

Benchmarking CXR embedding models for disease classification

Jiho Shin, Dominic Marshall, YingYing Fang

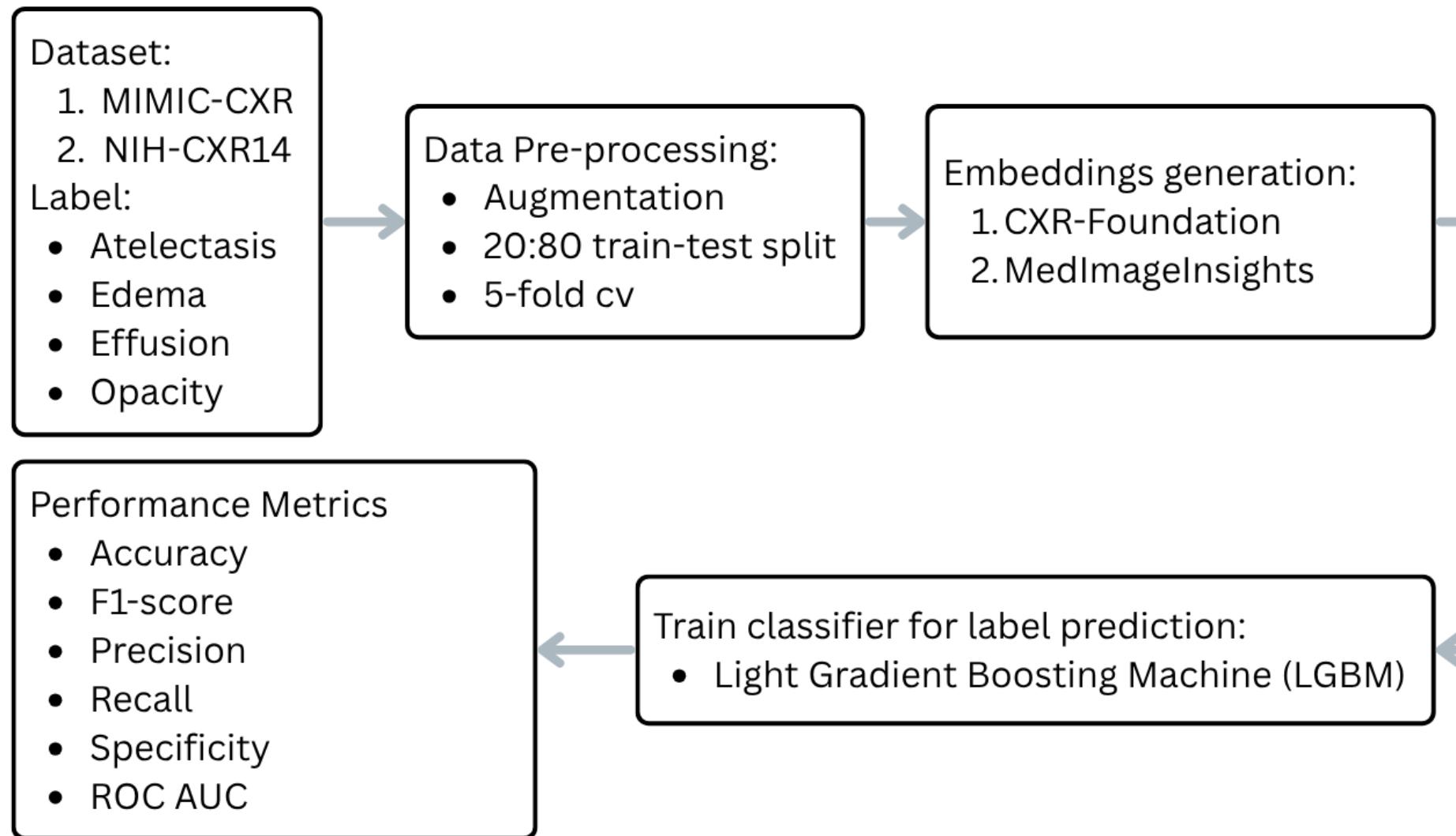
The bigger picture: Master Project

- Problem: Acute Respiratory Distress Syndrome (ARDS)
 - Heterogeneous, making diagnosis, prognosis and personalised treatment challenging
- Solution: Machine Learning – Deep Representation Learning for Clinical Clustering (ML-DRLCC)
 - A multimodal deep learning framework aimed at identifying clinically meaningful subtypes in ARDS

