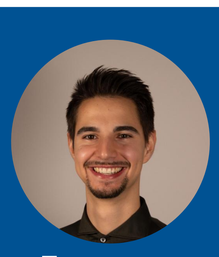
Linfeng Guo [1]



Francesco Vaccaro [1]

A Comparative Study of VLMs for Medical Image Analysis: CheXagent and MAIRA-2

[1] Technical University of Munich || Contact: Linfeng.Guo@tum.de or Francesco.Vaccaro@tum.de

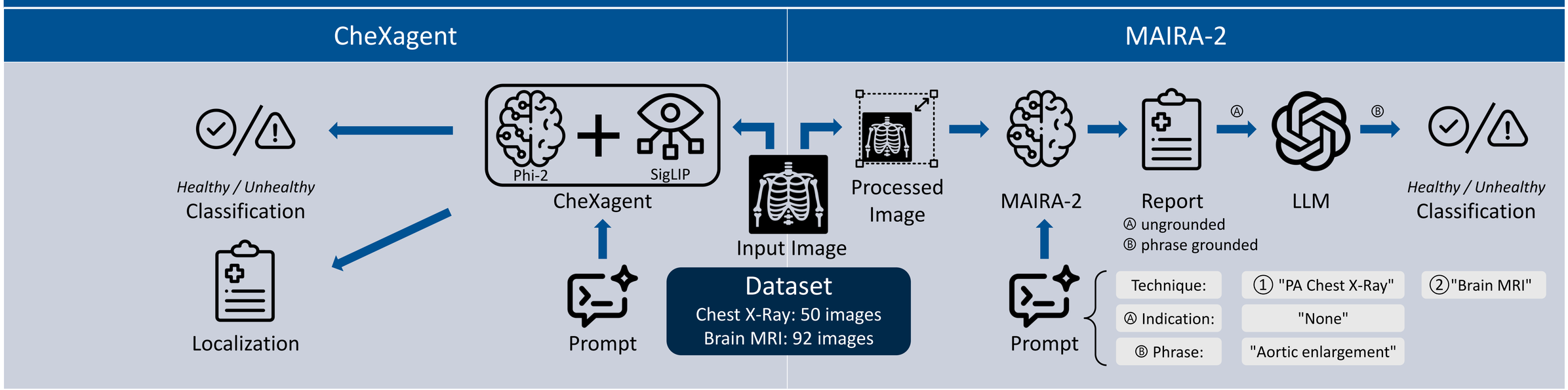
Workforce gaps & report quality – challenges in radiology

- Radiology faces workforce shortage in several countries
- Consistently high-quality radiology reports are critical in healthcare
- Traditional AI models focus on single domains and tasks
- VLMs offer multi-modal capabilities, enabling full-report generation
- Goal: Evaluate recently developed VLMs on radiology tasks to understand their practical utility

Model Comparsion

CheXagent	MAIRA-2
3.1 billion parameters	6.9 billion parameters
CheXinstruct	Training Data
8.5 million samples	0.3 million samples
Published	Published
January 2024	June 2024

