Stabilized Progressive Fine-Tuning (SPFT)

Vedant Jayant Padole

ASU School of Computing and Augmented Intelligence

DOI: https://doi.org/10.51584/IJRIAS.2026.11060121

Received: 18 May 2026; Accepted: 23 May 2026; Published: 29 June 2026

ABSTRACT

Foundation models have demonstrated strong performance in medical imaging; however, their ability to generalize across heterogeneous datasets, imaging modalities, and downstream tasks remains limited. Existing approaches often rely on modality-specific architectures, task-specific training pipelines, or large-scale retraining, which hinder scalability and practical deployment in real-world clinical settings.

In this work, we present a unified and reproducible optimization-driven framework for improving cross-dataset, cross-modality, and cross-task generalization of foundation models in medical imaging. Rather than introducing new architectures, we investigate how principled optimization strategies can enhance performance, stability, and transferability across diverse vision tasks. Building on the observation that pretrained encoders capture modality-agnostic representations, we propose a lightweight adaptation strategy, termed Stabilized Progressive Fine-Tuning (SPFT), which combines staged fine-tuning, progressive layer unfreezing, class-aware loss weighting, and Exponential Moving Average (EMA) stabilization.

We evaluate our approach across multiple datasets and tasks, including chest radiograph classification on ChestX-ray14 (Wang et al. 2017) and CheXpert (Irvin et al. 2019), dermoscopic image classification on HAM10000 (Tschandl et al. 2018), object detection on VinDr-CXR, and medical image segmentation using SAM-based models. Importantly, the same SPFT strategy is applied consistently across all tasks and architectures, including transformer-based and CNN-based models.

Experimental results demonstrate that our approach achieves strong and stable performance across classification (AUC up to 0.95), detection (mAP@50 up to 0.76), and segmentation (Dice up to 0.96), while maintaining low variance across runs. We further show that optimization plays a critical role in improving generalization and stability, even when architectural choices vary significantly.

These findings highlight that carefully designed optimization strategies can serve as a scalable and effective alternative to task-specific architectural modifications, enabling unified and robust medical imaging systems across diverse datasets and problem settings.

INTRODUCTION

Medical imaging has undergone rapid transformation with the emergence of large-scale foundation models, which leverage self-supervised learning to capture rich and transferable visual representations. Recent approaches such as RAD-DINO (Anonymous 2023) have demonstrated strong performance on chest X-ray classification tasks, achieving competitive results on widely used benchmarks such as ChestX-ray14 and CheXpert. However, despite these advances, effectively adapting foundation models across heterogeneous datasets, imaging modalities, and downstream tasks remains a significant challenge.

A central issue lies in the heterogeneity of medical data. Imaging modalities such as chest radiographs (grayscale) and dermatoscopic images (RGB) differ substantially in visual characteristics, label distributions, and noise patterns. Furthermore, even within the same modality, datasets such as ChestX-ray14 (Wang et al. 2017), CheXpert (Irvin et al. 2019), and VinDr-CXR exhibit notable distributional differences in annotation quality, label sparsity, and clinical variability. As a result, models trained on one dataset or task often fail to generalize

to others without substantial architectural modifications, task-specific pipelines, or large-scale retraining. While effective, such approaches increase computational cost and limit scalability in real-world clinical applications.

In this work, we investigate whether a unified training strategy can enable cross-dataset, cross-modality, and cross-task generalization of foundation models without modifying the underlying architecture. We hypothesize that pretrained encoders capture modality-agnostic representations, and that the primary challenge lies in adapting these representations through effective optimization rather than increasing model or pipeline complexity. Based on this insight, we propose a unified adaptation framework, termed Stabilized Progressive Fine-Tuning (SPFT), which combines staged fine-tuning, progressive layer unfreezing, class-aware optimization, and Exponential Moving Average (EMA) stabilization to improve robustness and generalization.

We evaluate our approach across multiple datasets and tasks spanning distinct medical imaging domains, including chest radiograph classification on ChestX-ray14 (Wang et al. 2017) and CheXpert (Irvin et al. 2019), dermoscopic image classification on HAM10000 (Tschandl et al. 2018), object detection on VinDr-CXR, and medical image segmentation using SAM-based models. These datasets and tasks exhibit substantial variation in modality, label structure, annotation granularity, and data distribution, providing a rigorous testbed for evaluating unified generalization.

Experimental results demonstrate consistent improvements over baseline models across classification, detection, and segmentation tasks, achieving strong performance while maintaining low variance across runs. Importantly, the same optimization strategy is applied across all models and tasks, without requiring task-specific architectural modifications or complex augmentation pipelines.

The contributions of this work are summarized as follows:

- We propose a unified optimization-driven framework (SPFT) for adapting foundation models across heterogeneous medical imaging datasets, modalities, and tasks without architectural modifications.
- We demonstrate strong cross-dataset, cross-modality, and cross-task generalization across classification (ChestX-ray14, CheXpert, HAM10000), detection (VinDr-CXR), and segmentation tasks, using a single training strategy.
- We show that principled optimization strategies can reduce reliance on task-specific pipelines and architectural complexity, highlighting the critical role of optimization in medical image modeling.
- We provide empirical insights into how training dynamics influence transferability, spatial reasoning, and stability across diverse medical imaging tasks and datasets.