Motivation

- Benchmarking CXR embedding models to select the optimal CXR image representation for integration into ML-DRLCC
- Potential of CXR embedding models for patient representation

Workflow



Dataset

Open-source datasets:

1. MIMIC CXR
2. NIH-CXR14

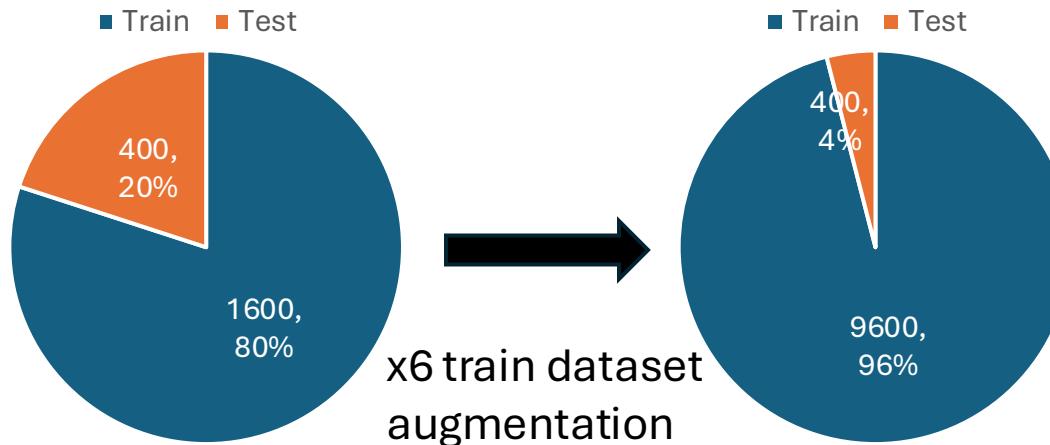
Disease labels

1. Opacity
2. Atelectasis
3. Edema
4. Effusion

For each disease label and dataset (MIMIC, NIH):

- 0: Absence of disease label → 1000 images
- 1: Presence of disease label → 1000 images

