The Adjustment of Hyperparameters and Model Architecture in Prediction Model of OVF Collapse Progression

Sibeen Kim

Short Bio

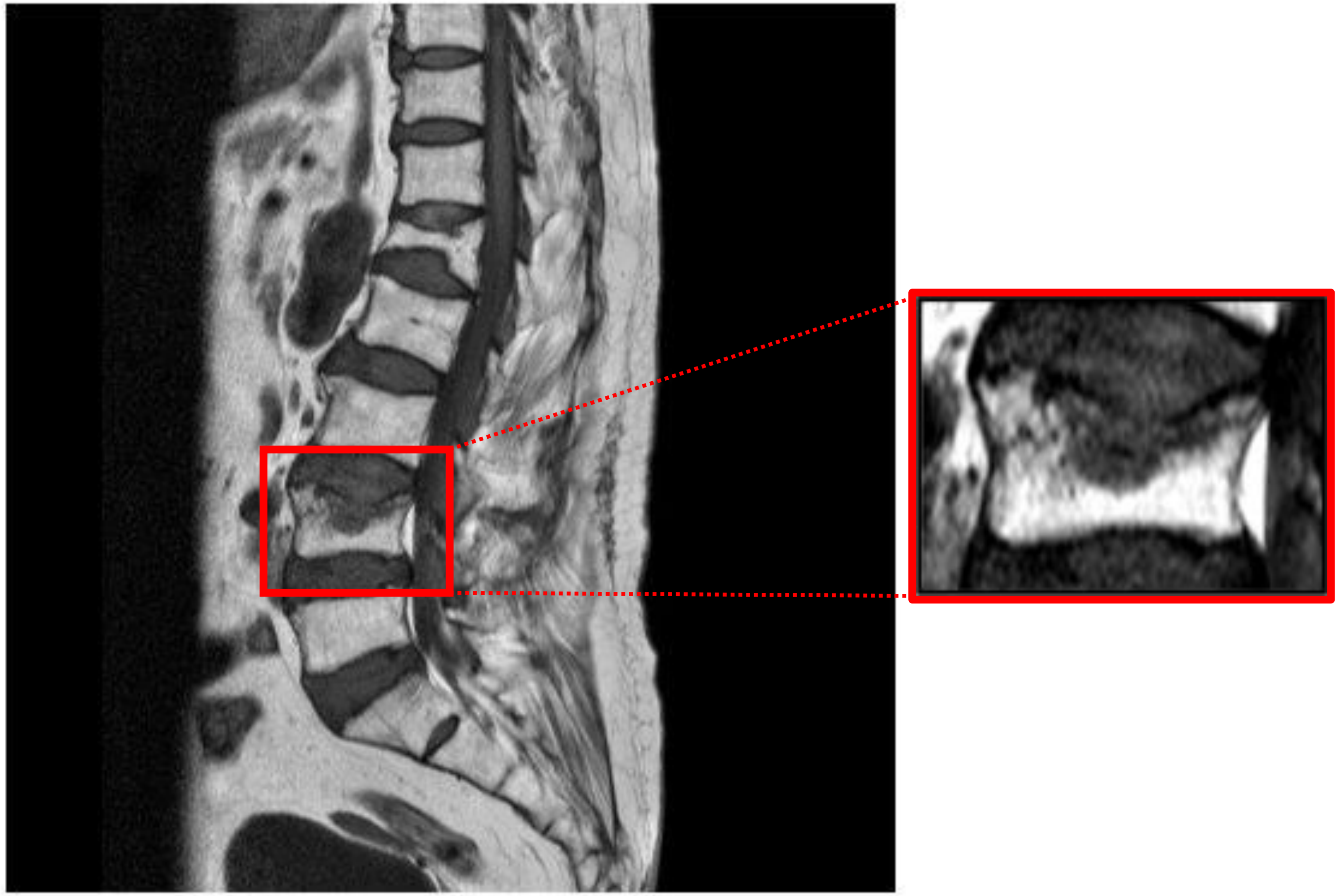
- B.S. in Biomedical Engineering, Korea University — Mar 2018 - Present

- Research Intern @ *LANIB, Severance Hospital — Apr 2023 - Present
- Research Intern @ Letsur — Jan 2024 - Present
- Research Intern @ **BREIN Lab, Korea University — Feb 2024 - Present

* Laboratory of Advanced Neuroimaging Biomarker research

** Brain Reverse Engineering by Intelligent Neuroimaging

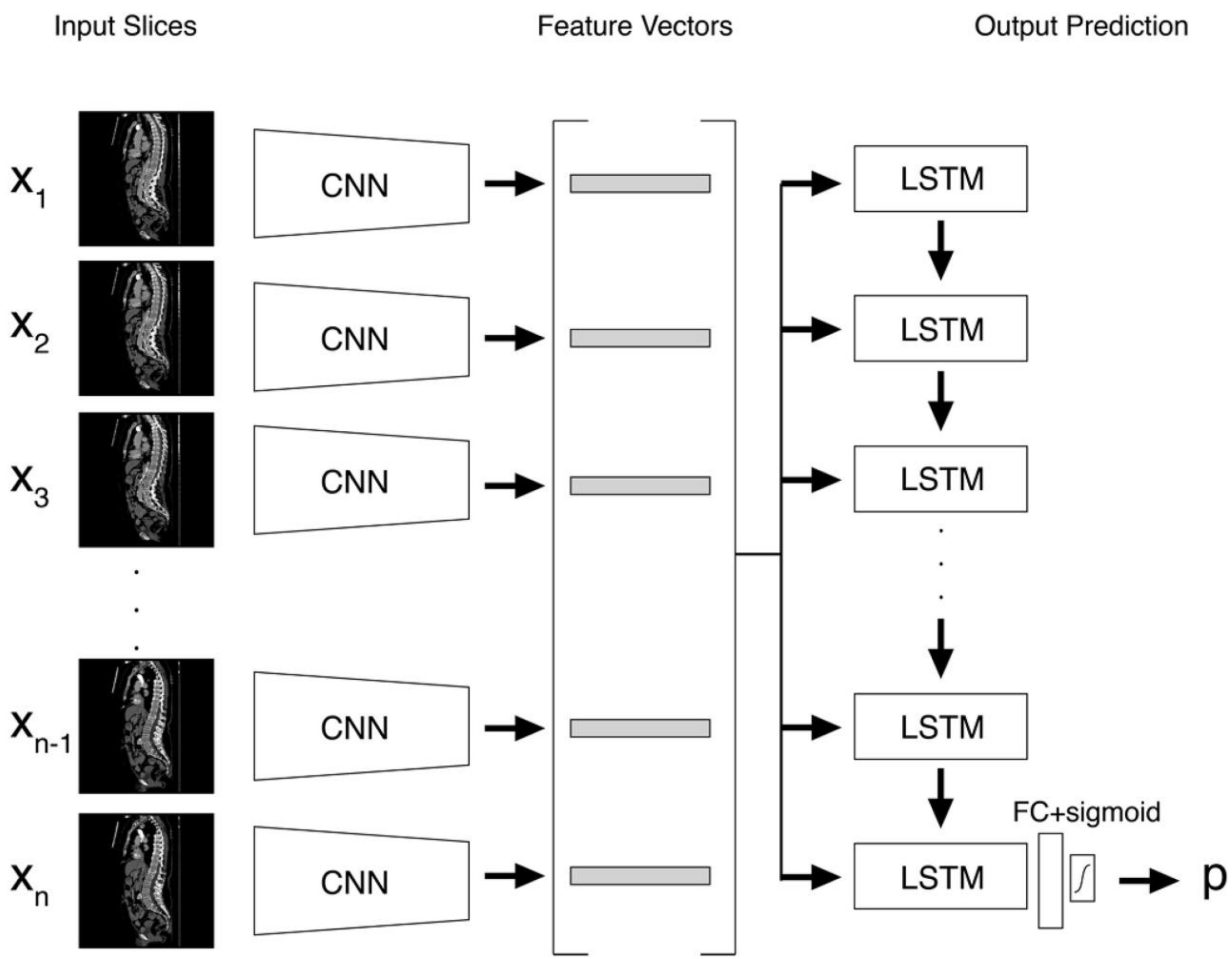
Osteoporotic Vertebral Compression Fracture

