] Z. W. , X. S. , Z. Q. , L. G. , and M. L , “Embedded prompt tuning: Towards enhanced calibration of pretrained models for medical images ,” *Medical Image Analysis* , vol. 97 , 2024 , doi: 10.1016/j.media.2024.103258.

[35] Y. B. , O. N. , K. N. , A. G. , and I. I , “Radiomics and deep learning characterisation of liver malignancies in CT images – A systematic review ,” *Computers in Biology and Medicine* , vol. 194 , 2025 , doi: 10.1016/j.compbiomed.2025.110491.

[36] F. Z. M. , F. M. G. , and S. O , “Unifying healthcare intelligence: How foundation models are transforming heterogeneous medical data integration ,” in *International conference on artificial intelligence computer data sciences and applications acdsa 2025* , 2025. doi: 10.1109/ACDSA65407.2025.11166593.

[37] Y. S. *et al.* , “Deep learning techniques for automated coronary artery segmentation and coronary artery disease detection: A systematic review of the last decade (2013-2024 ,” *Computer Methods and Programs in Biomedicine* , vol. 268 , 2025 , doi: 10.1016/j.cmpb.2025.108858.

[38] Z. C. , D. X. , and L. S , “Next-gen medical imaging: U-net evolution and the rise of transformers ,” *Sensors* , vol. 24 , 2024 , doi: 10.3390/s24144668.

[39] G. S. *et al.* , “3DSAM-adapter: Holistic adaptation of SAM from 2D to 3D for promptable tumor segmentation ,” *Medical Image Analysis* , vol. 98 , 2024 , doi: 10.1016/j.media.2024.103324.

[40] Z. L. , L. A. , and R. C , “Radio DINO: A foundation model for advanced radiomics and AI-driven medical imaging analysis ,” *Computers in Biology and Medicine* , vol. 195 , 2025 , doi: 10.1016/j.compbiomed.2025.110583.

[41] A. A. and A. S , “Generative AI for diagnostic medical imaging: a review ,” *Current Medical Imaging* , vol. 21 , 2025 , doi: 10.2174/0115734056369157250212095252.

[42] S. M. , L. M. , and L. E , “Deep learning-based medical object detection: a survey ,” *IEEE Access* , vol. 13 , 2025 , doi: 10.1109/ACCESS.2025.3553087.

[43] L. D. *et al.* , “Deep learning in medical image analysis: a comprehensive review of algorithms , trends , applications , and challenges ,” *CMES - Computer Modeling in Engineering and Sciences* , vol. 145 , 2025 , doi: 10.32604/cmes.2025.070964.

[44] Q. B. , L. F. , Z. C. , and Z. D , “A review of development and future directions of medical foundation models ,” *Shuju Caiji Yu Chuli Journal of Data Acquisition and Processing* , vol. 40 , 2025 , doi: 10.16337/j.1004-9037.2025.03.002.

[45] L. W. *et al.* , “DIAS: A dataset and benchmark for intracranial artery segmentation in DSA sequences ,” *Medical Image Analysis* , vol. 97 , 2024 , doi: 10.1016/j.media.2024.103247.

[46] J. T. *et al.* , “Scaling up self-supervised learning for improved surgical foundation models ,” *Medical Image Analysis* , vol. 108 , 2026 , doi: 10.1016/j.media.2025.103873.

[47] K. S. *et al.* , “Advanced feature selection techniques in medical imaging—a systematic literature review ,” *Computers , Materials and Continua* , vol. 85 , 2025 , doi: 10.32604/cmc.2025.066932.

[48] G. Z. *et al.* , “Domain-continual learning for multi-center anatomical detection via prompt-enhanced and densely-fused MedSAM ,” *Information Fusion* , vol. 126 , 2026 , doi: 10.1016/j.inffus.2025.103614.

[49] L. J. , L. K. , L. C. , and T. Y , “A review of few-shot image classification: Approaches , datasets and research trends ,” *Neurocomputing* , vol. 649 , 2025 , doi: 10.1016/j.neucom.2025.130774.

