Check for updates

Learning from routine health system data builds better neuroimaging AI models

We trained a three-dimensional visual foundation model directly on 5.24 million routine clinical computed tomography (CT) and magnetic resonance imaging (MRI) image series so that it learned a shared representation of neuroanatomy and disease. The model demonstrated state-of-the-art diagnosis, in contrast to foundation models that are trained on public Internet and medical data, and enabled preliminary report generation and triage in real health systems.

This is a summary of:
Kondepudi, A. et al. Health system learning enables generalist neuroimaging models. *Nat. Med.* https://doi.org/10.1038/s41591-026-04497-1 (2026).

A.K. and T.H. used ChatGPT to help prepare their contribution to this Research Briefing.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 31 July 2026

The problem

General-purpose artificial intelligence (AI) models have progressed rapidly because they learn from Internet-scale text, images and video[1]. However, much of the information needed for medicine is not available in the public domain. Neuroimaging data are especially difficult to access to scale because CT and MRI image series contain identifiable anatomy and are tightly governed by data privacy protections. Existing medical foundation models have generally been trained on fragmented public datasets or institution-specific collections rather than the large, heterogeneous clinical imaging archives through which radiologists develop expertise. We therefore asked whether the continuous stream of imperfect but clinically meaningful data produced during routine care could be harnessed to train medical AI tools[2]. Neuroimaging provided a stringent test case because clinical CT and MRI data vary substantially across scanners, protocols, modalities and disease presentations, yet they support a broad range of diagnostic and prognostic tasks.

The discovery

We developed NeuroVFM, a visual foundation model trained on the UM-NeuroImages dataset, which comprises 5.24 million three-dimensional image series from 566,915 neuroimaging studies (Fig. 1a). The model was trained without diagnostic labels or paired reports using the framework Vol-JEPA (volumetric adaptation of joint-embedding predictive architectures)[3]. Instead of reconstructing pixels, Vol-JEPA predicts masked regions in a learned representation space, encouraging the encoder to learn neuroanatomy and pathology without expert annotations. We then evaluated frozen NeuroVFM representations on temporally held-out studies, probing whether the learned manifold encodes neuroanatomy that enables clinical utility. This approach culminated in a silent prospective triage feasibility study. This study tested whether a general-purpose neuroimaging representation produced by training on health system data outperforms specialist and general-purpose imaging models trained on public medical datasets or Internet-scale data.

NeuroVFM outperformed models trained with report supervision[4], voxel reconstruction or public Internet data across 82 CT diagnostic tasks (with a mean area under the receiver operating characteristic (AUROC) of 92.7%), as well as across 74 MRI diagnostic tasks (with an AUROC of 92.5%) (Fig. 1b). NeuroVFM required fewer labeled examples than other foundation models to reach equivalent performance, and learned a shared CT and MRI latent space that preserved neuroanatomy and pathology across neuroimaging protocols, sequences and people examined. When paired with the open-source language model Qwen3-14B, NeuroVFM outperformed GPT-5 and Sonnet 4.5 in preliminary report generation, with higher key finding accuracy, lower hallucination rates and fewer laterality errors in an expert-verified test set. In a silent prospective study of 1,155 consecutive clinical studies, the NeuroVFM arm achieved 92.6% balanced triage accuracy, compared with 71.2% for GPT-5 (Fig. 1c). These results suggest that direct learning from health system data provides domain-specific visual perception that might enable real-world clinical deployment to improve patient care.

The implications

The broader implication of our study is that routine health system data may serve as a source of clinically grounded knowledge for medical AI. We propose NeuroVFM as part of a paradigm that combines specialist perception with general reasoning. Although NeuroVFM may misidentify rare pathologies or miss small lesions, our findings provide initial evidence that clinically grounded imaging models are able to convert private visual data into structured outputs. Frontier large language models may then be used to reason over these outputs to support clinical workflows.

This approach may be applied to other data-rich clinical specialties[5]. We plan to extend learning of health system data across institutions and modalities, including pathology, genomics and longitudinal outcomes. The long-term goal is to develop multimodal foundation models that encode increasingly comprehensive representations of disease, supporting diagnosis, prognosis and clinical decision-making.

Akhil Kondepudi & Todd Hollon
University of Michigan, Ann Arbor, MI, USA.