Data preprocessing and train-test split



x6 train dataset augmentation

- Augmentation



Original

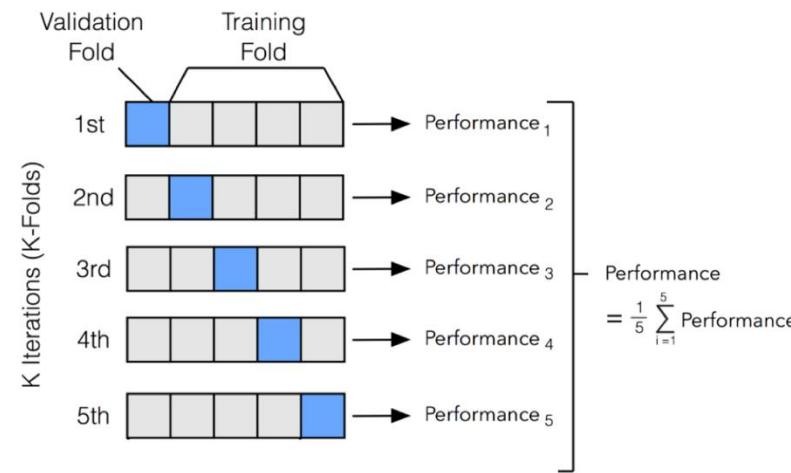


Bright

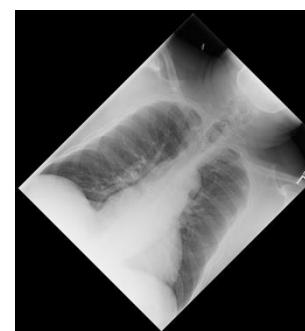


Dark

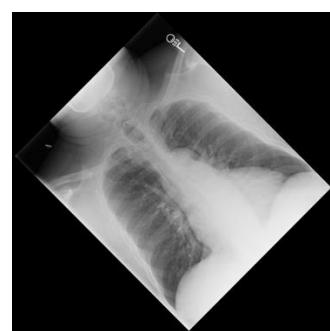
- K=5 fold cross-validation:



Contrast



Rot +45 deg



Rot -45 deg

Model 1: CXR-Foundation (2023)

Overview & Features

- ELIXR v2.0: rich 32×768 -dimensional vectors.

Benchmark Results

- Delivers **high AUC (89% for tube placement assessment)** on classification tasks and strong retrieval performance

Limitations

- Trained on US/India data only; may not generalise globally.
- Large data dimension (32x768)

Model 2: MedImageInsight (2024)

Overview & Features

- Embeddings: **1,024 vectors** used for similarity indexing

Benchmark Results

- On a radiographic tube-placement classification dataset (8,842 images), embeddings achieved **mean AUC of 93.8 %, vs. 89 % (CXR-Foundation)**

Strengths

- Bone-age estimation and other tasks achieved $AUC > 0.9$.
- Efficient and adaptable to lightweight classifiers.
- Compact representation (**1024 [MedImageInsight] vs 32x768 [CXR-foundation]**)

Results: Accuracy (95% CI: \pm error)

		MedImageInsights	CXR-foundation
MIMIC	Atelectasis	0.75 ± 0.0073	0.75 ± 0.0075
	Edema	0.84 ± 0.014	0.85 ± 0.024
	Effusion	0.90 ± 0.026	0.88 ± 0.0098
	Opacity	0.70 ± 0.016	0.70 ± 0.023
Average		0.80 ± 0.0087	0.80 ± 0.0088
NIH-14	Atelectasis	0.78 ± 0.016	0.75 ± 0.023
	Edema	0.85 ± 0.019	0.83 ± 0.013
	Effusion	0.83 ± 0.014	0.83 ± 0.0080
	Opacity	0.85 ± 0.019	0.89 ± 0.013
Average		0.83 ± 0.0087	0.82 ± 0.0076

Results: AUROC (95% CI: \pm error)

		MedImageInsights	CXR-foundation
MIMIC	Atelectasis	0.83 ± 0.0069	0.82 ± 0.013
	Edema	0.92 ± 0.011	0.92 ± 0.014
	Effusion	0.96 ± 0.012	0.94 ± 0.0060
	Opacity	0.78 ± 0.019	0.77 ± 0.017
Average		0.87 ± 0.0065	0.86 ± 0.0066
NIH-14	Atelectasis	0.86 ± 0.0079	0.82 ± 0.012
	Edema	0.92 ± 0.012	0.91 ± 0.015
	Effusion	0.90 ± 0.011	0.90 ± 0.011
	Opacity	0.92 ± 0.012	0.96 ± 0.010
Average		0.90 ± 0.0054	0.90 ± 0.0061

Results

Accuracy (95% CI: \pm error)

		MedImageInsights	CXR-foundation
MIMIC	Atelectasis	0.75 ± 0.0073	0.75 ± 0.0075
	Edema	0.84 ± 0.014	0.85 ± 0.024
	Effusion	0.90 ± 0.026	0.88 ± 0.0098
	Opacity	0.70 ± 0.016	0.70 ± 0.023
	Average	0.80 ± 0.0087	0.80 ± 0.0088
NIH-14	Atelectasis	0.78 ± 0.016	0.75 ± 0.023
	Edema	0.85 ± 0.019	0.83 ± 0.013
	Effusion	0.83 ± 0.014	0.83 ± 0.0080
	Opacity	0.85 ± 0.019	0.89 ± 0.013
	Average	0.83 ± 0.0087	0.82 ± 0.0076

