nature biomedical engineering

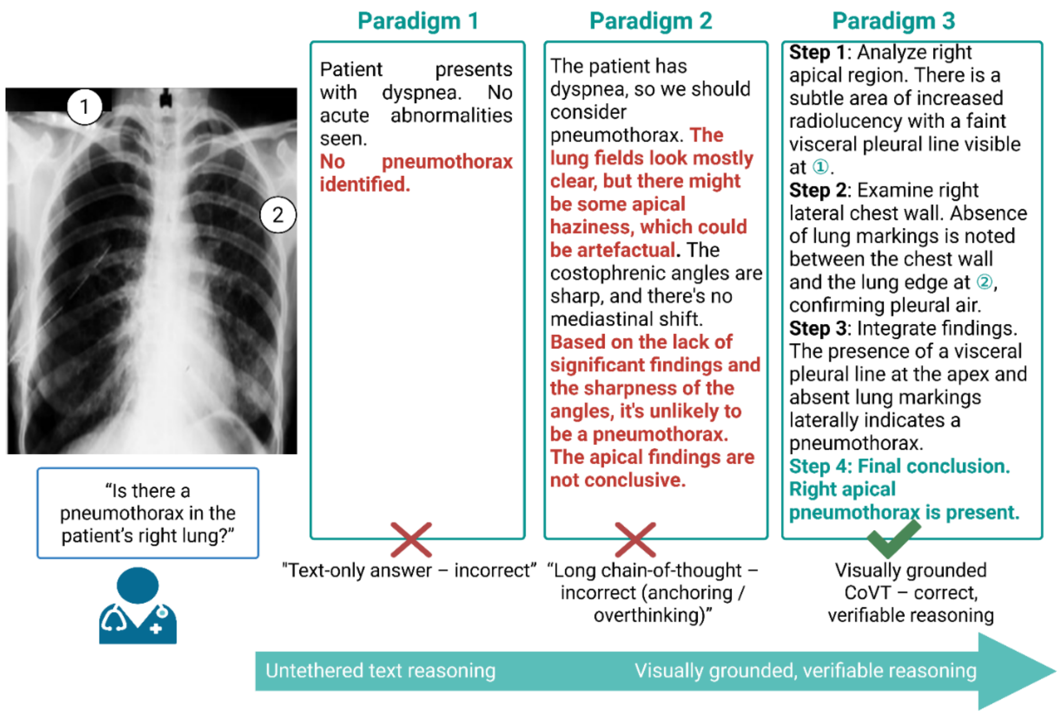
Supplementary information
https://doi.org/10.1038/s41551-026-01762-z

Foundation models in biomedical imaging: turning hype into reality

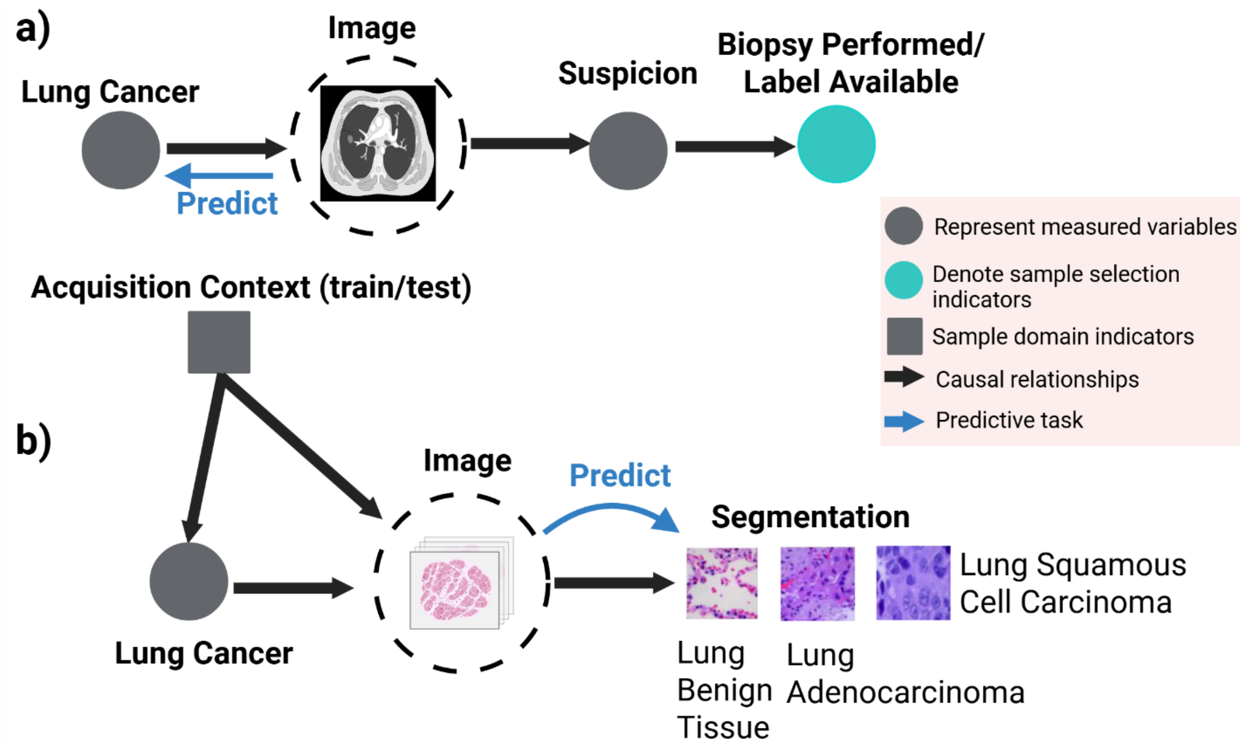
In the format provided by the authors and unedited

