Towards a unified foundation model for medical imaging

Lancet Digit Health 2026

Published Online https://doi.org/10.1016/j.landig.2026.101013

See Articles https://doi.org/10.1016/j.landig.2026.101007

Artificial intelligence (AI) has rapidly transformed medical imaging, enabling automated detection, classification, and risk prediction across an increasing number of clinical applications. However, in clinical practice, physicians routinely integrate information across imaging modalities, specialties, and diagnostic contexts. In their Article published in *The Lancet Digital Health*, Yang Zhou and colleagues introduce MerMED-FM, a multimodal, multidisease medical imaging foundation model trained on more than 3·3 million images across seven imaging modalities and more than ten clinical specialties. The study represents an important step towards a generalist AI framework that is capable of supporting cross-disciplinary imaging interpretation.[1]

Most existing medical imaging AI models have been developed as task-specific tools, typically focusing on a single modality such as chest radiography, pathology slides, retinal photographs, or CT imaging.[2,3] Although these specialist models can achieve excellent performance in narrow tasks, their fragmentation creates challenges for real-world use.[4] Hospitals often need to maintain multiple AI systems across departments, each requiring separate validation, governance, and technical integration. Zhou and colleagues attempt to address this limitation by developing a unified vision-only foundation model that can interpret diverse medical images within a single architecture. By pretraining the model on large-scale multi-institutional datasets covering radiology, pathology, ophthalmology, dermatology, and ultrasound and combining self-supervised learning with a memory-based training mechanism, MerMED-FM demonstrates competitive performance across multiple clinical imaging benchmarks. The conceptual transition from fragmented specialist models to generalist medical foundation models is shown in the figure.

One of the key contributions of the study by Zhou and colleagues is the demonstration that a single visual encoder can generalise across heterogeneous imaging modalities while maintaining competitive diagnostic performance. In particular, MerMED-FM achieved area under the receiver operating characteristic curve values exceeding 0·90 in several imaging benchmarks after fine-tuning with relatively small labelled datasets, thus highlighting the potential advantages of foundation model approaches in data-constrained clinical environments. The authors also report that MerMED-FM can match or outperform several state-of-the-art modality-specific models on select datasets, particularly in ophthalmic imaging and radiology tasks.[5–7] These findings reinforce the growing consensus that self-supervised pretraining on large-scale medical datasets might serve as a scalable strategy for developing widely applicable clinical AI systems.

Beyond technical performance, the study also highlights important implications for integrating AI into clinical workflows. Medical decision making is inherently multidisciplinary and often relies on the interpretation of multiple imaging modalities. For example, oncological diagnosis is generally based on findings of radiological imaging and histopathological analysis, whereas systemic diseases such as diabetes might require integration of retinal imaging with other organ-based evaluations.[5–8] A generalist imaging foundation model could serve as a shared diagnostic backbone across departments, aiding in screening, triage, and decision making within complex health-care environments.

Nevertheless, several important questions need to be addressed before such models can be widely adopted in clinical practice. First, although MerMED-FM demonstrates strong performance across multiple datasets, it does not consistently outperform specialist models in domains such as dermatology, pathology, and ultrasound. This observation highlights an ongoing trade-off between generalist and specialist models. Therefore, hybrid deployment strategies, where generalist foundation models serve as screening or triage tools and specialist models manage high-precision tasks, might represent a pragmatic path forward.

Second, the model was primarily trained on retrospective datasets and classification tasks. Real-world clinical implementation will require prospective validation studies to assess safety, workflow integration, and clinical impact. Importantly, future research should also evaluate whether such models can support true multimodal reasoning at the patient level, for example, by integrating CT, pathology slides, and ophthalmic imaging within a single diagnostic context.

Third, concerns regarding fairness, generalisability, and governance remain crucial. Although the authors conducted initial subgroup analyses by age and sex, broader

