Evaluation of Different Computer Vision Classification Techniques on Medical Datasets

Group3
Deshmukh, Sharvari
Ma, Dawson
Nguyen, Andy

Problem Statement

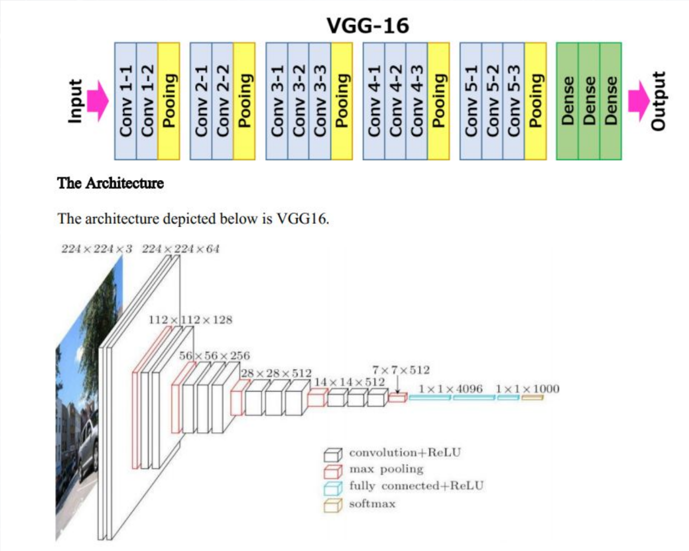
Objective: Identify effective architectures for medical image classification.
Focus: Comparison between Convolutional Neural Networks (CNNs) and Transformers.
Importance: Enhance accuracy and reliability in diagnostic tools.

Why CNNs and Transformers?

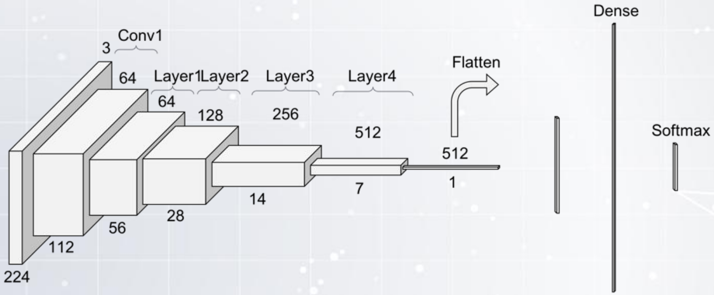
CNNs have been the traditional choice for image data

Transformers were made for NLP, but shows promise in image classification without spatial constraints