METHODOLOGY

Backbone

We propose a unified optimization-driven framework for adapting foundation models across heterogeneous medical imaging modalities. Our approach, termed **Stabilized Progressive Fine-Tuning (SPFT)**, focuses on improving generalization through training strategy design rather than architectural modifications or extensive data augmentation.

We build our approach on top of RAD-DINO (Anonymous 2023), a vision foundation model trained using self-supervised learning. Such models have been shown to learn rich and transferable representations that generalize well across downstream tasks (Caron et al. 2021; Oquab et al. 2023). In the context of medical imaging, RAD-DINO has demonstrated strong performance on chest X-ray classification benchmarks, making it a suitable choice for feature extraction.

Instead of training a model from scratch, we leverage the pretrained encoder to benefit from its learned visual representations, thereby reducing computational cost and improving generalization.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Architecture

We build our model on top of the RAD-DINO encoder, followed by a feature aggregation module and a task-specific multi-layer perceptron (MLP) classification head. Given an input image x, the encoder produces token embeddings:

$$\boldsymbol{H} = \text{Encoder}(x) \in \mathbb{R}^{N \times D}$$

where N is the number of tokens and $D = 768$ is the embedding dimension.

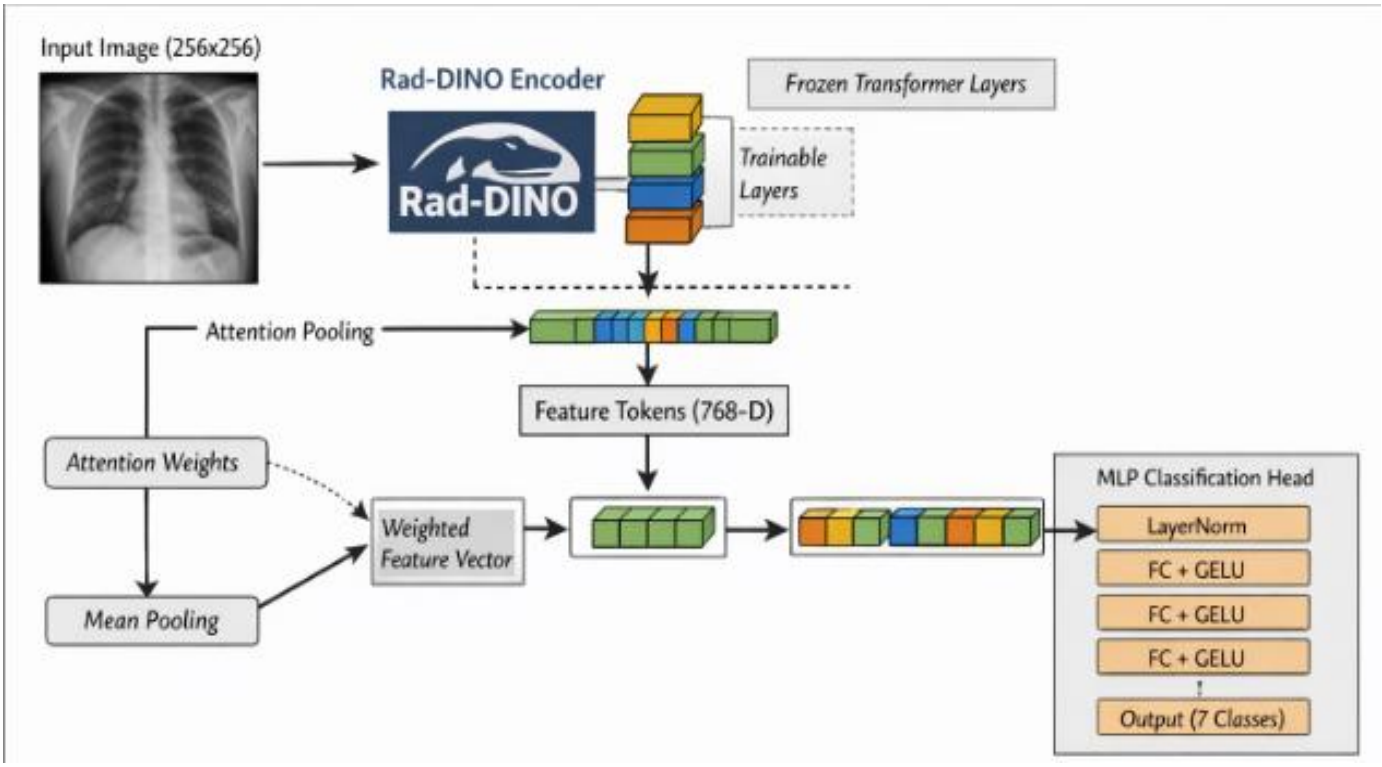
Unlike standard transformer-based approaches that rely on the [CLS] token, we aggregate spatial information using global average pooling over patch tokens after removing the [CLS] token. This design encourages the model to capture distributed spatial cues, which is particularly important in medical imaging where pathological patterns are often diffuse and not confined to a single region:

$$\boldsymbol{z} = \frac{1}{N-1}\sum_{i=2}^{N} \boldsymbol{H}_i$$

The aggregated feature vector $\boldsymbol{z}$ is passed through a three-layer MLP head:

$$\boldsymbol{y} = f_{\text{MLP}}(\boldsymbol{z})$$

The MLP consists of a layer normalization layer, two hidden layers of dimension 512 with GELU activation, dropout with a rate of 0.2, and a final linear layer mapping to output classes.