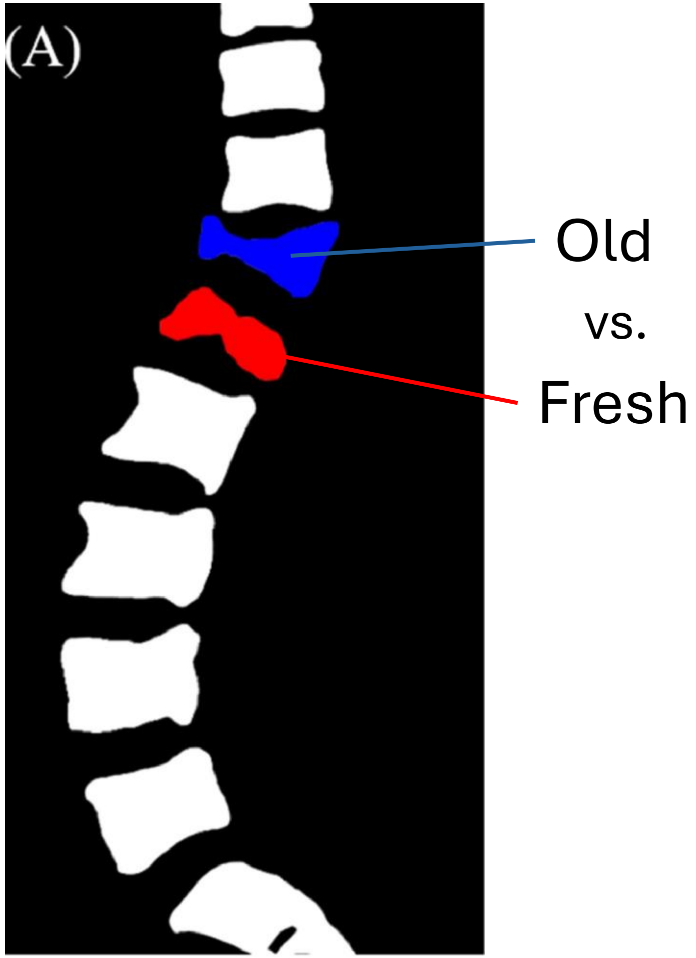


Vertebral Collapse

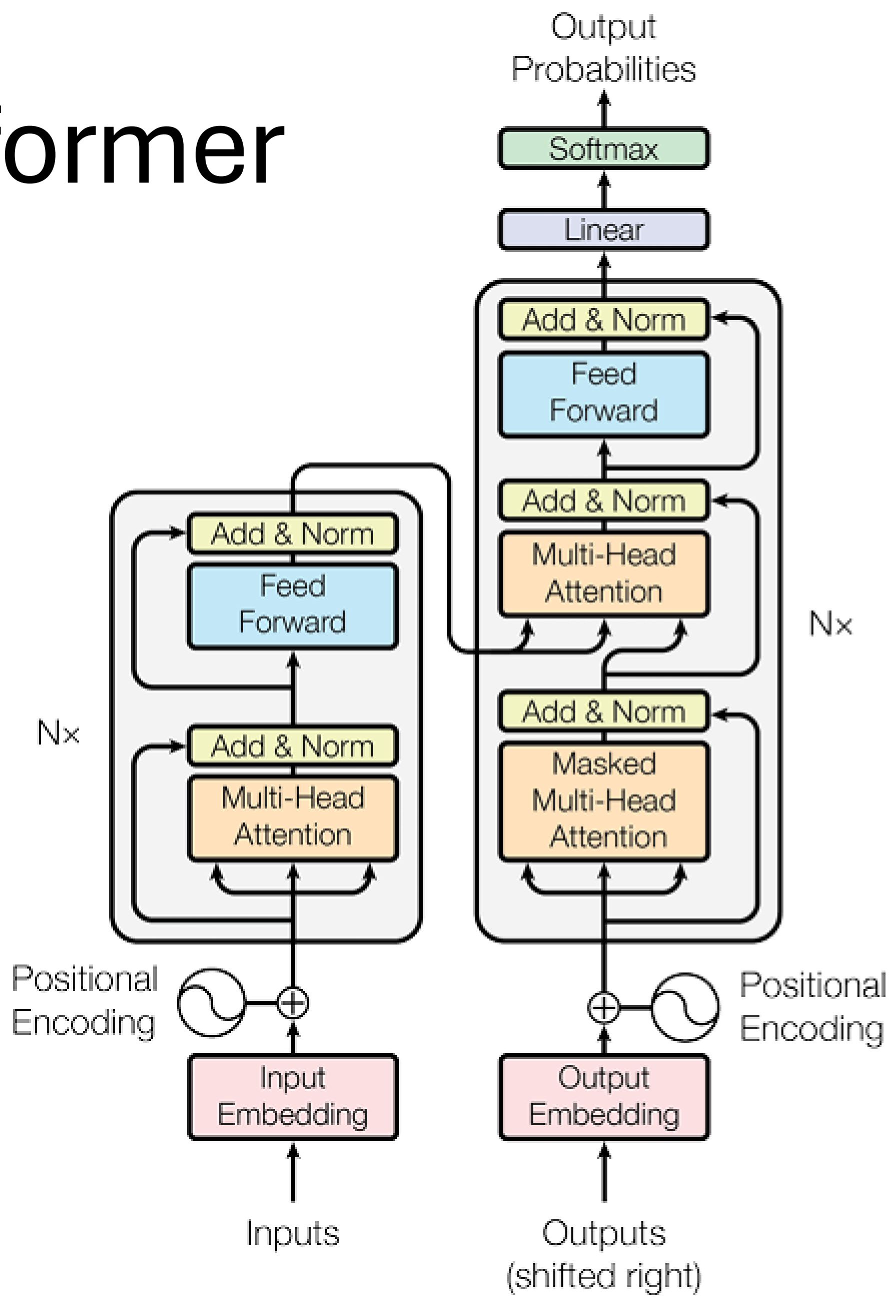
7-37%

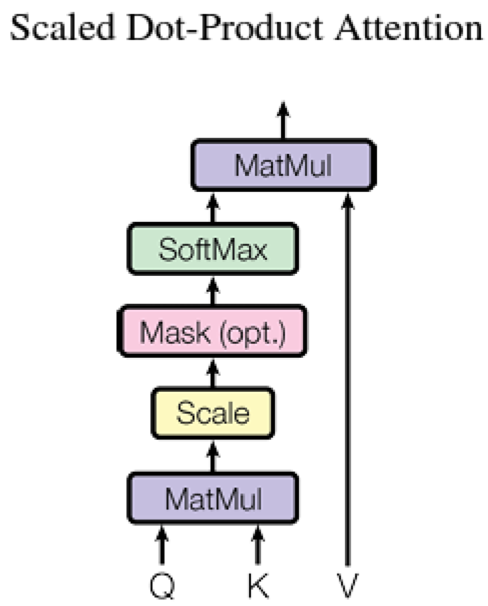
CNN for OVF Diagnosis



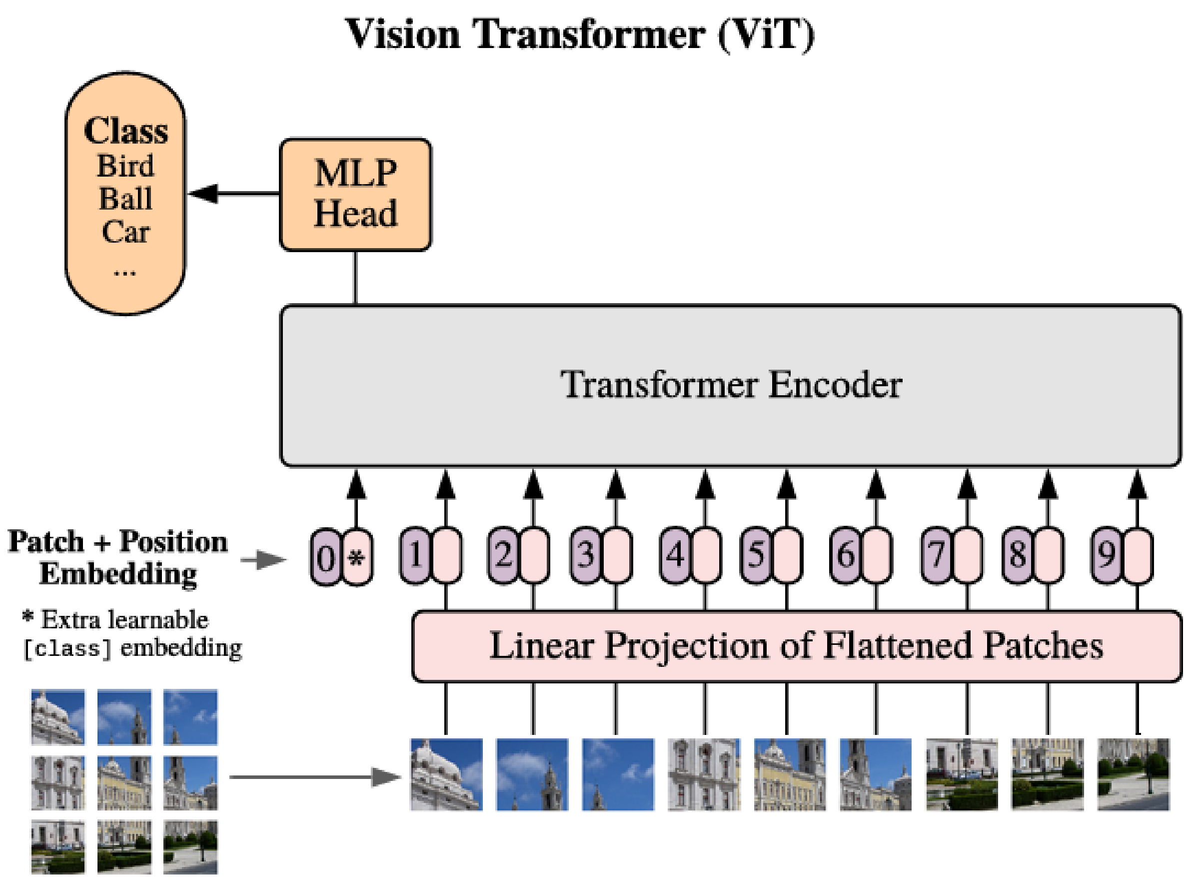


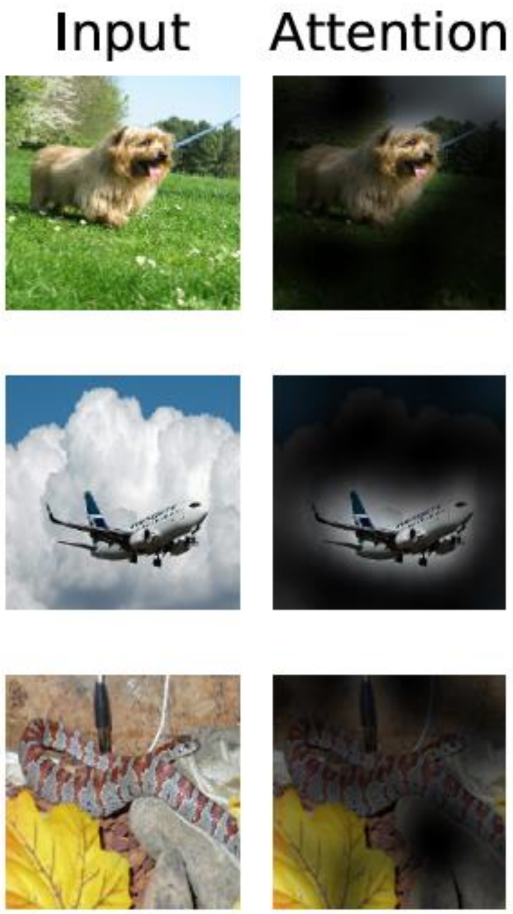
Transformer



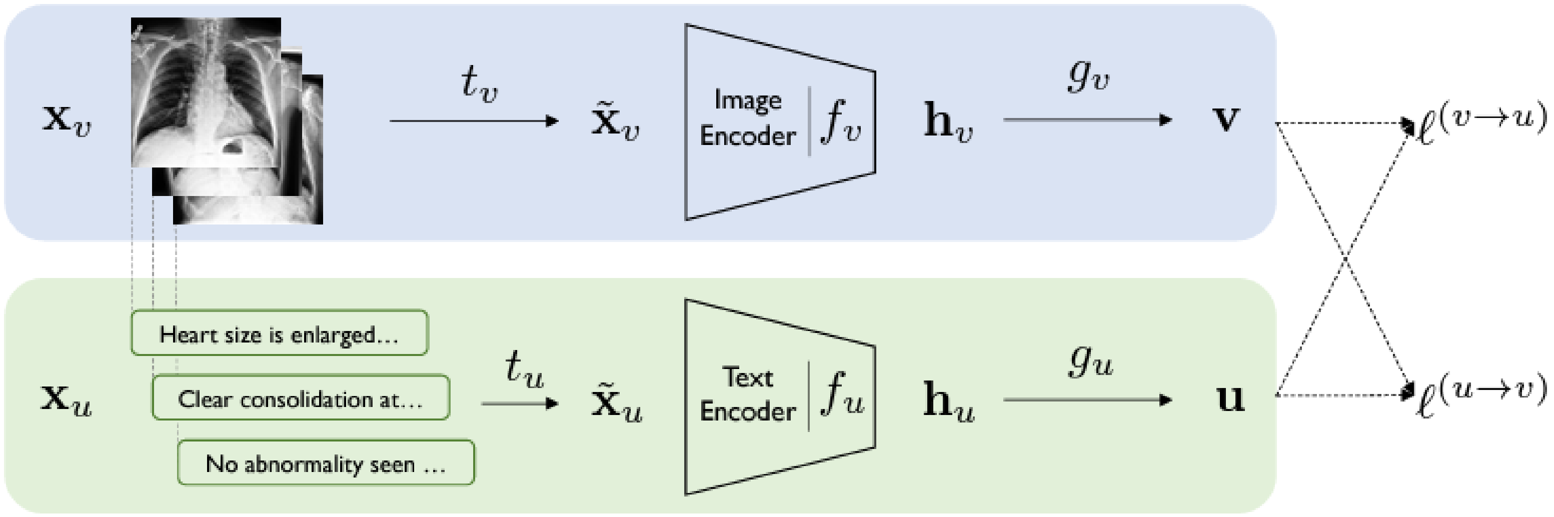