CNN

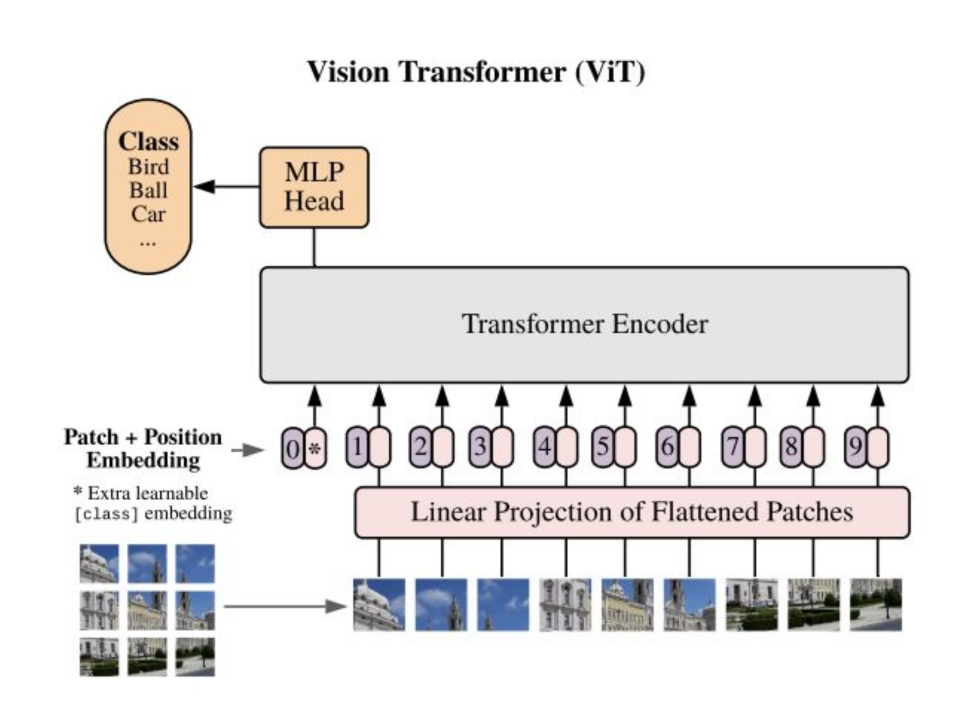


VGG



ResNet

Transformer

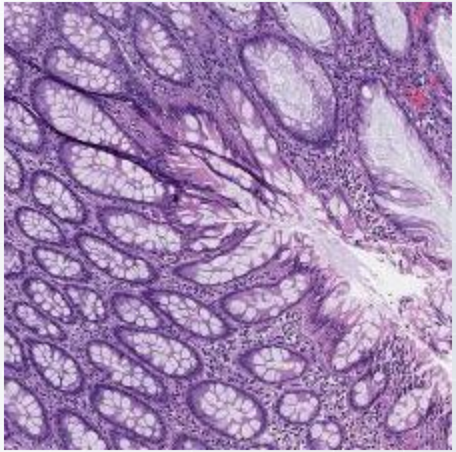


Datasets Used

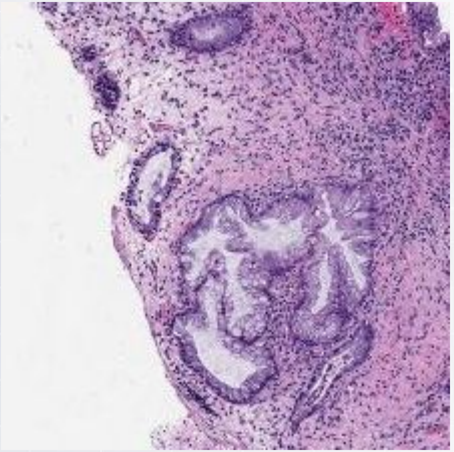
Dataset	Details
MHIST	- Binary classification dataset of classes: Hyperplastic Polyp (HP) & Sessile Serrated Adenoma (SSA) 3,152 fixed-size images of colorectal polyps, each with a gold-standard label determined by the majority vote of seven board-certified gastrointestinal pathologists. 400 MB size with ResNet-18 baseline - Used for Histopathology image classification tasks such as how dataset size, network depth, transfer learning, and high-disagreement examples affect model performance.
LC25000	25,000 color images in 5 classes - Each class contains 5,000 images of the following histologic entities: colon adenocarcinoma, benign colonic tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue. 1.85 GB zip with no current baseline model defined - All images are de-identified, HIPAA compliant, validated

MHIST

(224 x 224 pixels)



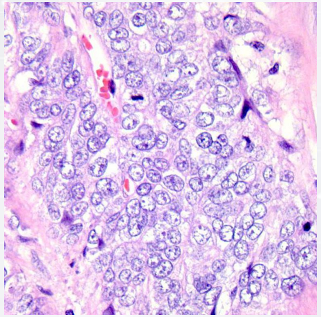
HP



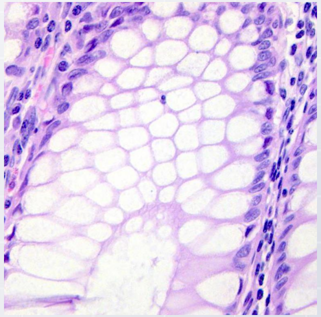
SSA

LC25000

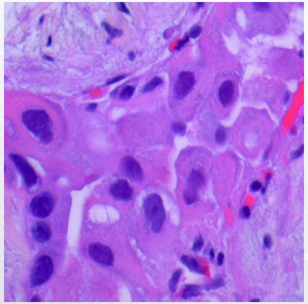
(768 x 768 pixels)