Overview of the proposed RAD-DINO fine-tuning architecture. The encoder is partially unfrozen, and token representations are aggregated via global average pooling after removing the [CLS] token. The resulting feature vector is passed through an MLP head for multi-label classification.

Selective Layer Unfreezing

Fine-tuning all layers of a large foundation model can lead to overfitting, particularly when the target dataset is limited or exhibits distributional shifts. Prior work has shown that partial fine-tuning can effectively balance the

preservation of pretrained representations with task-specific adaptation (Howard and Ruder 2018; Kornblith et al. 2019).

Motivated by this, we adopt a selective unfreezing strategy, where only the top k layers of the encoder are unfrozen while earlier layers remain fixed. This enables the model to adapt higher-level semantic features that are more task-specific, while retaining low-level representations learned during pretraining, thereby improving generalization and training stability.

Hybrid Class Reweighting Strategy

Medical imaging datasets often exhibit significant class imbalance, which can bias model training toward majority classes and degrade performance on clinically important but underrepresented conditions. To address this, we propose a hybrid class reweighting strategy that combines distribution-based initialization with adaptive refinement driven by model performance.

Initial Class Weights

We first compute class weights based on label distribution:

$$w_c^{(0)} = \frac{N_{\text{neg},c}}{N_{\text{pos},c}}$$

where $N_{\text{pos},c}$ and $N_{\text{neg},c}$ denote the number of positive and negative samples for class c, respectively.

2.4.0.2 Normalization

To ensure numerical stability and balanced scaling across classes, we normalize the weights:

$$w_c = \frac{w_c^{(0)}}{\frac{1}{C}\sum_{c=1}^{C} w_c^{(0)}}$$

Heuristic Adjustment & Adaptive Refinement

We further adjust class weights for clinically significant or empirically difficult classes based on validation behavior. To account for dynamic training behavior, class weights are iteratively refined based on validation AUC, increasing emphasis on underperforming classes during training.

The resulting weighted binary cross-entropy loss is defined as:

$$\mathcal{L} = -\sum_{c=1}^{C} w_c \left[y_c \log(p_c) + (1 - y_c)\log(1 - p_c)\right]$$

Exponential Moving Average

To improve training stability and reduce variance across updates, we maintain an exponential moving average (EMA) of model parameters:

$$\theta_{\text{EMA}} = \alpha\theta_{\text{EMA}} + (1 - \alpha)\theta$$

where α is the decay factor controlling the contribution of historical parameters. EMA acts as a temporal smoothing mechanism, mitigating the effect of noisy gradient updates during training.

Experimental Setup

Datasets

We evaluate our framework across multiple datasets spanning diverse medical imaging modalities and tasks, including classification, detection, and segmentation.

ChestX-ray14: Contains over 112,120 frontal-view chest radiographs annotated with 14 disease labels. It exhibits significant class imbalance and label noise.

CheXpert: Comprises 224,316 chest radiographs from 65,240 patients with uncertainty-aware annotations for 14 observations.

HAM10000: Consists of 10,015 multi-source dermatoscopic RGB images across 7 diagnostic categories of skin lesions.

VinDr-CXR: Provides 18,000 chest radiographs annotated with bounding boxes for 22 local abnormalities and 6 global labels.

TBX-11K: Contains 11,200 chest X-rays annotated with bounding boxes specifically tracking active tuberculosis conditions.

Metrics

We utilize macro-averaged Area Under the Receiver Operating Characteristic Curve (AUC) as the primary classification metric. Object detection is evaluated using mean Average Precision at IoU thresholds of 0.50 (mAP@50) and 0.50:0.95 (mAP@50-95). Segmentation quality is evaluated via the Sorensen-Dice coefficient.

Implementation & Hyperparameter Details

To guarantee reproducibility, we apply rigorous standardized protocols across configurations. All datasets are divided into fixed 80/10/10 splits for training, validation, and testing, ensuring patient-level stratification to eliminate data leakage.

Images for classification and segmentation are resized to 224×224 pixels, whereas object detection pipelines utilize 640×640 pixels. Models are optimized using AdamW with mixed-precision training. The parameter decay factor α for EMA stabilization is fixed at 0.999. Fine-tuning is sustained for a maximum of 50 epochs, constrained by an early stopping patience window of 7 validation epochs monitoring macro AUC.