Vision Transformer





Contrastive Visual Representation Learning from Text

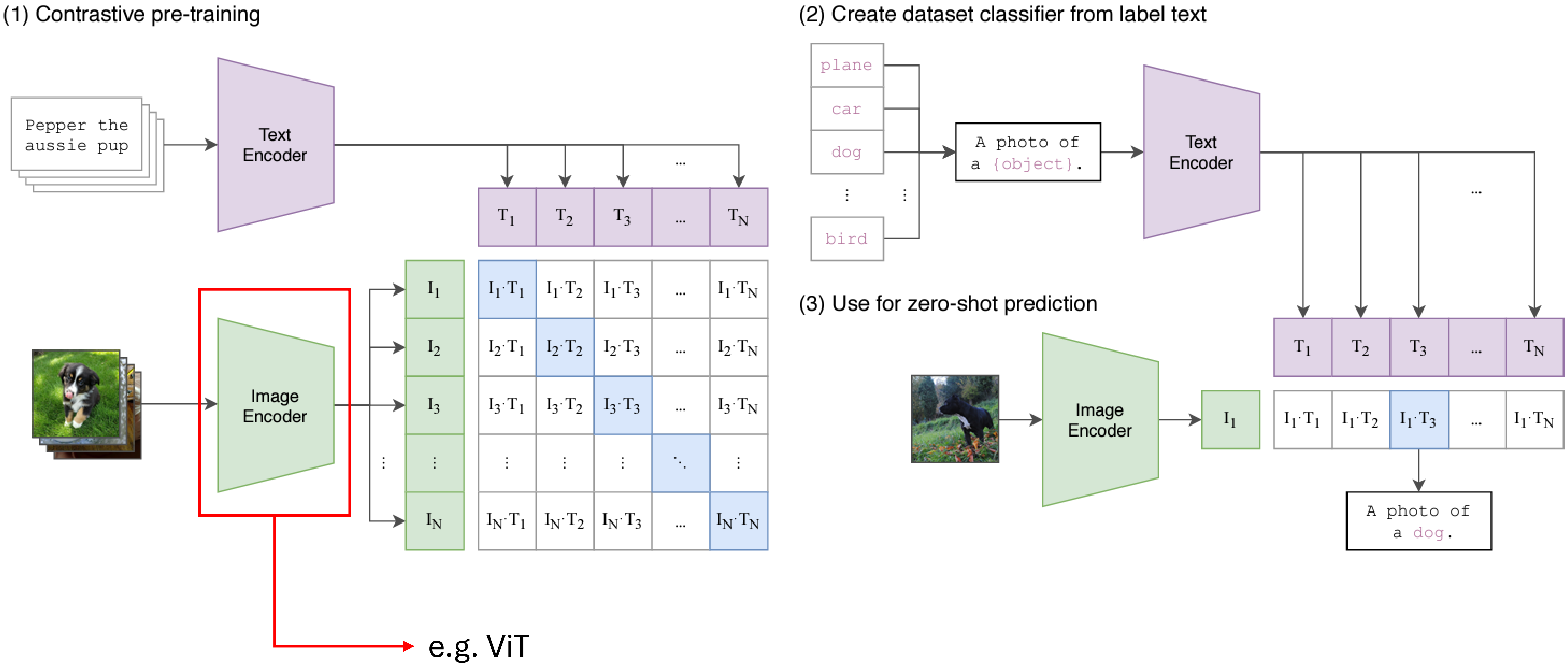


$$\ell_i^{(v\to u)} = -\log \frac{\exp(\langle \mathbf{v}_i, \mathbf{u}_i \rangle / \tau)}{\sum_{k=1}^{N} \exp(\langle \mathbf{v}_i, \mathbf{u}_k \rangle / \tau)}$$