Supplementary Materials

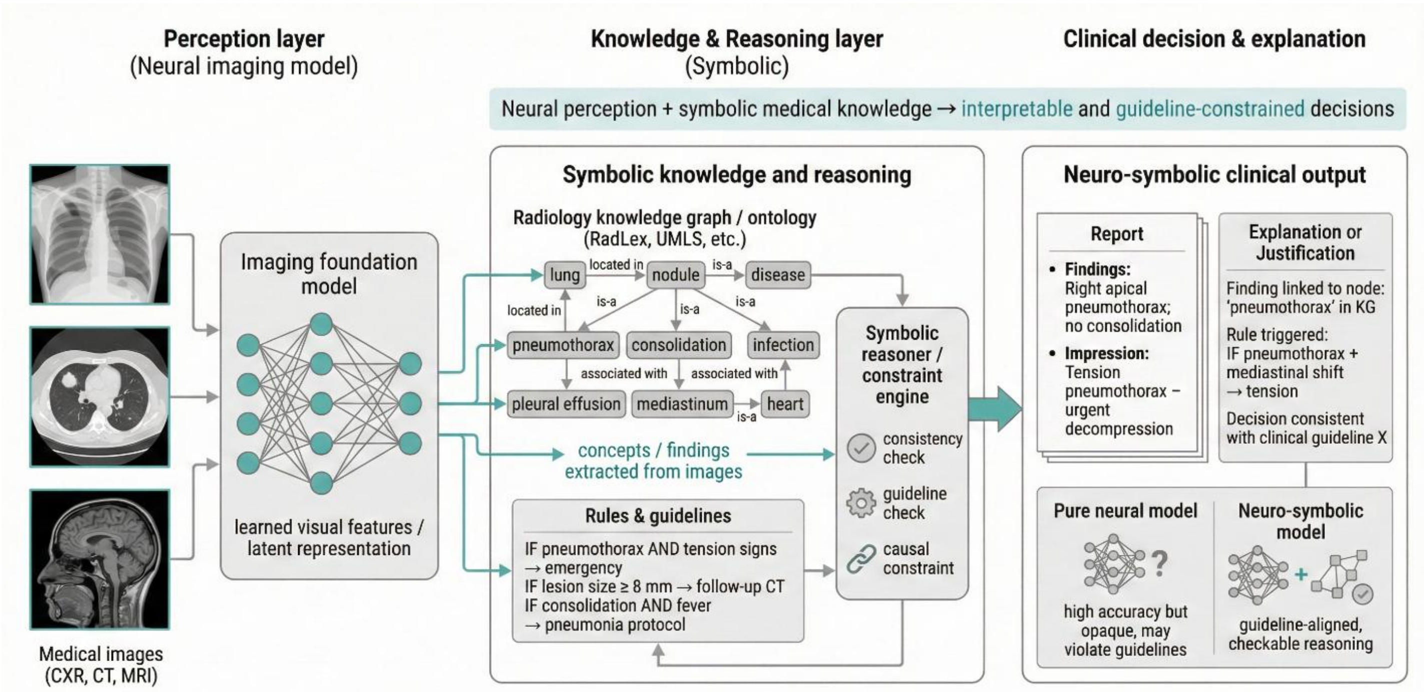
A. Figures



Suppl Fig. 1: Comparing text-only reasoning and Chain-of-Visual-Thought (CoVT) for pneumothorax detection. Given a chest radiograph and the query "Is there a pneumothorax in the patient's right lung?", a short text-only report (Paradigm 1) and a long, unconstrained chain-of-thought explanation (Paradigm 2) both arrive at the wrong conclusion due to failing to attend to key apical and lateral visual cues. In contrast, Paradigm 3 (CoVT) decomposes the decision into explicit, stepwise statements, each grounded in specific image regions, leading to the correct diagnosis of a right apical pneumothorax. The bottom bar illustrates the conceptual shift from untethered text reasoning to visually grounded, verifiable reasoning.



Suppl Fig. 2: **Causality in biomedical imaging.** (a) Lung cancer prediction (blue arrow) as a statistical inference task where the underlying disease causally drives clinical suspicion, which in turn triggers biopsy and label generation (black arrows), revealing how selection bias can confound model training. (b) Lung tumor segmentation where the acquisition context serves as a domain indicator affecting both disease manifestation and image generation. Filled circles represent measured variables, cyan circles denote sample selection indicators, and gray squares represent sample domain indicators. Blue arrows indicate the direction of the predictive task, while black arrows represent causal relationships.



Suppl Fig. 3: Neuro-symbolic AI combines imaging foundation models with ontologies, rules, and logical reasoning to improve robustness, causal consistency, and clinical trust.

B. Supplementary Information

Supplementary Note 1. To improve the auditability of radiology report generation, CoVT-CXR was developed using a carefully annotated chest X-ray dataset in which reasoning steps in the report are aligned with specific image regions. This design encourages the model to learn a direct association between semantic claims, such as the presence of cardiomegaly, and the image evidence supporting those claims. As a result, the reasoning process becomes not only more interpretable but also visually auditable.