EXPERT OPINION

"This is a remarkable study that demonstrates the potential of modular, foundation model-based AI platforms trained on domain-specific, real-life data. The approach clearly has immediate implications beyond neuroimaging." **Albert H. Kim, Washington University School of Medicine, St. Louis, MO, USA.**

FIGURE

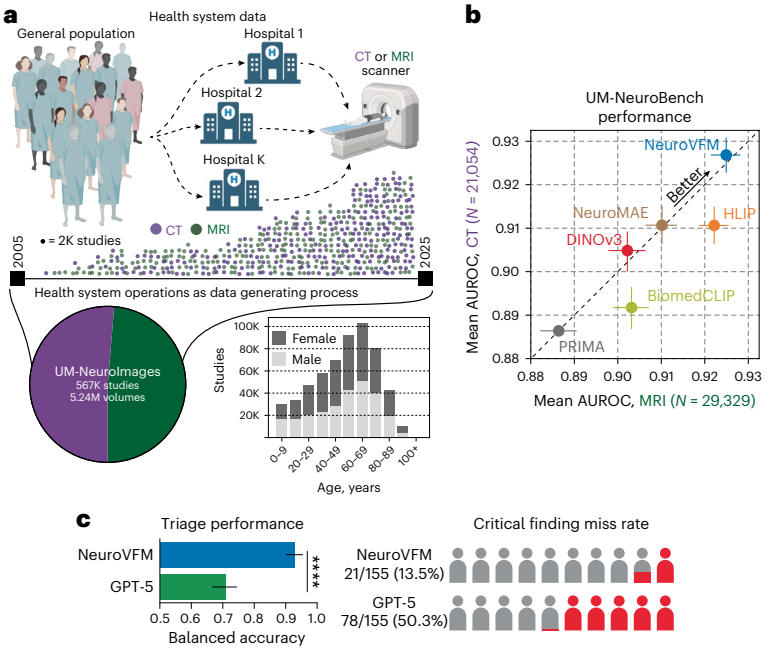

Fig. 1 | Training of NeuroVFM on routine neuroimaging data of health system archives yields improved performance. a, NeuroVFM was trained on all available clinical neuroimaging examinations in health system archives of Michigan Medicine at the University of Michigan (UM), including CT and MRI data acquired over more than 30 years. An analysis of the age and gender of people whose data were used is in the bottom right plot. **b**, On a prospective test set of more than 50,000 neuroimaging studies spanning more than 150 diagnostic tasks, NeuroVFM achieved state-of-the-art performance across both CT data and MRI data. Comparisons were made against other foundation models, such as DINOv3, NeuroMAE, HLIP, BiomedCLIP and PRIMA, and results are presented as mean AUROC values. *N*, number of studies. **c**, In a silent prospective feasibility study of clinical triage, reports based on NeuroVFM representations were used to prioritize studies by clinical urgency. Relative to reports generated by GPT-5, NeuroVFM-supported prioritization was far more accurate (left), with approximately fourfold fewer urgent cases missed (right). © 2026, Kondepudi, A. et al., CCBY 4.0.

BEHIND THE PAPER

This project grew from a simple observation: although AI systems are able to reason fluently about medicine, they rarely learn from the clinical data that define everyday practice. Building NeuroVFM required treatment of the health system as the learning environment, encompassing the full spectrum of raw, heterogeneous, and non-standardized CT and MRI acquisitions encountered in routine care. The pivotal moment came when NeuroVFM representations aligned anatomy and pathology across individuals, neuroimaging modalities and protocol sequences without explicit labels specifying these correspondences. That result changed how we thought about the project: NeuroVFM was no longer simply a diagnostic model; it provided evidence that learning on health system data is able to produce a reusable clinical representation. Rigorously proving that claim became one of the hardest parts of the study. We needed evaluations that were clinically meaningful, prospectively assessed, and fair to both specialist models and frontier models optimized mainly for two-dimensional images. **A.K. & T.H.**

REFERENCES

1. Radford, A. et al. Learning transferable visual models from natural language supervision. In *Proc. 38th International Conference on Machine Learning* 8748–8763 (2021).
   This paper established contrastive image-text pretraining (CLIP) as a foundation for modern vision-language models.
2. Jiang, L. Y. et al. Health system-scale language models are all-purpose prediction engines. *Nature* **619**, 357–362 (2023).
   This study demonstrated that large-scale health system data can support broadly useful clinical prediction models.
3. Assran, M. et al. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 15619–15629 (2023).
   This paper introduced I-JEPA, the self-supervised representation learning method adapted here for volumetric medical imaging.
4. Lyu, Y. et al. Learning neuroimaging models from health system-scale data. *Nat. Biomed. Eng.* https://doi.org/10.1038/s41551-025-01608-0 (2026).
   This paper showed strong performance with neuroimaging learning from MRI–report pairs at health system scale.
5. Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
   This perspective outlines the rationale for generalist medical AI models that learn across heterogeneous clinical data modalities and support multiple downstream tasks.

FROM THE EDITOR

"This work shows that frontier AI general models (such as OpenAI's GPT-5 and Meta's DINOv3) underperform on neuroimaging tasks and that learning directly from uncurated data generated during routine clinical care at health systems yields high-performance, generalist neuroimaging models." **Editorial Team, *Nature Medicine*.**