Hyperparameter Configuration Matrix Across Models

Hyperparameter	Classification	Detection	Segmentation
Base Architecture	RAD-DINO	YOLOv8 / RT-DETR	SAMv2
Encoder LR	3×10^{-6}	1×10^{-5}	5×10^{-6}
Head/Decoder LR	5×10^{-5}	1×10^{-3}	1×10^{-4}
Batch Size	64	16	8
Unfrozen Layers (k)	Top 10 Layers	All Backbone	Memory Attention
LR Scheduler	Cosine Anneal	Linear Decay	Cosine Anneal

Results

Results on HAM10000

Table 2 and Table 3 detail performance statistics across three independent runs on the HAM10000 dataset. Our optimization framework achieves a macro AUC of 0.9320 ± 0.0168. Performance metrics show strong stability

across both high-frequency classes and sparse target entities, such as *bcc* (0.9537 ± 0.0123) and *vasc* (0.9548 ± 0.0182).

Performance comparison across multiple runs on HAM10000

Metric / Class	Run 1	Run 2	Run 3
Macro AUC	0.9243	0.9164	0.9554
Actinic Keratoses (akiec)	0.9468	0.9379	0.9604
Basal Cell Carcinoma (bcc)	0.9379	0.9558	0.9675
Benign Keratosis (bkl)	0.9185	0.8755	0.9318
Dermatofibroma (df)	0.9207	0.8935	0.9873
Melanoma (mel)	0.8554	0.8893	0.9056
Melanocytic Nevi (nv)	0.9396	0.9285	0.9566
Vascular Lesions (vasc)	0.9515	0.9342	0.9788

Mean $\pm$ standard deviation across three runs on HAM10000

Metric / Class	Mean $\pm$ Std
Macro AUC	0.9320 ± 0.0168
Actinic Keratoses (akiec)	0.9484 ± 0.0092
Basal Cell Carcinoma (bcc)	0.9537 ± 0.0123
Benign Keratosis (bkl)	0.9086 ± 0.0238
Dermatofibroma (df)	0.9338 ± 0.0397
Melanoma (mel)	0.8834 ± 0.0209
Melanocytic Nevi (nv)	0.9416 ± 0.0116
Vascular Lesions (vasc)	0.9548 ± 0.0182

Compared to standard baselines in Table 4, our unified optimization pipeline significantly outperforms ResNet50 (0.8388) and DenseNet121 (0.8967) without structural alterations.

Comparison with baseline models on HAM10000

Model	AUCROC
InceptionV3 (Szegedy et al. 2016)	0.8589
ResNet50 (He et al. 2016)	0.8388
DenseNet121 (Huang et al. 2017)	0.8967
ALBEF (Only Images) (Li et al. 2021)	0.9136
Ours (Best SPFT)	**0.9554**

Results on ChestXray14

Table 5 and Table 6 present performance across three independent runs on the ChestXray14 dataset. SPFT achieves a macro AUC of 0.8314 ± 0.0015, maintaining minimal variance across evaluations.

Performance across three independent runs on ChestXray14

Class	Run 1	Run 2	Run 3
Macro AUC	0.8331	0.8302	0.8309
Atelectasis	0.7857	0.7936	0.7977

Cardiomegaly	0.9014	0.8956	0.8999
Effusion	0.8351	0.8380	0.8365
Infiltration	0.7094	0.7088	0.7064
Mass	0.8447	0.8416	0.8471
Nodule	0.7785	0.7804	0.7804
Pneumonia	0.7507	0.7428	0.7471
Pneumothorax	0.8950	0.8912	0.8921
Consolidation	0.7680	0.7635	0.7718
Edema	0.8627	0.8598	0.8636
Emphysema	0.9388	0.9359	0.9364
Fibrosis	0.8468	0.8278	0.8382
Pleural Thickening	0.8009	0.7917	0.7998
Hernia	0.9461	0.9515	0.9152

Mean and standard deviation of AUC on ChestXray14

Class	**Mean ± Std**
Macro AUC	0.8314 ± 0.0015
Atelectasis	0.7923 ± 0.0060
Cardiomegaly	0.8990 ± 0.0030
Effusion	0.8365 ± 0.0015
Infiltration	0.7082 ± 0.0016
Mass	0.8445 ± 0.0028
Nodule	0.7798 ± 0.0011
Pneumonia	0.7469 ± 0.0040
Pneumothorax	0.8928 ± 0.0020
Consolidation	0.7678 ± 0.0042
Edema	0.8620 ± 0.0020
Emphysema	0.9370 ± 0.0015
Fibrosis	0.8376 ± 0.0095
Pleural Thickening	0.7975 ± 0.0049
Hernia	0.9376 ± 0.0196