Supplementary Note 2. The Radiology Knowledge Graph system (ReXKG) extends evaluation from the surface text level to the knowledge level.[1] It works by first extracting structured information (entities and their relationships) from an AI-generated report to construct a KG. This generated KG is then compared against a ground-truth KG derived from a human expert's report. Using novel metrics that assess the similarity of nodes (entities), edges (relationships), and subgraphs, ReXKG can evaluate whether the AI model has correctly understood not just the key findings but also their clinical context and interrelations, a much deeper and more meaningful assessment of its reasoning capabilities.[1]

Supplementary Note 3. Neuro-symbolic AI (NSAI) represents an emerging strategy for combining data-driven learning with explicit knowledge-based reasoning[2]. This hybrid approach can help address the black box problem by combining the perceptual strengths of neural networks with the explicit structure of symbolic AI. In a typical NSAI architecture for medical imaging, a neural network component acts as a perception engine, analyzing raw image pixels to identify objects and features. This structured information is then passed to a symbolic reasoning engine, which operates on formal logic, ontologies, and rule-based systems to validate the findings, draw inferences, and construct a transparent, auditable explanation for the final output or recommendation. This hybrid direction may combine the perceptual strengths of neural models with the transparency and constraint-based reasoning of symbolic systems. This trajectory suggests that for FMs to become more trustworthy tools in medicine, they cannot rely on emergent intelligence from scale alone; they must be supported by explicit reasoning structures, verifiable knowledge, and auditable constraints.

References

1. Zhang, X., Acosta, J. N., Zhou, H.-Y. & Rajpurkar, P. Uncovering knowledge gaps in radiology report generation models through knowledge graphs. *arXiv [cs.AI]* (2024).
2. Bhuyan, B. P., Ramdane-Cherif, A., Tomar, R. & Singh, T. P. Neuro-symbolic artificial intelligence: a survey. *Neural Comput. Appl.* **36**, 12809–12844 (2024).