AUROC (95% CI: \pm error)

		MedImageInsights	CXR-foundation
MIMIC	Atelectasis	0.83 ± 0.0069	0.82 ± 0.013
	Edema	0.92 ± 0.011	0.92 ± 0.014
	Effusion	0.96 ± 0.012	0.94 ± 0.0060
	Opacity	0.78 ± 0.019	0.77 ± 0.017
	Average	0.87 ± 0.0065	0.86 ± 0.0066
NIH-14	Atelectasis	0.86 ± 0.0079	0.82 ± 0.012
	Edema	0.92 ± 0.012	0.91 ± 0.015
	Effusion	0.90 ± 0.011	0.90 ± 0.011
	Opacity	0.92 ± 0.012	0.96 ± 0.010
	Average	0.90 ± 0.0054	0.90 ± 0.0061

- Two models exhibit almost identical performance while MedImageInsight model slightly outperforms CXR foundation in general
- Similar performance metrics across disease labels (within ± 0.04 difference), while they differ significantly across different disease labels

Discussion

- The MedImageInsight model might be the ideal option as it has a smaller embedding dimension (1024) compared to the CXR Foundation (32x768), with similar performance.
 - Low computational cost
 - Theoretically, more advantageous for clustering (lower noise level)
 - Lower risk of overfitting
- Model architecture and feature extraction method may not be the limiting factor — the nature of the disease label itself may drive performance differences.