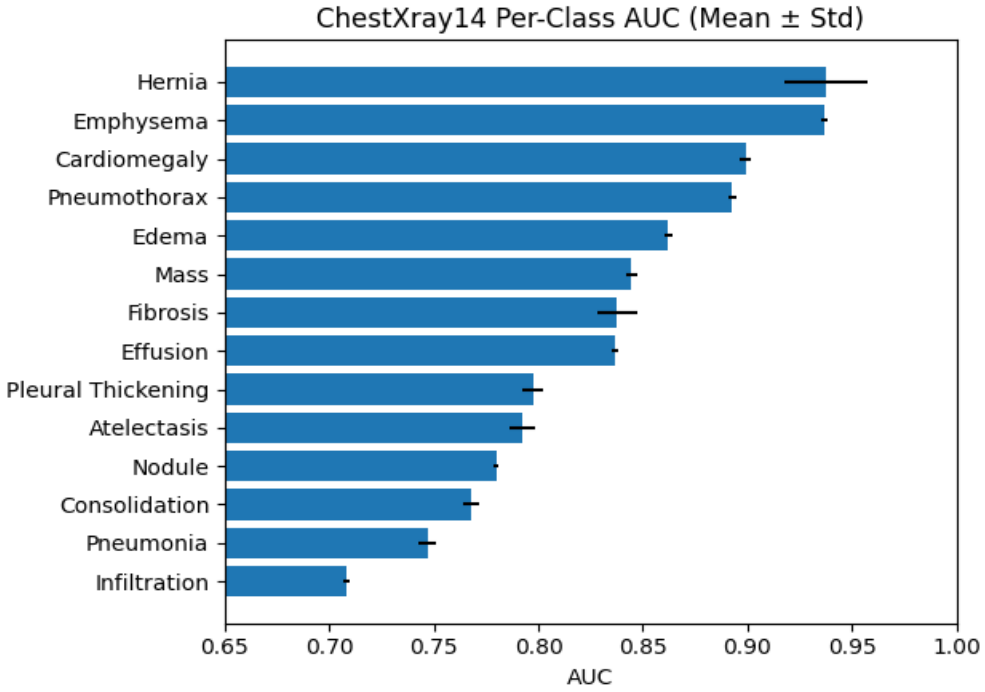
Table 7 details the results of varying the fine-tuning depth. Limiting parameter relaxation to the terminal 10 layers outperforms full unfreezing configurations by preventing over-adaptation.

Comparison of different fine-tuning strategies on ChestXray14

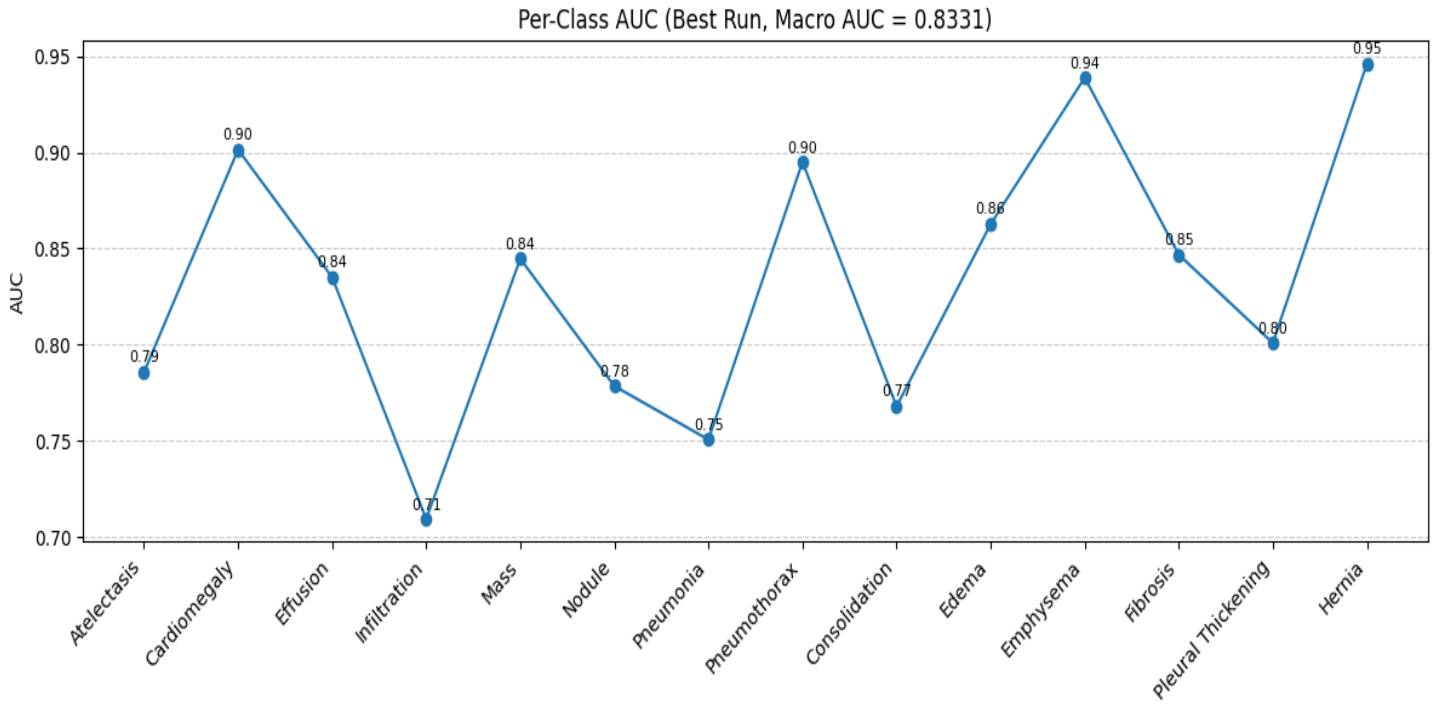
Class	**SPFT-10 layers**	**SPFT-all layers**
Macro AUC	**0.8331**	0.8320
Atelectasis	0.7857	0.7766
Cardiomegaly	0.9014	0.8941
Effusion	0.8351	0.8302
Infiltration	0.7094	0.7077
Mass	0.8447	0.8321
Nodule	0.7785	0.7584

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Pneumonia	0.7507	0.7369
Pneumothorax	0.8950	0.8882
Consolidation	0.7680	0.7508
Edema	0.8627	0.8594
Emphysema	0.9388	0.9315
Fibrosis	0.8468	0.8232
Pleural Thickening	0.8009	0.7806
Hernia	0.9461	0.9518



Per-class AUC on ChestXray14. Error bars denote standard deviation across runs.



