Leveraging Foundation Model Embeddings in Adapter Training for Radiography Classification

Xue Li, PhD Candidate, Student, Electrical and Computer Engineering, University of Wisconsin-Madison
Jameson Merkow, PhD; Noel Codella, PhD; Alberto Pang, PhD; Naiteek Sangani, MS; Alexander Ersoy, MS; Christopher Burt; John Garrett, PhD; Richard Bruce, MD; Joshua Warner, MD, PhD, CIIP; Tyler Bradshaw, PhD; Ivan Tarapov, MS; Matthew Lungren, MD; Alan McM

Introduction

Foundation models, pre-trained on vast datasets, have revolutionized machine learning by providing transferable embeddings for diverse downstream applications, like image classification, search, and report generation. In this study, we utilize embeddings from both general-purpose and medical domain-specific foundation models to train adapters for the multi-class classification of radiography images, with a focus on tube placement to enhance diagnostic accuracy and efficiency.

Hypothesis

Foundation model embeddings can be effectively utilized to train lightweight adapter models for multi-class classification of radiography images.

Methods

8,842 radiographs were labeled across seven categories. They were rescaled into a [0, 255] intensity range, and then fed into six foundation models for embedding extraction: DenseNet121, BiomedCLIP, Med-Flamingo, MedImageInsight, Rad-DINO, and CXR-Foundation model. Adapters were trained on the training dataset using traditional machine learning models, including K-Nearest Neighbors (KNN), logistic regression (LR), Support Vector Machines (SVM), random forest (RF), and Multi-Layer Perceptron (MLP). Model parameters were optimized based on performance on the validation dataset, and the trained adapters were subsequently evaluated on the test dataset to assess classification performance.

Results

Mean area under the curve (mAUC) metrics and computational efficiency were computed for each foundation model embedding paired with various adapter models, as summarized in Figure 1 and Table 1, respectively. MedImageInsight achieved the highest mAUC of 93.85% with SVM while Med-Flamingo embeddings performed the worst, peaking at 77.49% with RF. Rad-DINO and CXR-Foundation model embeddings delivered strong results, achieving mAUC values of 91.12% and 89.02%, respectively, both with SVM. DenseNet121 and BiomedCLIP showed moderate performance, with mAUCs of 81.84% and 83.04%. Training and inference times for all adapters were within seconds, except for SVM and RF, which required under one minute for training.

Conclusion

Foundation model embeddings, like MedImageInsight, can effectively train lightweight adapter models for multi-class radiography classification, achieving high accuracy and computational efficiency supporting practical deployment.

Figure(s)

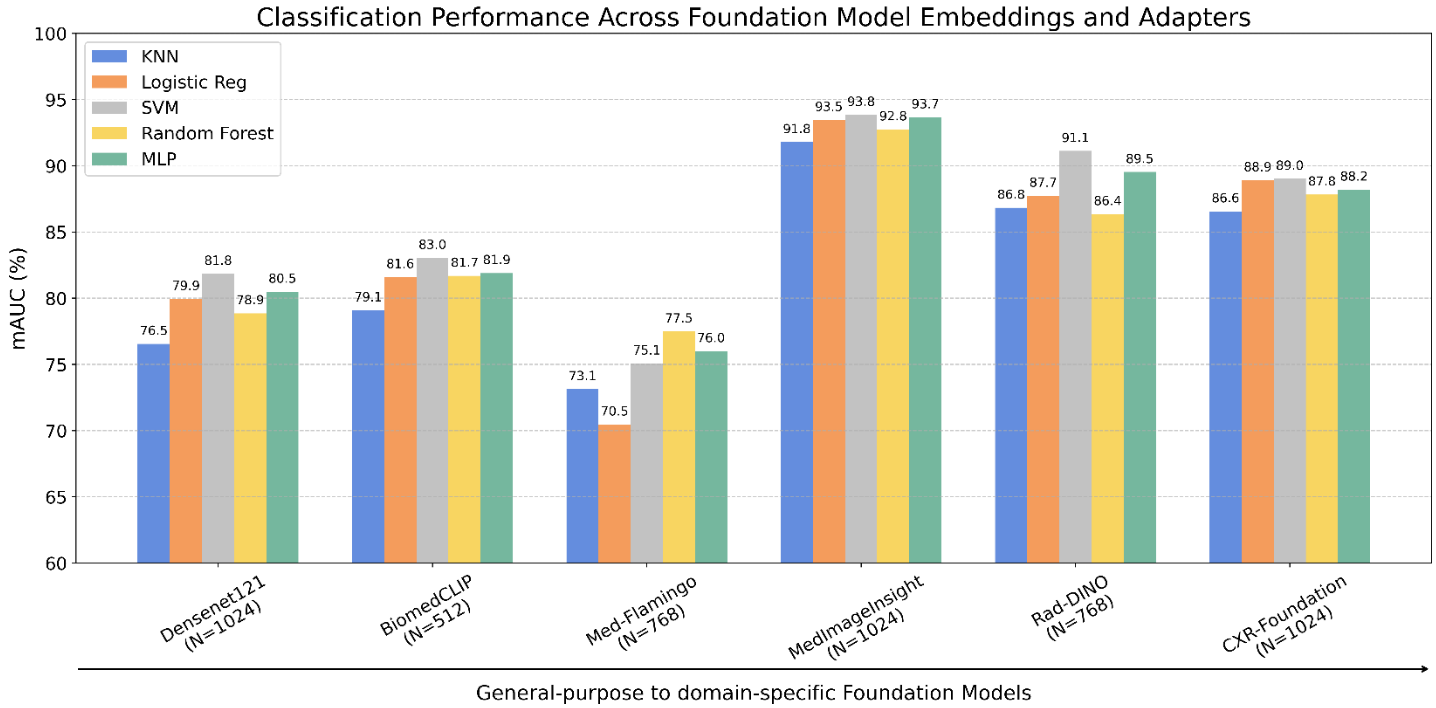

Figure 1. Performance comparison of adapter models (KNN, logistic regression, SVM, random forest, and MLP) paired with embeddings from general-purpose to domain-specific foundation models, evaluated by mAUC across seven classes including normal, endotracheal tube, nasogastric tube, Pneumoperitoneum, pneumothorax, rib fracture, and vascular lines. 'N' denotes the embedding size used for adapter training.

Adapters	KNN	LR	SVM	RF	MLP
Training	0.004	5.717	41.555	32.361	5.147
Inference	0.112	0.003	1.909	0.096	0.006

* KNN= K-Nearest Neighbors; LR=Logistic Regression; SVM=Support Vector Machines; RF=Random Forest; MLP=Multi-Layer Perceptron

Table 1. Average training and inference times (in seconds) on CPU for each adapter model

Keywords

Artificial Intelligence/Machine Learning; Imaging Research