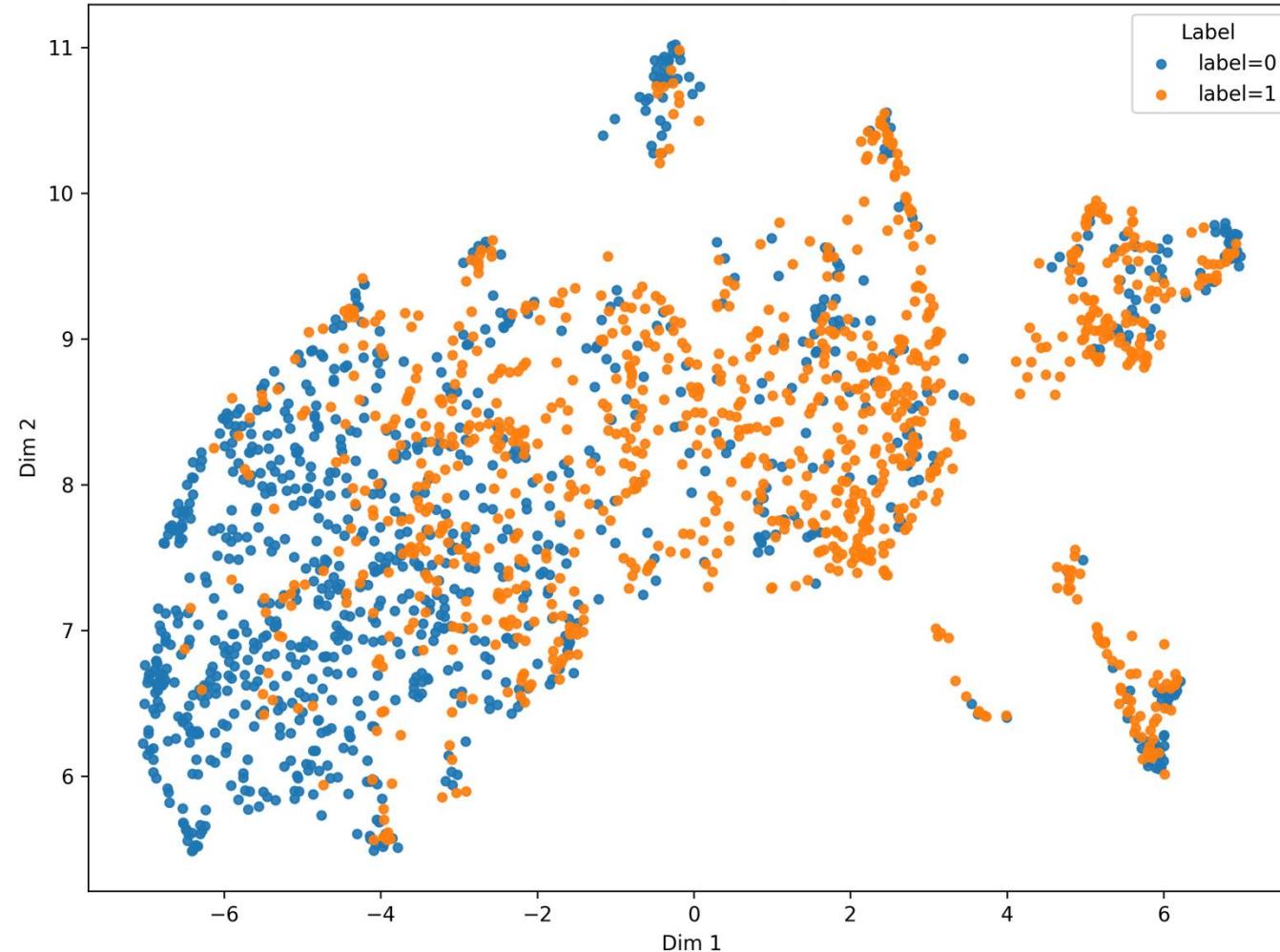
Clustering: MedImageInsight & MIMIC

1. **K-Nearest Neighbour (KNN)** from CXR embeddings (cosine similarity, k=15).
2. Applied **Leiden community detection** to find patient subgroups in the embedding space.
3. Dimensionality reduced with **UMAP** for visualization

Clustering

MedImageInsights/MIMIC/Atelectasis

Patient Embeddings colored by Label (UMAP)