Per-class AUC for the best-performing model on ChestXray14.

Results on CheXpert

Table 8 presents the performance of our method on the CheXpert dataset compared to existing models. Our single optimization-based strategy yields a Macro AUC of 0.8420, matching single-model benchmarks without architectural alterations or multimodal training.

Comparison of representative models on the CheXpert dataset

Model	Type	Mean AUC
DenseNet-121 Baseline (Irvin et al. 2019)	CNN (Supervised)	0.8280
CheXzero (Tiu et al. 2022)	Vision-Language	0.7510
GLoRIA (Huang et al. 2021)	Vision-Language	0.8730
BioViL (Boecking et al. 2022)	Vision-Language	0.8820
MedCLIP (Wang et al. 2022)	Vision-Language	0.8750
SISTA (Ma et al. 2025)	Representation Learning	0.8710
Ours (SPFT)	Optimization-based	**0.8420**

Performance on the CheXpert validation set (AUC)

Class	AUC
Atelectasis	0.8277
Cardiomegaly	0.7884
Consolidation	0.8801
Edema	0.9175
Enlarged Cardiomediastinum	0.5146
Fracture	N/A
Lung Lesion	0.6738
Lung Opacity	0.9256
No Finding	0.8739
Pleural Effusion	0.9209
Pleural Other	0.8584
Pneumonia	0.8092
Pneumothorax	0.8850
Support Devices	0.9705
Macro AUC	**0.8420**

Table 9 breaks down the results by class. Undefined validation rows containing zero active evaluations (*Fracture*) were omitted from final aggregation scores.

Unified Evaluation Across Tasks

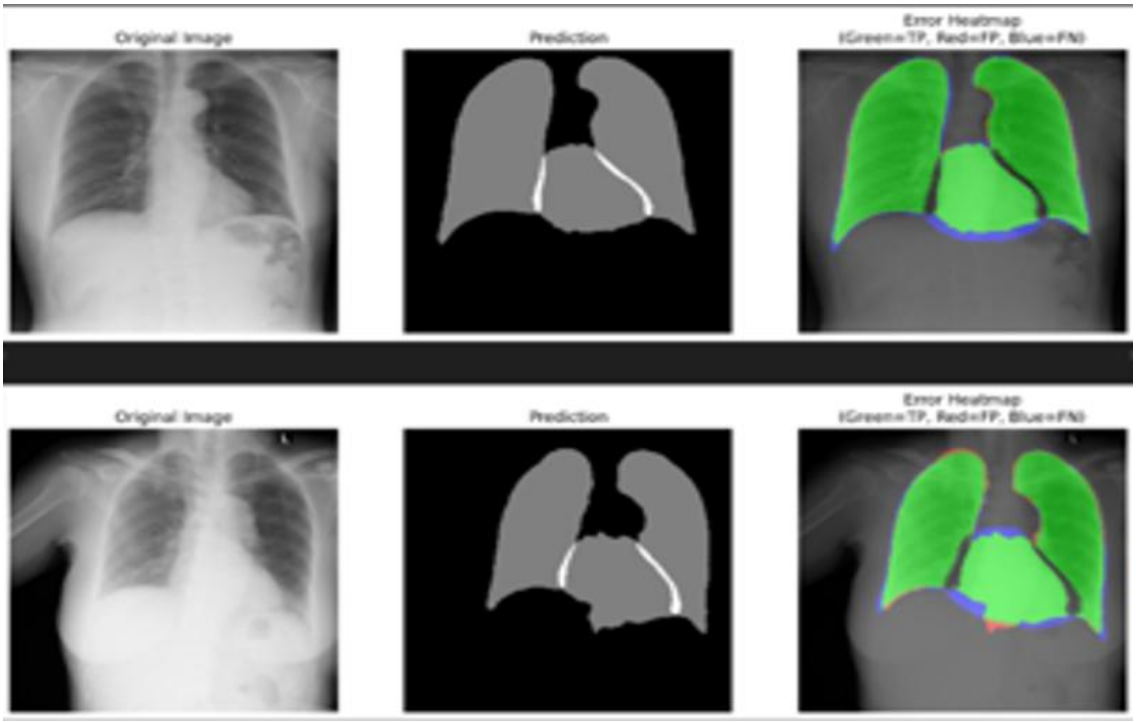
A key aspect of this work is the structural validation of SPFT across varied localization tasks, using consistent learning constraints.