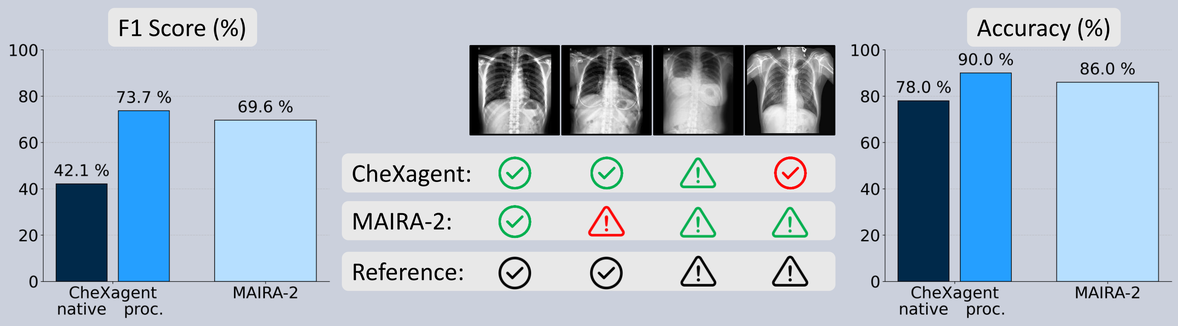
Method



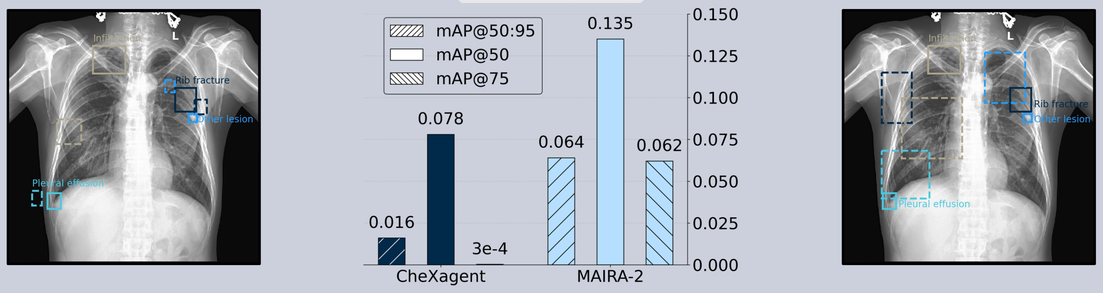
Results & Evaluation

Chest X-Ray

Classification: Healthy vs. Unhealthy

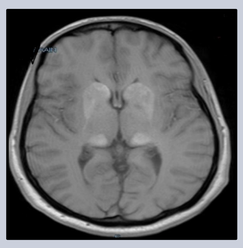


Localization: given Disease



Brain MRI

Reference: T1 sequence showing high signal in basal ganglia bilaterally



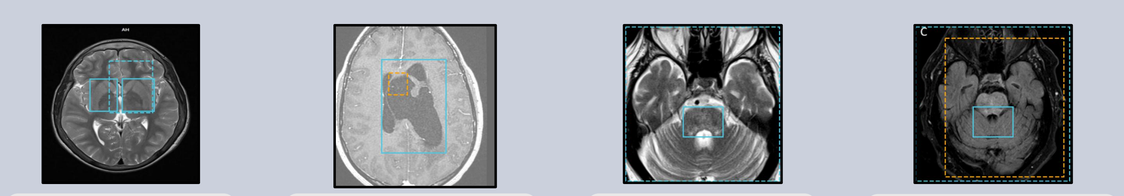
Description Generation

CheXagent: **Chest X-ray** showing bilateral hyperdense bones.

MAIRA-2: A right **chest tube** has been placed. The right **pneumothorax** has resolved. **Lungs** are hyperexpanded. **Heart** size is normal.

Both models fail at cross-domain generalization

Localization: all abnormalities



Conclusion

- Strong in-domain performance for classification tasks
- Grounding tasks still require improvements for possible clinical applications
- Cross-domain generalization with brain MRI completely fails

Selected References

[1] S. Bannur *et al.*, "MAIRA-2: Grounded Radiology Report Generation," Sep. 20, 2024, *arXiv*: arXiv:2406.04449. doi: 10.48550/arXiv.2406.04449.
[2] Z. Chen *et al.*, "A Vision-Language Foundation Model to Enhance Efficiency of Chest X-ray Interpretation," Dec. 18, 2024, *arXiv*: arXiv:2401.12208. doi: 10.48550/arXiv.2401.12208..
[3] H. Q. Nguyen et al., "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations," *Scientific Data*, vol. 9, no. 1, p. 429, 2022.
[4] C. I. Bercea et al., "NOVA: A Benchmark for Anomaly Localization and Clinical Reasoning in Brain MRI," *arXiv preprint* arXiv:2505.14064, May 2025.

CompAI
Computational Imaging and AI in Medicine