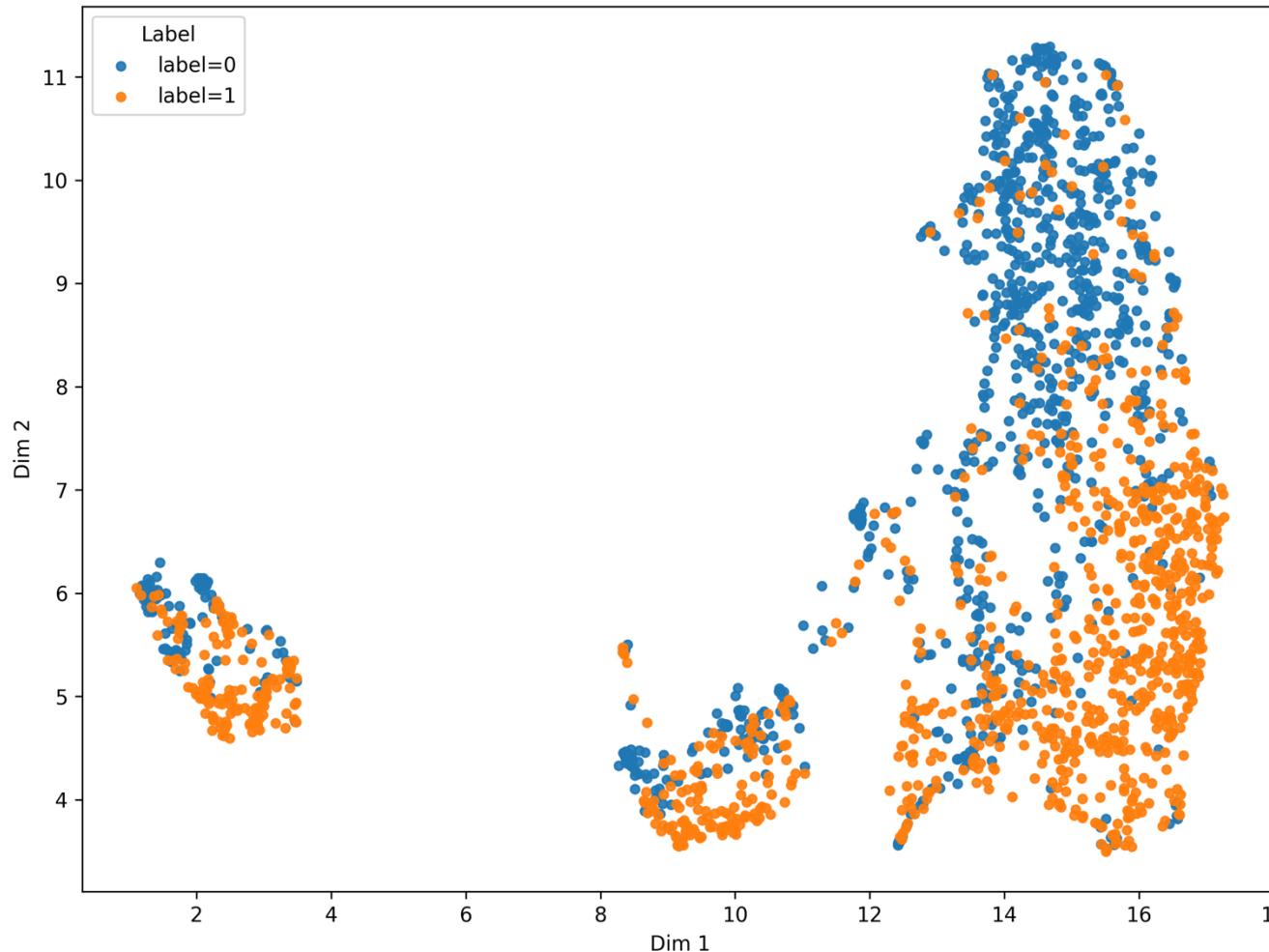


		MedImageInsights
MIMIC	Atelectasis	0.75 ± 0.0073
	Edema	0.84 ± 0.014
	Effusion	0.90 ± 0.026
	Opacity	0.70 ± 0.016