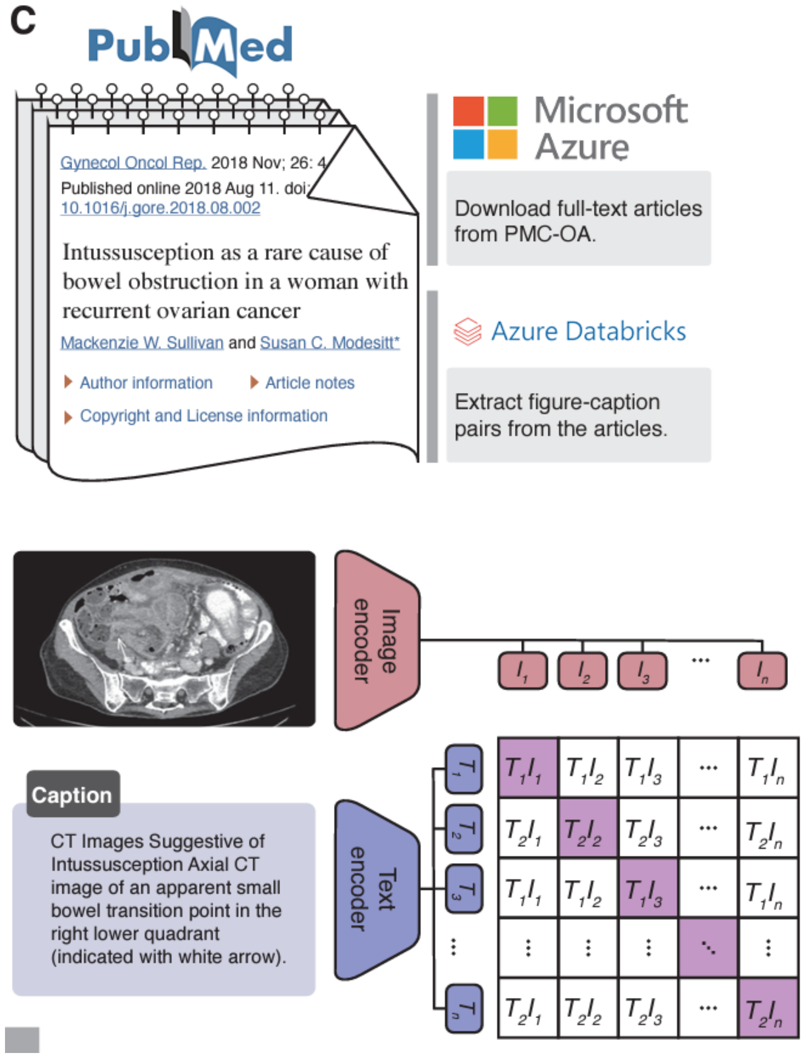
$$\ell_i^{(u\to v)} = -\log \frac{\exp(\langle \mathbf{u}_i, \mathbf{v}_i \rangle / \tau)}{\sum_{k=1}^{N} \exp(\langle \mathbf{u}_i, \mathbf{v}_k \rangle / \tau)}$$

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left(\lambda \ell_i^{(v\to u)} + (1-\lambda) \ell_i^{(u\to v)} \right)$$

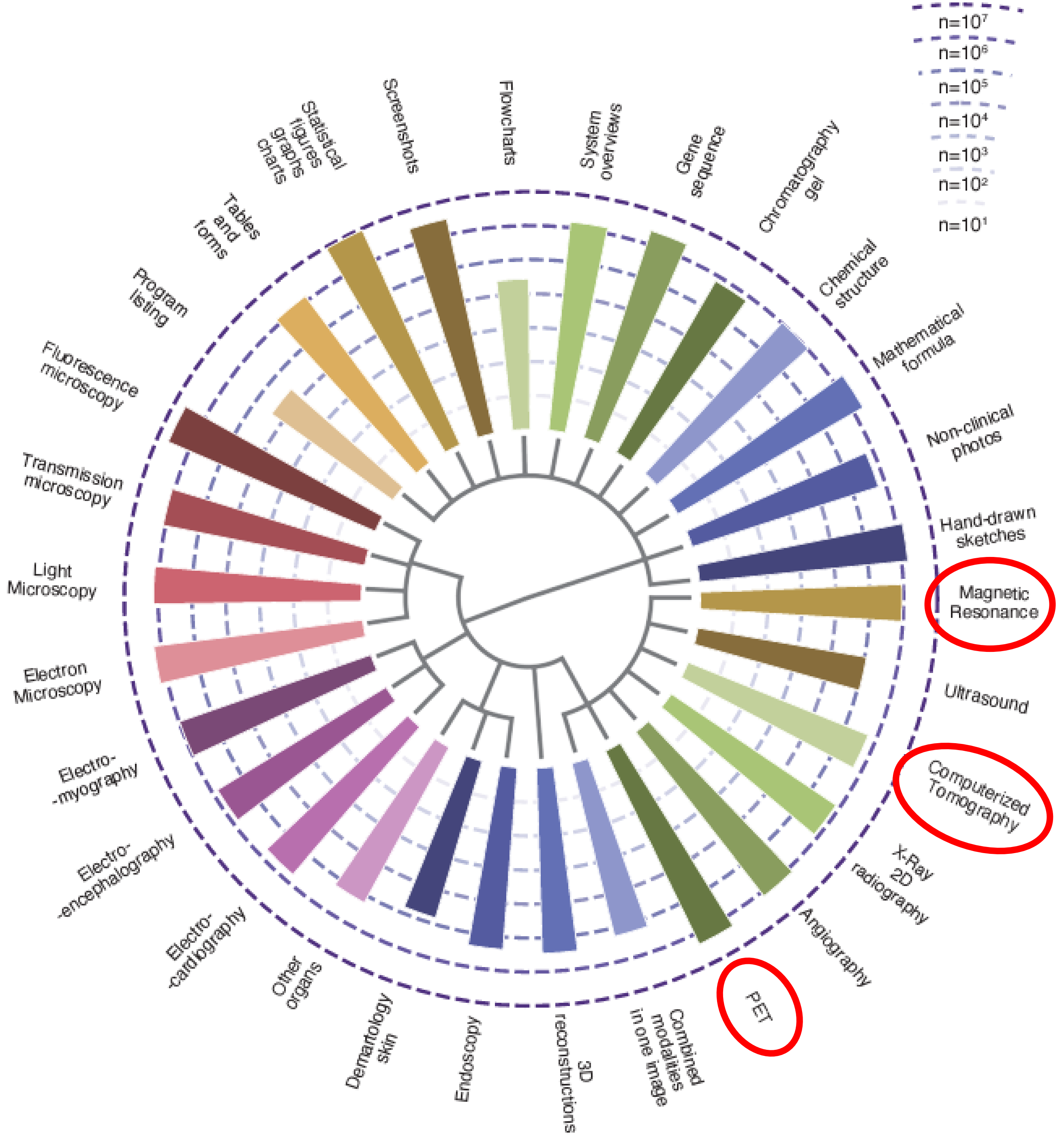
CLIP


(1) Contrastive pre-training
Pepper the aussie pup
Text Encoder
Image Encoder
e.g. ViT
(2) Create dataset classifier from label text
plane
car
dog
bird
A photo of a {object}.
Text Encoder
(3) Use for zero-shot prediction
Image Encoder
A photo of a dog.