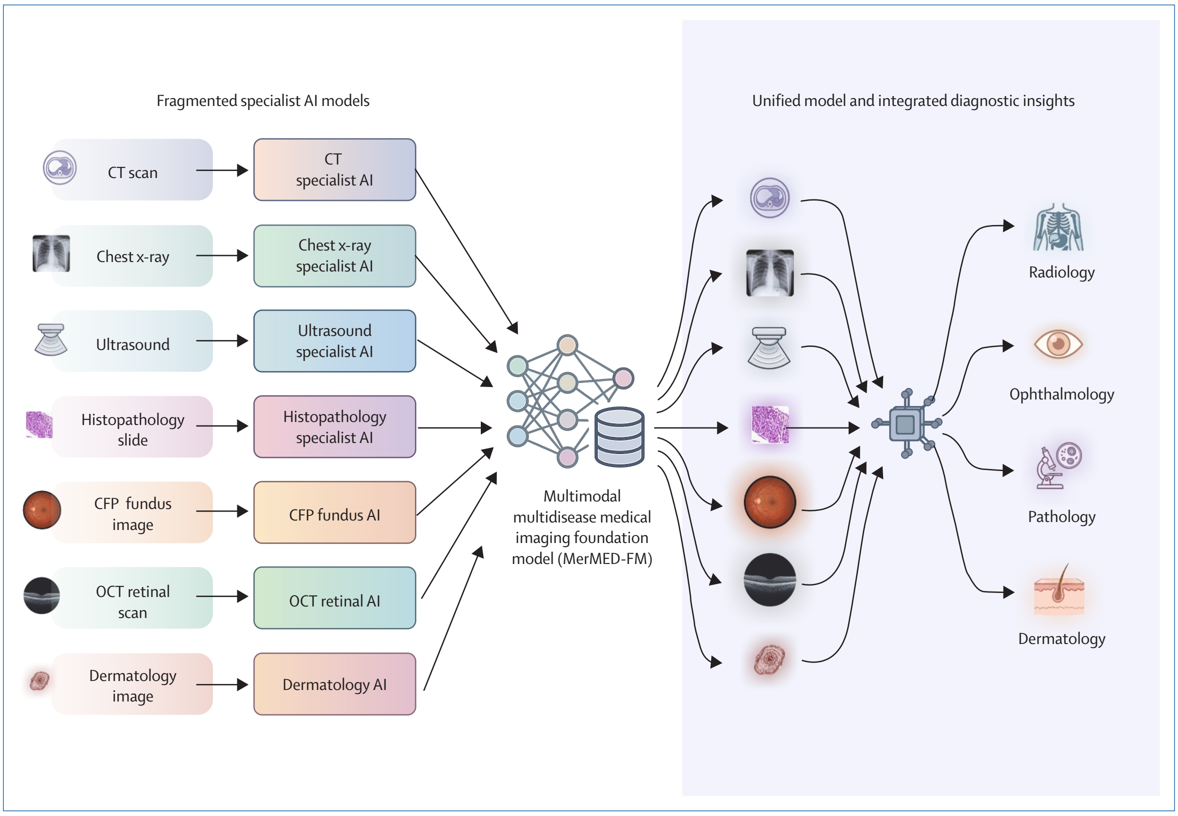

***Figure:* Conceptual transition from fragmented modality-specific AI systems to a unified multimodal multidisease medical imaging foundation model capable of interpreting seven imaging modalities**

AI=artificial intelligence. CFP=colour fundus photography. OCT=optical coherence tomography.

evaluations across diverse populations and health-care systems are needed to ensure equitable performance. In addition, regulatory frameworks and clinical governance structures will need to evolve to support increasingly generalist AI models that operate across multiple specialties.[9]

Finally, the development of large-scale foundation models raises broader questions regarding data access, model transparency, and global health equity. Training such models generally requires vast datasets and substantial computational resources, which could centralise innovation within a limited number of institutions. Ensuring that these technologies remain accessible and beneficial across diverse health-care settings is crucial.[10] The open-access release of the model will also promote transparency and global collaboration in medical AI, enabling researchers worldwide to evaluate and improve the system.

In summary, the study by Zhou and colleagues represents a notable advance towards unified AI systems capable of supporting medical imaging interpretation across specialties. Although further validation and refinement are required, multimodal medical foundation models such as MerMED-FM might represent an important step towards scalable, generalist AI infrastructures for future health-care systems.

We declare no competing interests.

*Kai Jin, Andrzej Grzybowski
jinkai@zju.edu.cn

Eye Center of Second Affiliated Hospital, School of Medicine, Zhejiang University, 310009 Hangzhou, China (KJ); Zhejiang Provincial Key Laboratory of Ophthalmology, Zhejiang Provincial Clinical Research Center for Eye Diseases, Zhejiang Provincial Engineering Institute on Eye Diseases, Hangzhou, China (KJ); Department of Ophthalmology, University of Warmia and Mazury, Olsztyn, Poland (AG); Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland (AG)

1 Zhou Y, Quek CWN, Zhou J, et al. Multimodal, multi-disease medical imaging foundation model (MerMED-FM). *Lancet Digit Health* 2026; **XX:** XX–XX.
2 Zhang S, Xu Y, Usuyama N, et al. A multimodal biomedical foundation model trained from fifteen million image–text pairs. *NEJM AI* 2025; **2:** AIoa2400640.
3 Zhou Y, Chia MA, Wagner SK, et al. A foundation model for generalizable disease detection from retinal images. *Nature* 2023; **622:** 156–63.

4 Jin K, Yu T, Ying GS, et al. A systematic review of vision and vision-language foundation models in ophthalmology. *Adv Ophthalmol Pract Res* 2025; **6:** 8–19.

5 Ding T, Wagner SJ, Song AH, et al. A multimodal whole-slide foundation model for pathology. *Nat Med* 2025; **31:** 3749–61.

6 Yan S, Yu Z, Primiero C, et al. A multimodal vision foundation model for clinical dermatology. *Nat Med* 2025; **31:** 2691–702.

7 Qiu J, Wu J, Wei H, et al. Development and validation of a multimodal multitask vision foundation model for generalist ophthalmic artificial intelligence. *NEJM AI* 2024; **1:** AIoa2300221.

8 Jiao J, Zhou J, Li X, et al. USFM: a universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis. *Med Image Anal* 2024; **96:** 103202.

9 Ning Y, Teixayavong S, Shang Y, et al. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. *Lancet Digit Health* 2024; **6:** e848–56.

10 Khosravi B, Purkayastha S, Erickson BJ, Trivedi HM, Gichoya JW. Exploring the potential of generative artificial intelligence in medical image synthesis: opportunities, challenges, and future directions. *Lancet Digit Health* 2025; **7:** 100890.