Clustering

MedImageInsights/MIMIC/Edema

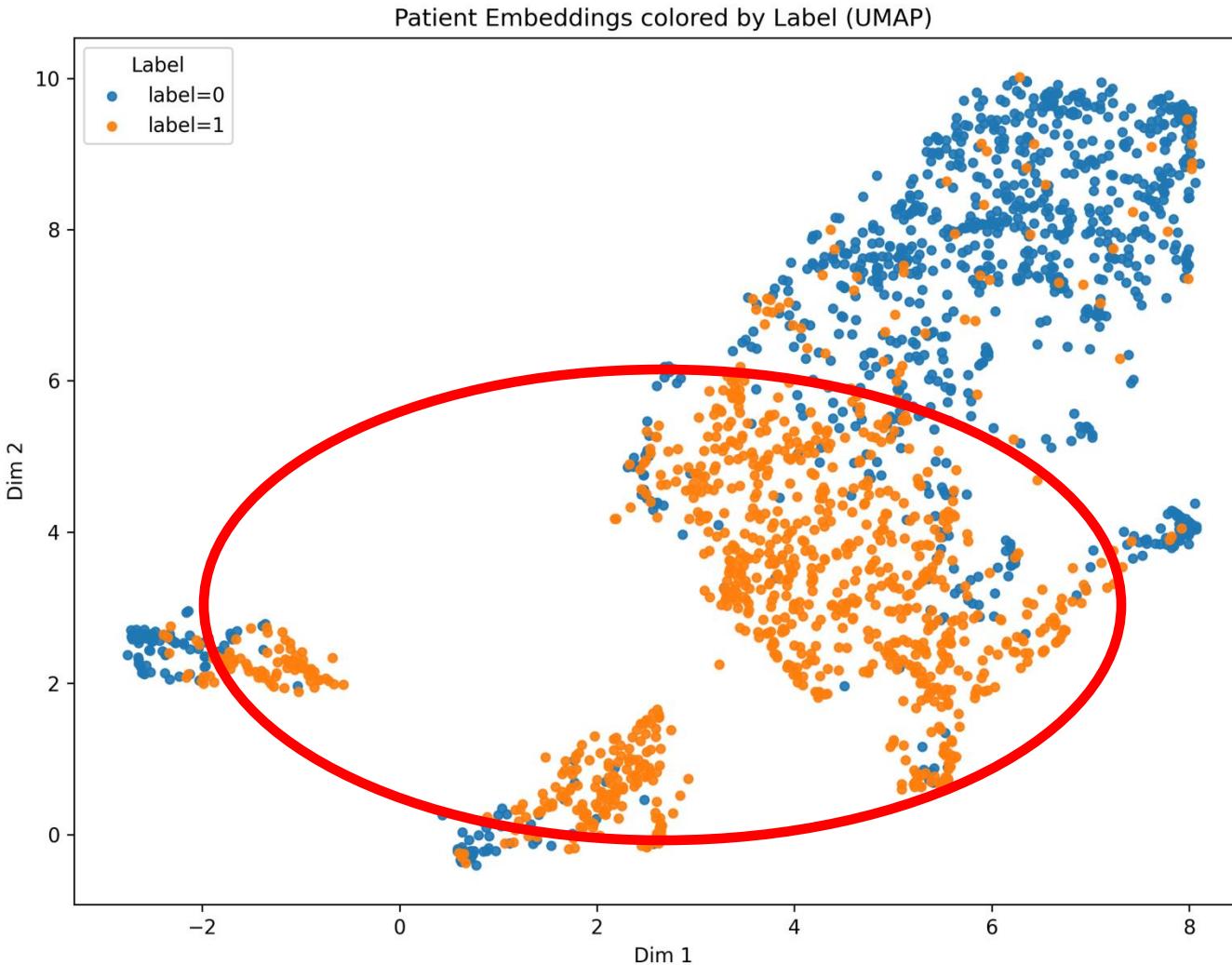
Patient Embeddings colored by Label (UMAP)



		MedImageInsights
MIMIC	Atelectasis	0.75 ± 0.0073
	Edema	0.84 ± 0.014
	Effusion	0.90 ± 0.026
	Opacity	0.70 ± 0.016