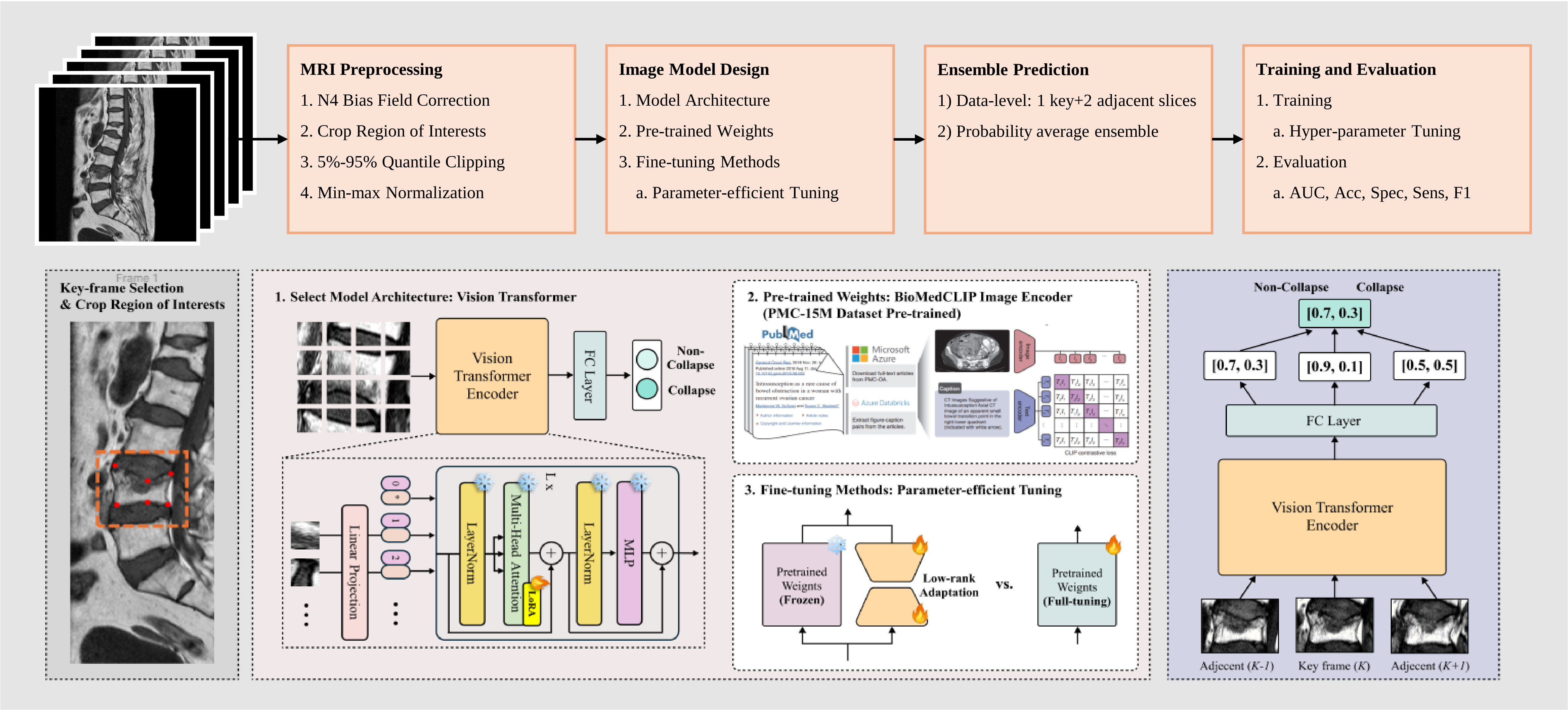
BiomedCLIP



c
PubMed
Gynecol Oncol Rep. 2018 Nov; 26: 4
Published online 2018 Aug 11. doi:
10.1016/j.gore.2018.08.002
Intussusception as a rare cause of bowel obstruction in a woman with recurrent ovarian cancer
Mackenzie W. Sullivan and Susan C. Modesitt*
Author information
Article notes
Copyright and License information
Microsoft Azure
Download full-text articles from PMC-OA.
Azure Databricks
Extract figure-caption pairs from the articles.
Image encoder
Caption
CT Images Suggestive of Intussusception Axial CT image of an apparent small bowel transition point in the right lower quadrant (indicated with white arrow).
Text encoder



n=10^7
n=10^6
n=10^5
n=10^4
n=10^3
n=10^2
n=10^1
Screenshots
Flowcharts
System overviews
Gene sequence
Chromatography gel
Chemical structure
Mathematical formula
Non-clinical photos
Hand-drawn sketches
Magnetic Resonance
Ultrasound
Computerized Tomography
X-Ray 2D radiography
Angiography
PET
Combined modalities in one image
3D reconstructions
Endoscopy
Dermatology skin
Other organs
Electro-cardiography
Electro-encephalography
Electro-myography
Electron Microscopy
Light Microscopy
Transmission microscopy
Fluorescence microscopy
Program listing
Tables and forms
Statistical figures graphs charts



MRI Preprocessing
1. N4 Bias Field Correction
2. Crop Region of Interests
3. 5%-95% Quantile Clipping
4. Min-max Normalization
Image Model Design
1. Model Architecture
2. Pre-trained Weights
3. Fine-tuning Methods
a. Parameter-efficient Tuning
Ensemble Prediction
1) Data-level: 1 key+2 adjacent slices
2) Probability average ensemble
Training and Evaluation
1. Training
a. Hyper-parameter Tuning
2. Evaluation
a. AUC, Acc, Spec, Sens, F1
Key-frame Selection
& Crop Region of Interests
1. Select Model Architecture: Vision Transformer
Vision Transformer Encoder
FC Layer
Non-Collapse
Collapse
Linear Projection
LayerNorm
Multi-Head Attention
LoRA
LayerNorm
MLP
L x
2. Pre-trained Weights: BioMedCLIP Image Encoder
(PMC-15M Dataset Pre-trained)
PubMed
Microsoft Azure
Azure Databricks
3. Fine-tuning Methods: Parameter-efficient Tuning
Pretrained Weights (Frozen)
Low-rank Adaptation
vs.
Pretrained Weights (Full-tuning)
Non-Collapse
Collapse
[0.7, 0.3]
[0.7, 0.3]
[0.9, 0.1]
[0.5, 0.5]
FC Layer
Vision Transformer Encoder
Adjecent (K-1)
Key frame (K)
Adjecent (K+1)

Demography

Table 1. Demographic data of enrolled patients

	Non-Collapse (n=125) Mean (SD) or n (%)	Collapse (n=120) Mean (SD) or n (%)	*P*-value
Age (years)	72.0 (9.98)	73.6 (8.67)	0.174
T-score of BMD (Lumbar)	-2.72 (1.11)	-2.74 (1.04)	0.852
Sex (female)	103 (82.4)	94 (78.3)	0.423
Level (Lumbar)	87 (69.6)	73 (60.3)	0.150
History of Medication for osteoporosis			
Before OVCF Dx.	28 (22.4)	22 (18.3)	0.430
After OVCF Dx.	99 (79.2)	90 (75.0)	0.492

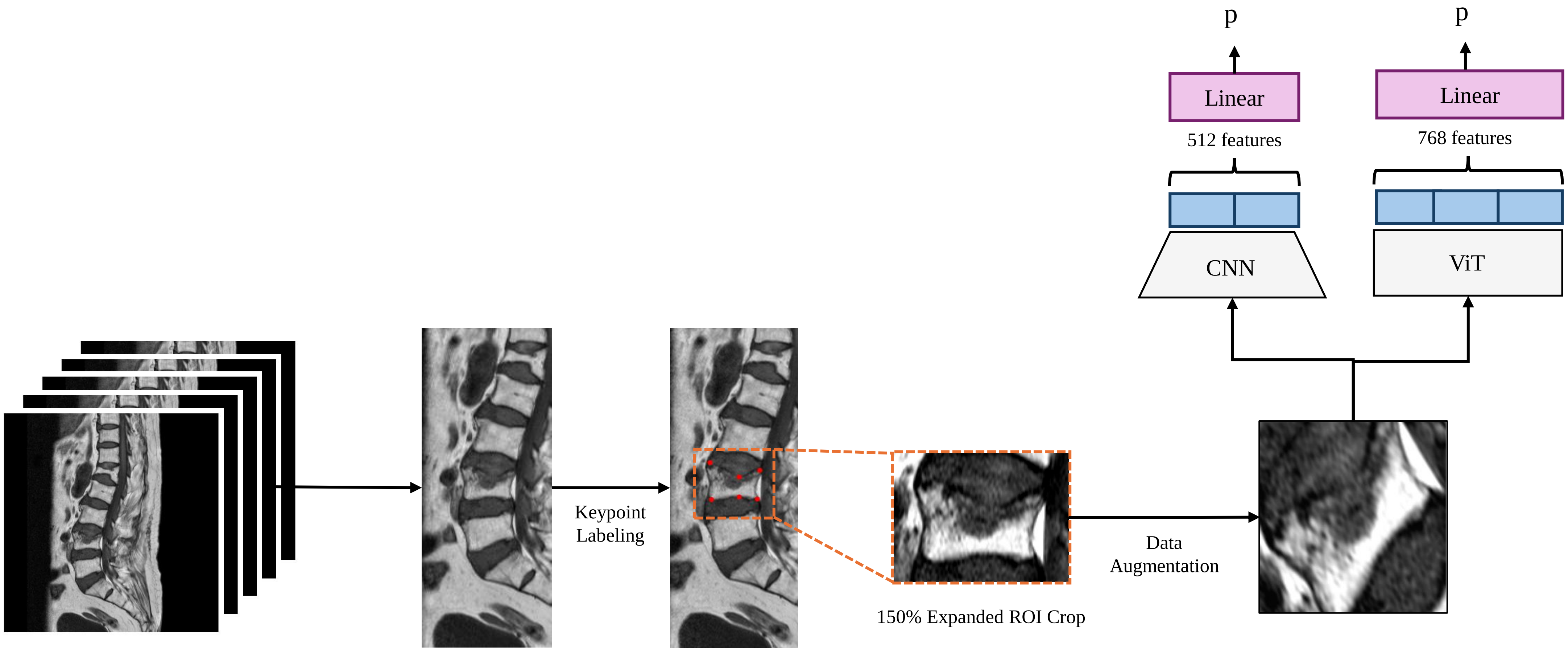
Table 2. Dataset splits across participating institutions