[50] W. A. *et al.* , “Sam2Rad: A segmentation model for medical images with learnable prompts ,” *Computers in Biology and Medicine* , vol. 187 , 2025 , doi: 10.1016/j.compbiomed.2025.109725.

[51] D. F. , P. J. , K.-C. J. , G. E. , and B. C , “A review of deep learning for brain tumor analysis in MRI ,” *npj Precision Oncology* , vol. 9 , 2025 , doi: 10.1038/s41698-024-00789-2.

[52] Y. B. *et al.* , “Comparative analysis of convolutional neural networks and transformer architectures for breast cancer histopathological image classification ,” *Frontiers in Medicine* , vol. 12 , 2025 , doi: 10.3389/fmed.2025.1606336.

[53] V. S. *et al.* , “Value of artificial intelligence in neuro-oncology ,” *The Lancet Digital Health* , vol. 7 , 2025 , doi: 10.1016/j.landig.2025.100876.

[54] F. Y. , M. S. , I. H. , T. A. , F. K. , and K. E , “Artificial intelligence in clinical data analysis: A review of large language models , foundation models , digital twins , and allergy applications ,” *Allergology International* , vol. 74 , 2025 , doi: 10.1016/j.alit.2025.06.005.

[55] S. N. *et al.* , “A lightweight and extensible cell segmentation and classification model for H&E-stained cancer whole slide images ,” *Computers in Biology and Medicine* , vol. 199 , 2025 , doi: 10.1016/j.compbiomed.2025.111326.

[56] T. A. , N. A. , G. M. , B. R. , A. T. , and L. E , “Foundation models in healthcare: Opportunities , risks & strategies forward ,” in *Conference on human factors in computing systems proceedings* , 2023. doi: 10.1145/3544549.3583177.

[57] T. N. *et al.* , “Generative AI and foundation models in radiology: Applications ,” *Opportunities , and Potential Challenges. Radiology* , vol. 317 , 2025 , doi: 10.1148/radiol.242961.

[58] A. D. , T. Y. , and D. A , “Artificial intelligence: The new doctor in personalized medicine ,” *Current Pharmacogenomics and Personalized Medicine* , vol. 22 , 2025 , doi: 10.2174/0118756921371741250410152407.

[59] I. A. *et al.* , “Radiomics analysis for clinical decision support in nuclear medicine ,” *Seminars in Nuclear Medicine* , vol. 49 , 2019 , doi: 10.1053/j.semnuclmed.2019.06.005.

[60] C. G. and H. K , “Artificial intelligence augmented cerebral nuclear imaging ,” *Seminars in Nuclear Medicine* , vol. 55 , 2025 , doi: 10.1053/j.semnuclmed.2025.05.005.

[61] F. D. , K. M. , and H. N , “From lab to pocket: A novel continual learning-based mobile application for screening COVID-19 ,” *Engineering Applications of Artificial Intelligence* , vol. 163 , 2026 , doi: 10.1016/j.engappai.2025.112960.

[62] V. T. *et al.* , “Privacy-preserving continual learning methods for medical image classification: a comparative analysis ,” *Frontiers in Medicine* , vol. 10 , 2023 , doi: 10.3389/fmed.2023.1227515.

[63] T. Muazu , M. Yingchi , A. U. Muhammad , M. Ibrahim , O. Samuel , and P. Tiwari , “IoMT: A Medical Resource Management System Using Edge Empowered Blockchain Federated Learning ,” *IEEE Trans. Netw. Serv. Manage.* , vol. 21 , no. 1 , pp. 517–534 , Feb. 2024 , doi: 10.1109/TNSM.2023.3308331.

[64] J. Z. *et al.* , “The applications of artificial intelligence in radiographic imaging ,” *Chinese Science Bulletin* , vol. 70 , 2025 , doi: 10.1360/TB-2024-1352.

[65] L. X. , P. L. , W. Y. , and Z. W , “Open challenges and opportunities in federated foundation models towards biomedical healthcare ,” *Biodata Mining* , vol. 18 , 2025 , doi: 10.1186/s13040-024-00414-9.