Clustering

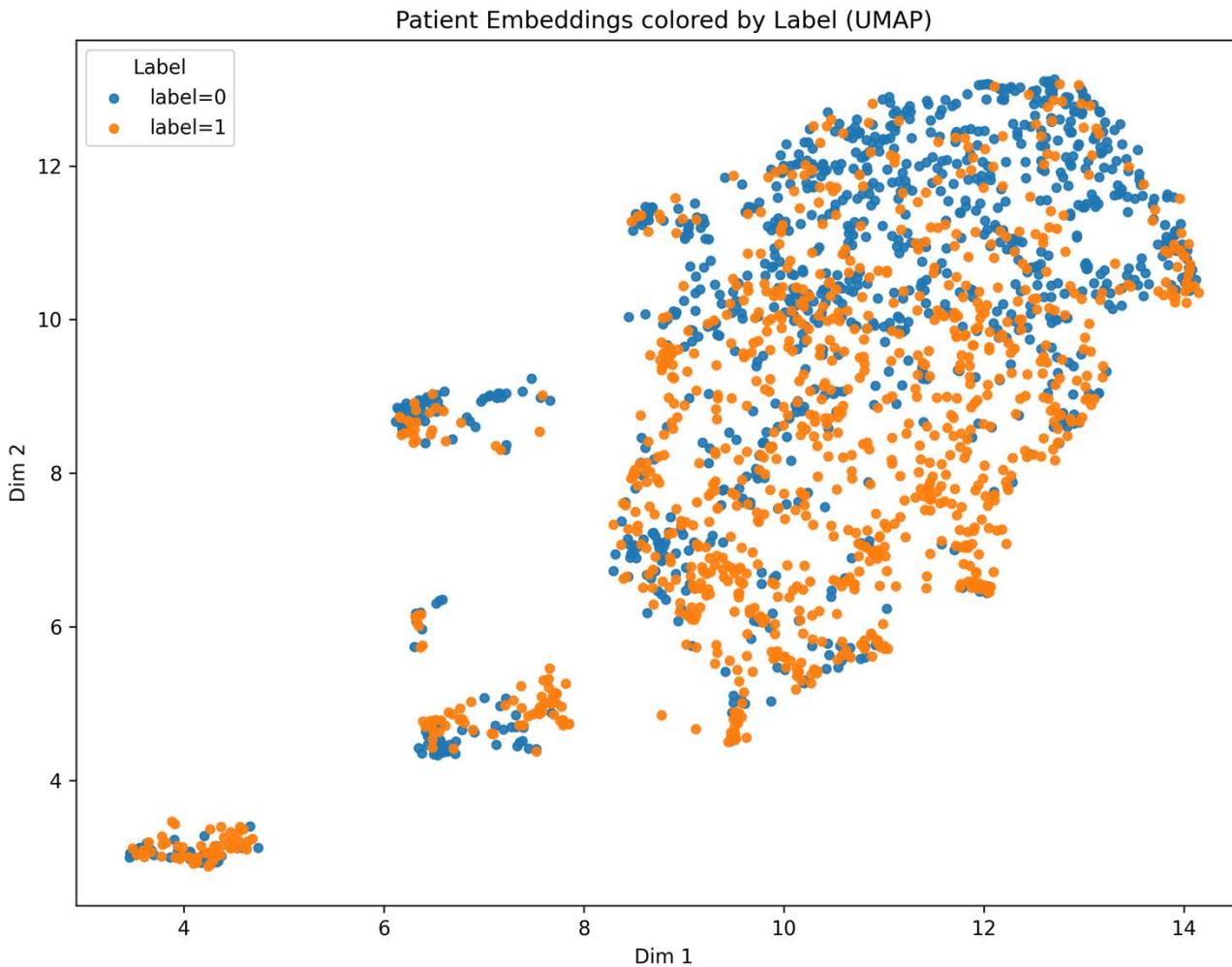
MedImageInsights/MIMIC/Effusion



		MedImageInsights
MIMIC	Atelectasis	0.75 ± 0.0073
	Edema	0.84 ± 0.014
	Effusion	0.90 ± 0.026
	Opacity	0.70 ± 0.016

Clustering

MedImageInsights/MIMIC/Opacity



	MedImageInsights
MIMIC	Atelectasis
	0.75 ± 0.0073
	Edema
	0.84 ± 0.014
	Effusion
	0.90 ± 0.026
	Opacity
	0.70 ± 0.016

Model architecture and feature extraction method may not be the limiting factor

- Integration with clinical data?
- Improve training dataset quality (e.g. Better augmentation)?

Future Steps

- Use cxr embedding model to fuse with clinical data to find meaningful patient clusters for Acute Respiratory Distress Syndrome (ARDS)
- Investigate how to combine clinical data with cxr embeddings
 - Variational Auto encoders (VAEs)
 - M4survive

arXiv > cs > arXiv:2503.10057

Search...
Help | Ad

Computer Science > Computer Vision and Pattern Recognition
[Submitted on 13 Mar 2025]

Multi-Modal Mamba Modeling for Survival Prediction (M4Survive): Adapting Joint Foundation Model Representations

Ho Hin Lee, Alberto Santamaría-Pang, Jameson Merkov, Matthew Lungren, Ivan Tarapov