	BRMH	KNUH	HUDSHH	KUDH	SCHH	
Train	109	55	36	0	0	200 (81.6%)
Test	0	0	0	30	15	45 (18.4%)

Abbreviations: BRMH, Boramae Hospital; SNUH, Seoul National University Hospital; KNUH, Kangwon National University Hospital; HUDSHH, Hallym University Dongtan Sacred Heart Hospital; KUDH, Keimyung University Dongsan Hospital; SCHH, Soon Chun Hyang University Hospital

Preprocessing

Preprocessing



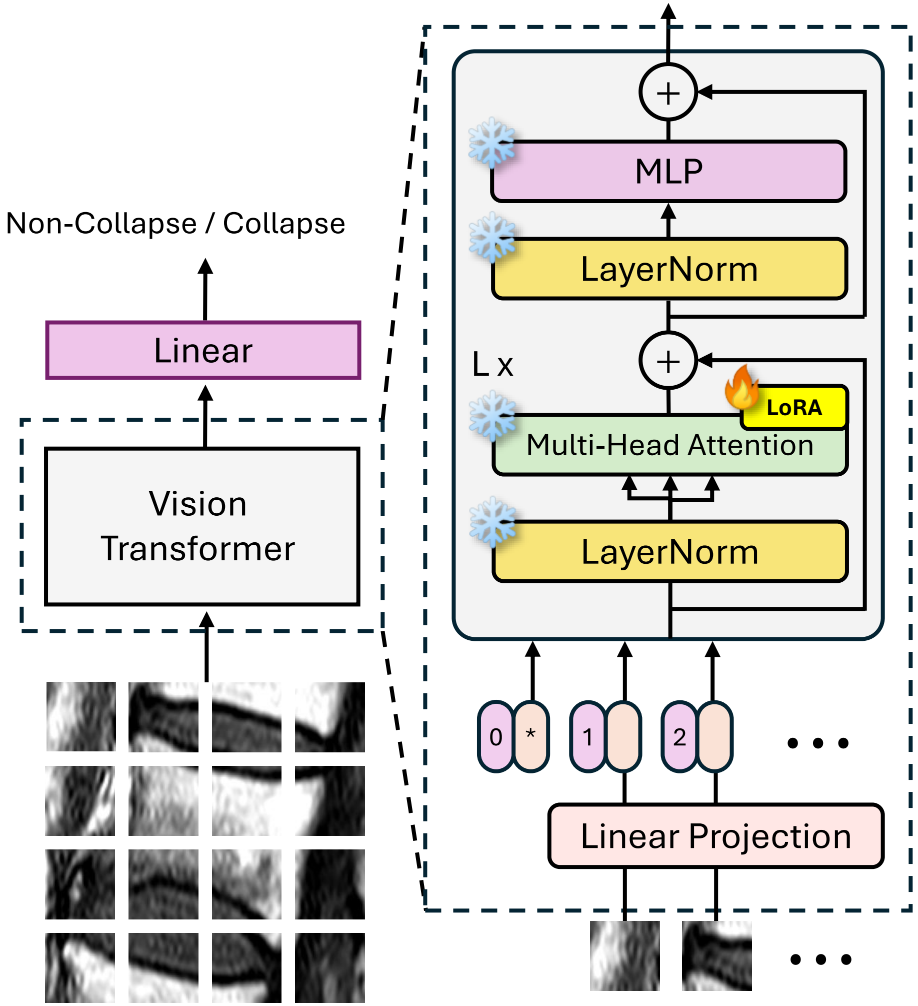
Model Design: Pretrained Weights

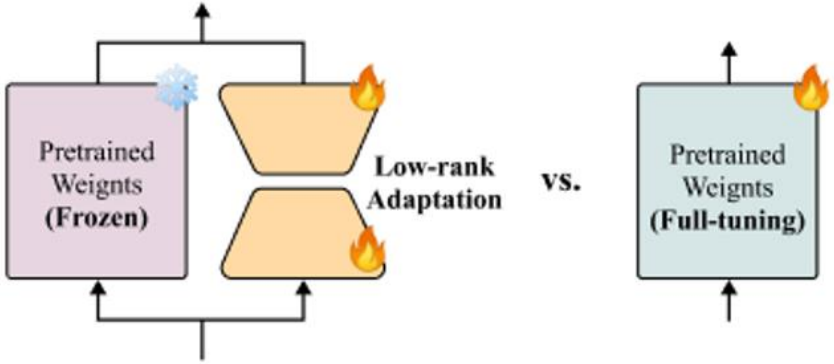
	Internal Validation			External Validation		
	AUC	Spec	Sens	**AUC**	Spec	Sens
CNN scratch	0.7830	0.7844	0.6699	0.7097	0.6963	0.6611
CNN ImageNet	**0.7973**	0.7833	0.7123	**0.8014**	0.7074	0.7833

Model Design: Pretrained Weights

	Internal Validation			**External Validation**		
	AUC	Spec	Sens	**AUC**	Spec	Sens
ViT Scratch	0.8159	0.7996	0.7086	0.7979	0.7630	0.6944
ViT ImageNet	0.8185	0.8642	0.6517	0.7825	0.7333	0.7333
ViT PMC	**0.8269**	0.8057	0.7306	**0.8051**	0.7296	0.7556

Model Design: Fine-tuning Methods





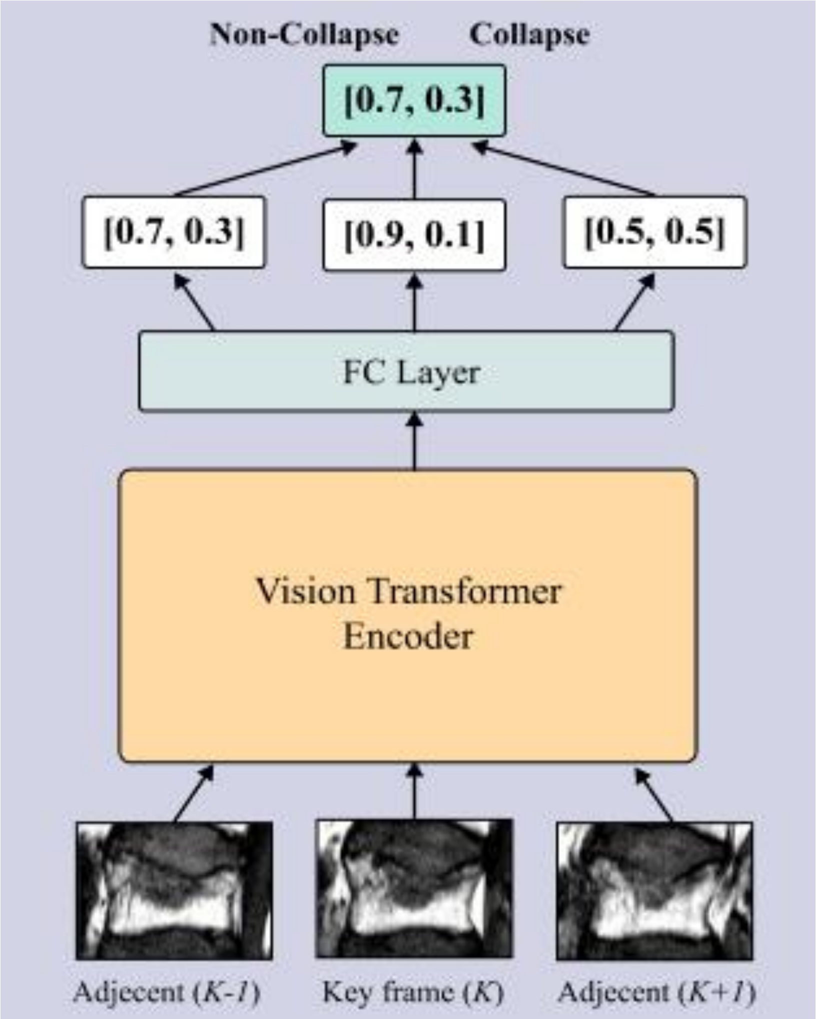
Model Design: Fine-tuning Methods

	Internal Validation			**External Validation**		
	AUC	Spec	Sens	**AUC**	Spec	Sens
ViT PMC	0.8269	0.8057	0.7306	0.8051	0.7296	0.7556
ViT PMC LoRA	**0.8404**	0.8557	0.7012	**0.8113**	0.6963	0.8111

Model Design: Architecture

	# Trainable Parameters	Internal Validation			External Validation		
		AUC	Spec	Sens	**AUC**	Spec	Sens
CNN ImageNet	11.2 M	0.7973	0.7833	0.7123	0.8014	0.7074	0.7833
ViT PMC LoRA	**0.886 M**	**0.8404**	0.8557	0.7012	**0.8113**	0.6963	0.8111