Segmentation Results (VinDr-CXR)

We integrate the SPFT training strategy into a SAMv2 segmentation pipeline on the VinDr-CXR dataset. Instead of using a classification head, the target parameter blocks are optimized against a combination of focal and dice losses. The model achieves competitive Dice scores across major anatomical structures:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

- Heart: 0.91
- Lungs: 0.95 (Left), 0.96 (Right)



Segmentation results on VinDr-CXR. Predicted masks align well with anatomical structures such as lungs and heart.

Detection Results

Object detection capabilities are evaluated by incorporating the SPFT optimization constraints directly into the backbones of YOLOv8 and RT-DETR during dense coordinate training.

Detection performance comparison under unified SPFT training

Model	mAP@50	mAP@50-95	Precision	Recall
YOLOv8	0.76	0.37	0.83	0.67
RT-DETR	0.16	0.07	0.19	0.36

Detection results on VinDr-CXR. The SPFT-enhanced model produces more precise and stable bounding boxes compared to baseline models.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

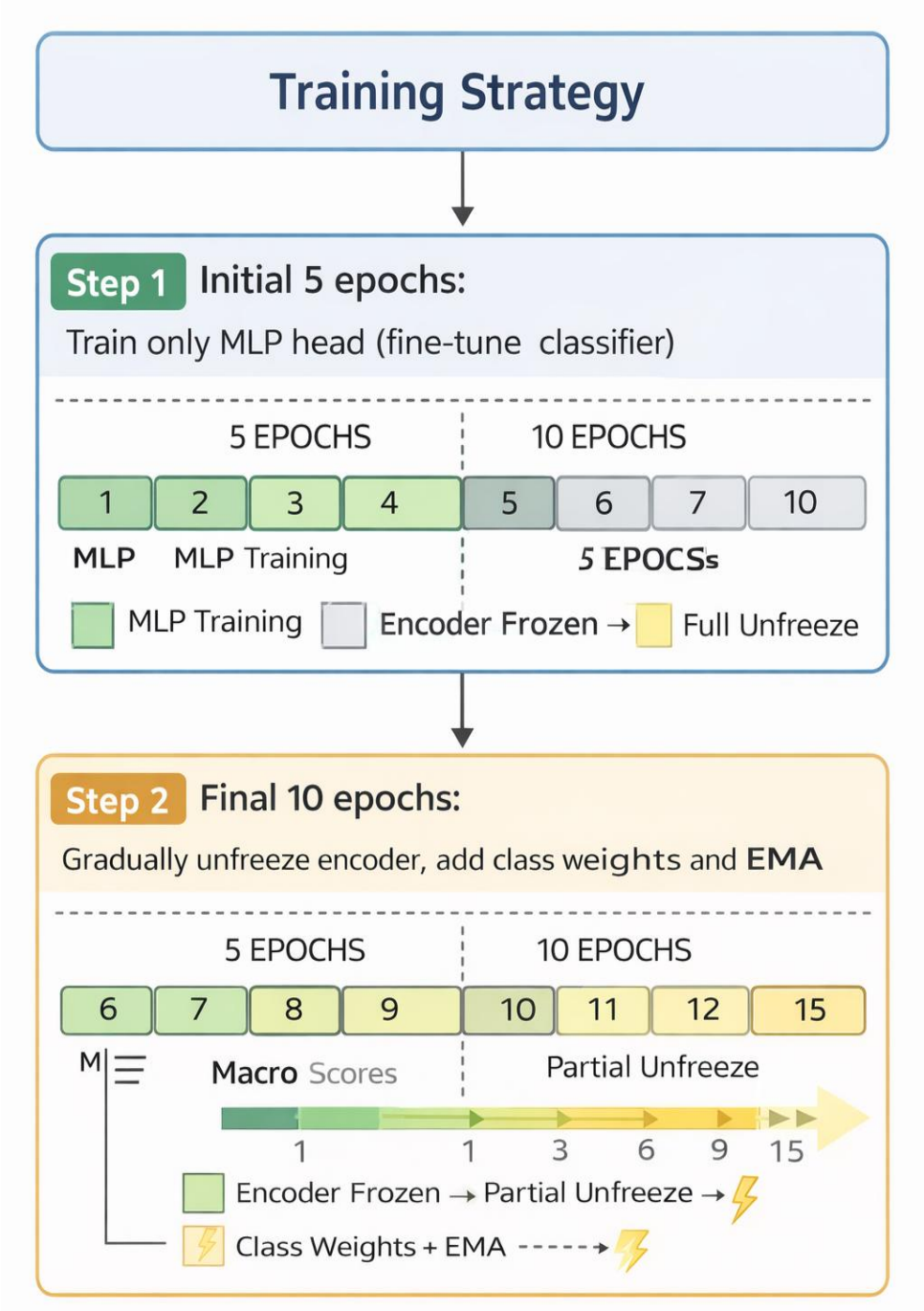
As shown in Table 10, YOLOv8 outperforms the transformer-based RT-DETR. This difference highlights the benefit of CNN spatial inductive biases when operating under sparse annotation limits typical of medical data.

Key Insight: Optimization vs Architecture

The core adaptation logic of SPFT remains identical across tasks. However, performance boundaries vary based on the underlying architecture's suitability for a given task. While SPFT stabilizes representation adaptation, selecting an appropriate model architecture remains vital for handling data constraints.