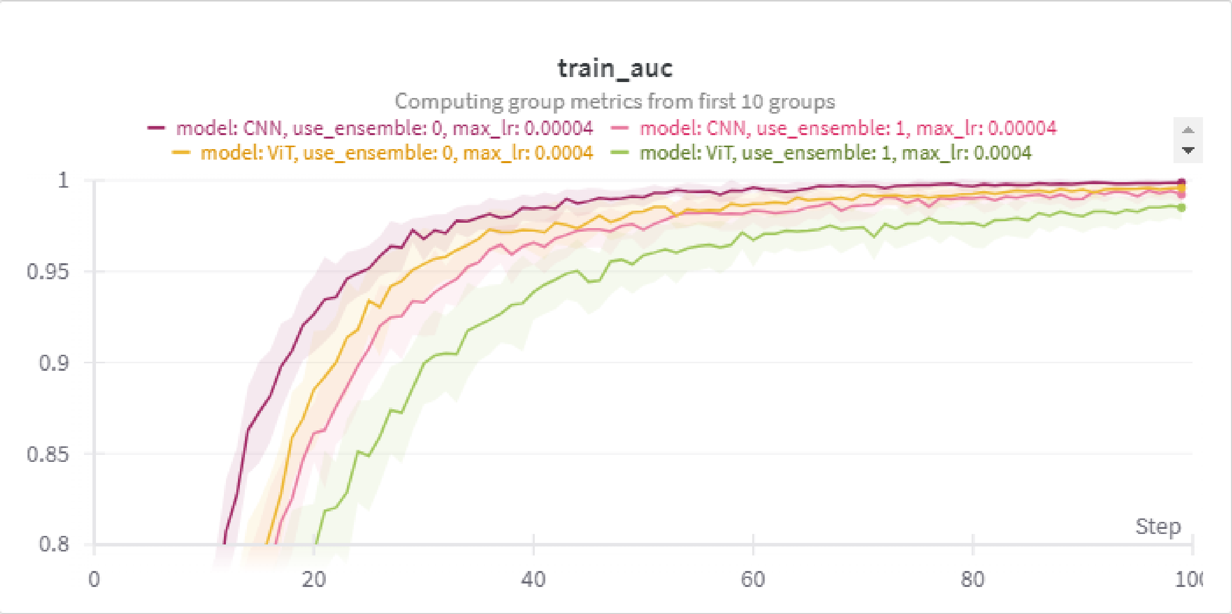
Additional Technique

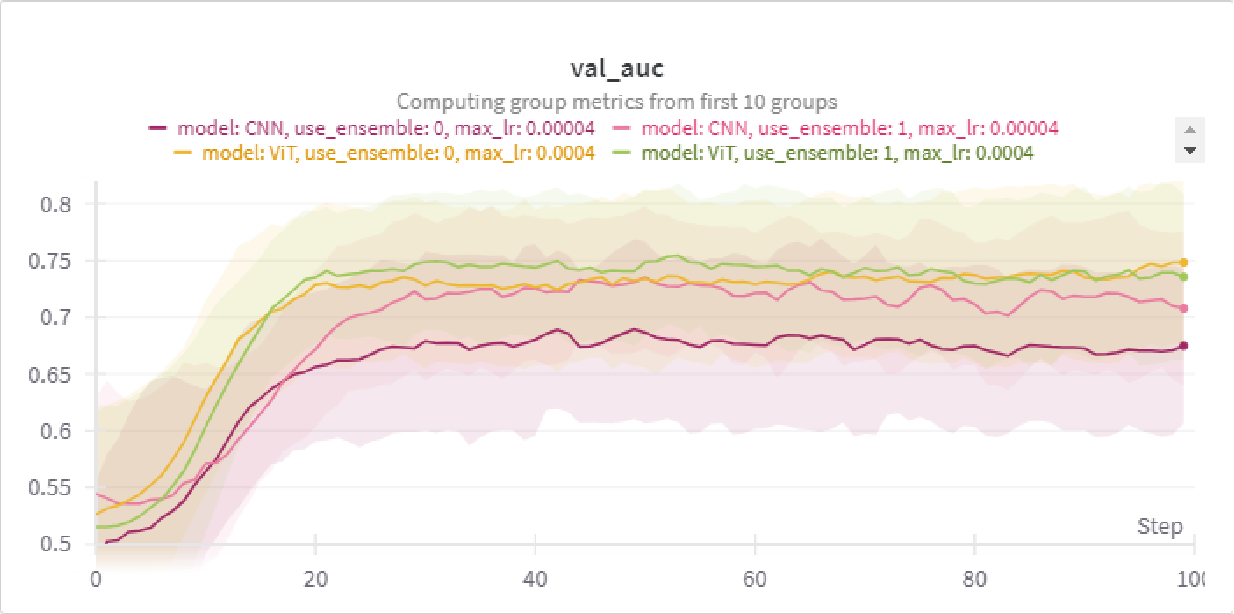


	Internal Validation		
	AUC	Spec	Sens
ViT PMC LoRA	0.8404	0.8557	0.7012
ViT PMC LoRA + ensemble	**0.8539**	0.8230	0.7739

	External Validation		
	AUC	Spec	Sens
ViT PMC LoRA	0.8113	0.6963	0.8111
ViT PMC LoRA + ensemble	**0.8656**	0.8111	0.7611

Hyperparameter Tuning

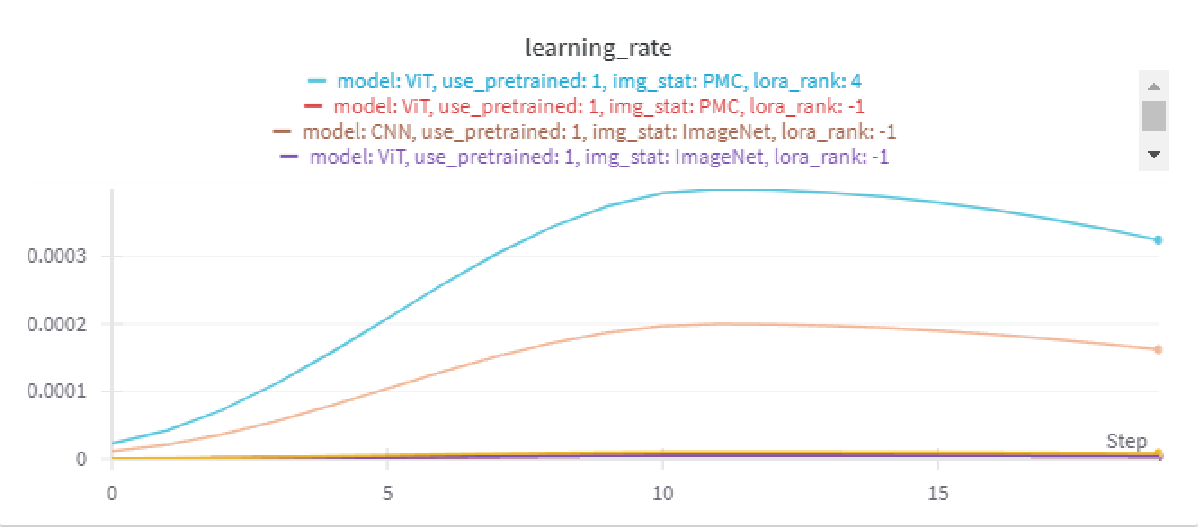


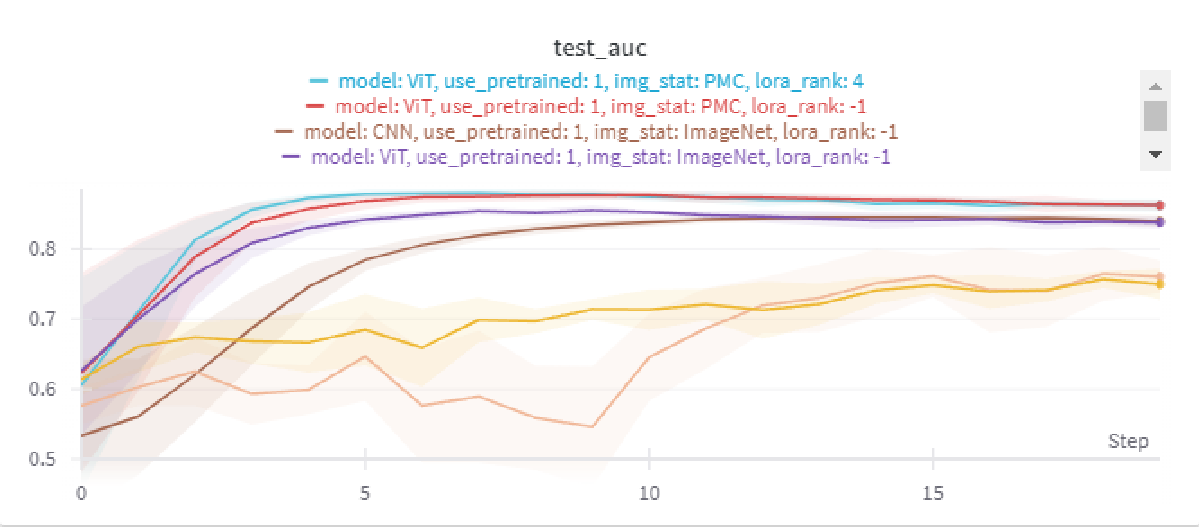


Hyperparameter Tuning

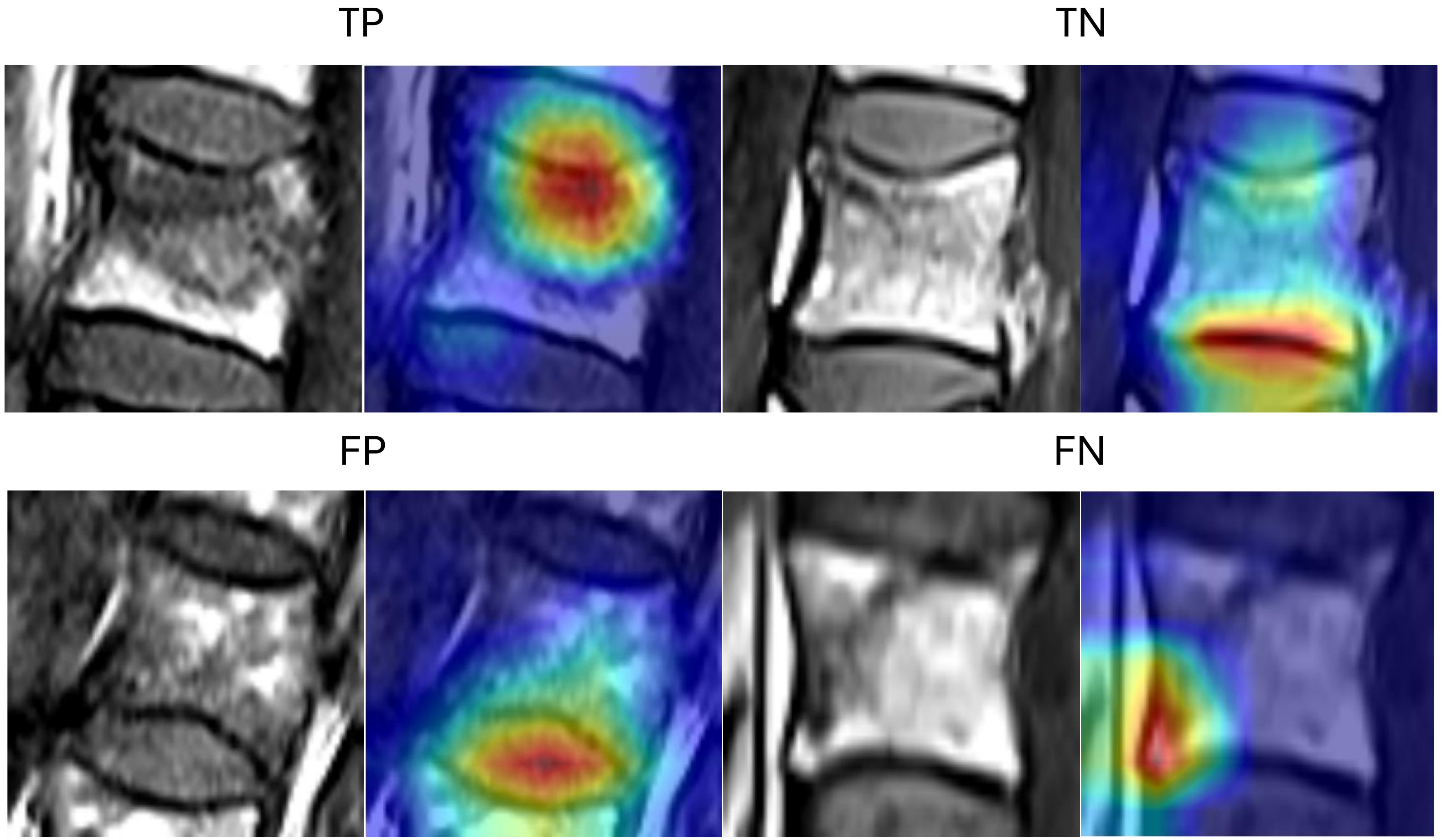
- (1) Short epoch + Broad range fixed learning rate
  - Objective: Identify “possible” range (underfitting X, overfitting X)
- (2) Longer epoch + Narrow range fixed learning rate + Other h.p. tune
  - e.g. weight decay, batch size, lora_rank
  - Objective: Stabilize training… Find good candidates!
- (3) Scheduler selection + Scheduler h.p. tune (e.g. warmup)
  - Objective: Find best setting!
- (4) Excessive epoch + early stop + seed
  - Objective: Find best score!

Hyperparameter Tuning

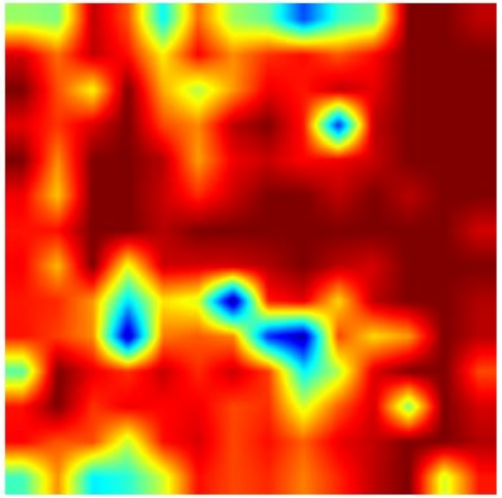
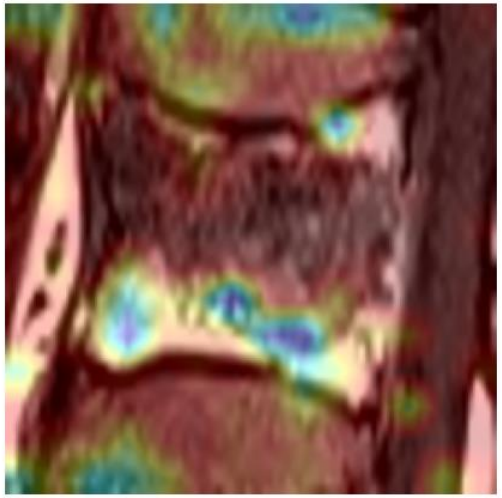
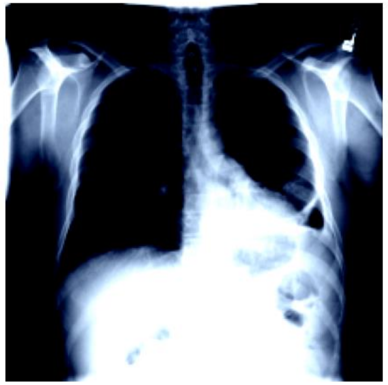
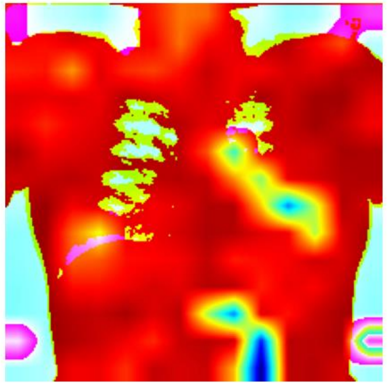




Grad-CAM



Attention Rollout

Key Results

	# Trainable Parameters	Internal Validation			External Validation		
		AUC	Spec	Sens	**AUC**	Spec	Sens
CNN ImageNet	11.2 M	0.7973	0.7833	0.7123	0.8014	0.7074	0.7833
ViT PMC LoRA	**0.886 M**	**0.8404**	0.8557	0.7012	**0.8113**	0.6963	0.8111

	Internal Validation			External Validation		
	AUC	Spec	Sens	**AUC**	Spec	Sens
ViT PMC LoRA	0.8404	0.8557	0.7012	0.8113	0.6963	0.8111
ViT PMC LoRA + ensemble	**0.8539**	0.8230	0.7739	**0.8656**	0.8111	0.7611

Reference

- Yabu, A. *et al*. Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images. *Spine J* **21**, 1652-1658, doi:10.1016/j.spinee.2021.03.006 (2021).
- Tomita, N., Cheung, Y. Y. & Hassanpour, S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med* **98**, 8-15, doi:10.1016/j.compbiomed.2018.05.011 (2018).
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. NeurIPS, 2017.
- Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. In *International Conference on Learning Representations* (ICLR, 2021).
- Zhang, Y., Jiang, H., Miura, Y., Manning, C. D. & Langlotz, C. P. Contrastive learning of medical visual representations from paired images and text. In *Machine Learning for Healthcare Conference* (eds Lipton, Z. et al.) 2–25 (PMLR, 2022).
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*.
- E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.

Thank You!