Ablation Study

To analyze the contribution of each component in the proposed training strategy, we conduct a systematic ablation study on ChestXray14 following a two-stage process (Figure 6).



Overview of the proposed two-stage training strategy. The model is first trained using only the MLP head, followed by progressive unfreezing of encoder layers with class-aware weighting and EMA stabilization.

Ablation study of the proposed training strategy on ChestXray14

Method	Unf.	Wts.	EMA	AUC
MLP Head Only	×	×	×	0.8095
+ Prog. Unfreezing	✓	×	×	0.8216
+ Class Weights	✓	✓	×	0.8306
+ EMA (Full)	✓	✓	✓	**0.8331**

Table 11 shows that incremental integration of progressive unfreezing (+1.21%), hybrid class weighting (+0.90%), and parameter smoothing via EMA (+0.25%) yields consistent performance gains.

Effect of Fine-Tuning Across Datasets

On HAM10000, the frozen encoder achieved performance close to the fine-tuned configuration. This indicates that when initial parameters align well with the target domain, further parameter relaxation offers minimal benefit. Conversely, optimization strategies such as SPFT provide more substantial improvements on distribution-shifted target tasks.

Reproducibility

All implementation assets, configuration scripts, and parameter sets will be released in a public repository upon manuscript acceptance.

DISCUSSION

Our results show that carefully designed optimization strategies can improve model adaptation without architectural modifications or intensive data augmentation. The ablation study highlights how progressive unfreezing, class-aware weighting, and EMA work together to stabilize training dynamics and improve overall performance.

Limitations and Clinical Applicability

While our approach shows strong generalization, several limitations must be considered before real-world clinical deployment:

- **Dataset Bias and Label Noise:** Standard benchmarks contain localized annotation biases and label noise. Relying solely on these datasets without external clinical validation can lead to overfitting on institutional artifacts.
- **Absence of Uncertainty Quantification:** Our current framework provides deterministic outputs. Safe clinical deployment requires active confidence tracking and uncertainty analysis to identify out-of-distribution samples.
- **Lack of Clinical Context:** The model operates purely on pixel data, missing secondary indicators like patient history, longitudinal metrics, and multi-modal records.

Future work will focus on integrating automated uncertainty quantification and validating our approach on external clinical datasets to ensure real-world robustness.

CONCLUSION

In this work, we present a unified optimization-driven framework (SPFT) for adapting foundation models across heterogeneous medical imaging tasks. By combining progressive layer unfreezing, hybrid class reweighting, and EMA stabilization, our approach achieves competitive performance across classification, detection, and

segmentation tasks without modifying the underlying architecture. These findings suggest that optimization-driven adaptation is a scalable and practical approach for deploying foundation models in complex clinical settings.

REFERENCE

1. Anonymous. 2023. “RAD-DINO: Exploring Vision Foundation Models for Radiology.” arXiv Preprint.
2. Boecking, Benedikt et al. 2022. “BioViL: A Vision-Language Model for Biomedical Applications.” arXiv Preprint arXiv:2206.07036.
3. Caron, Mathilde et al. 2021. “Emerging Properties in Self-Supervised Vision Transformers.” ICCV.
4. He, Kaiming et al. 2016. “Deep Residual Learning for Image Recognition.” CVPR.
5. Howard, Jeremy, and Sebastian Ruder. 2018. “Universal Language Model Fine-Tuning for Text Classification.” ACL.
6. Huang, Gao et al. 2017. “Densely Connected Convolutional Networks.” CVPR.
7. Huang, Shih-Cheng et al. 2021. “GLoRIA: Global-Local Representation Learning for Medical Images.” ICCV.
8. Irvin, Jeremy, Pranav Rajpurkar, Michael Ko, et al. 2019. “CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison.” Proceedings of the AAAI Conference on Artificial Intelligence 33: 590–97.
9. Kornblith, Simon et al. 2019. “Better Fine-Tuning by Reducing Representational Collapse.” ICML.
10. Li, Junnan et al. 2021. “Align Before Fuse: Vision and Language Representation Learning with Momentum Distillation.” NeurIPS.
11. Ma, DongAo, Jiaxuan Pang, Michael B. Gotway, and Jianming Liang. 2025. “A Fully Open AI Foundation Model Applied to Chest Radiography.” Nature 643 (8071): 488–98. https://doi.org/10.1038/s41586-025-09079-8.
12. Oquab, Maxime et al. 2023. “DINOv2: Learning Robust Visual Features Without Supervision.” arXiv.
13. Szegedy, Christian et al. 2016. “Rethinking the Inception Architecture for Computer Vision.” CVPR.
14. Tiu, Ekin et al. 2022. “CheXzero: A Zero-Shot Learning Model for Chest x-Ray Classification.” Nature Biomedical Engineering.
15. Tschandl, Philipp et al. 2018. “The HAM10000 Dataset.” Scientific Data.
16. Wang, Xiaosong, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. 2017. “ChestX-Ray8: Hospital-Scale Chest x-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
17. Wang, Zifeng et al. 2022. “MedCLIP: Contrastive Learning from Unpaired Medical Images and Text.” EMNLP.