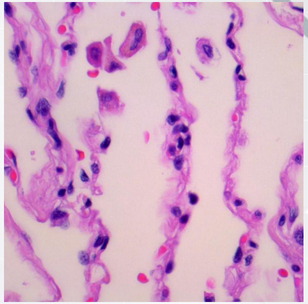
Colon_ACA



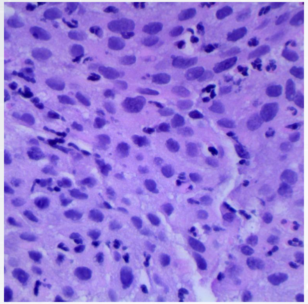
Colon_N



LUNG_ACA



LUNG_N



LUNG_SCC

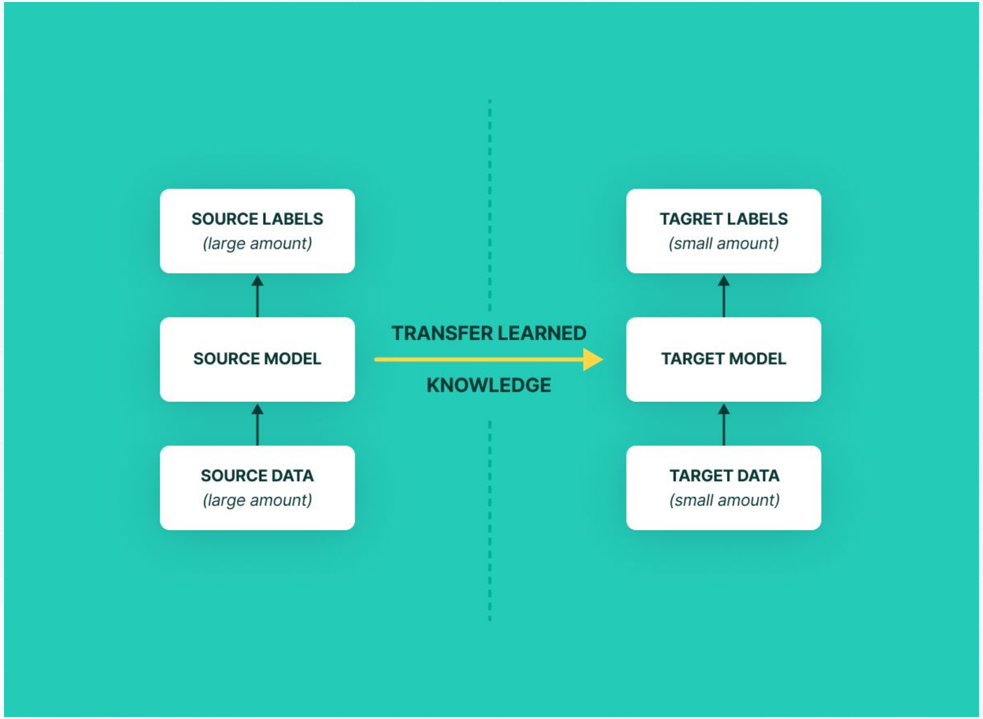
Problems specific to Classification of Medical Data

- Dataset Size, Quality and Scarcity
- Annotation and Labeling
- Inter- and Intra-Class Variability
- Transferability and Generalization
- Robustness to Noise and Artifacts
- Ethical and Legal Concerns

Proposed solutions

- Leveraging Pretrained Weights
- Dataset Augmentation Techniques
- Implementation of Fine-grained models
- Hybrid architectural approach

Transfer Learning



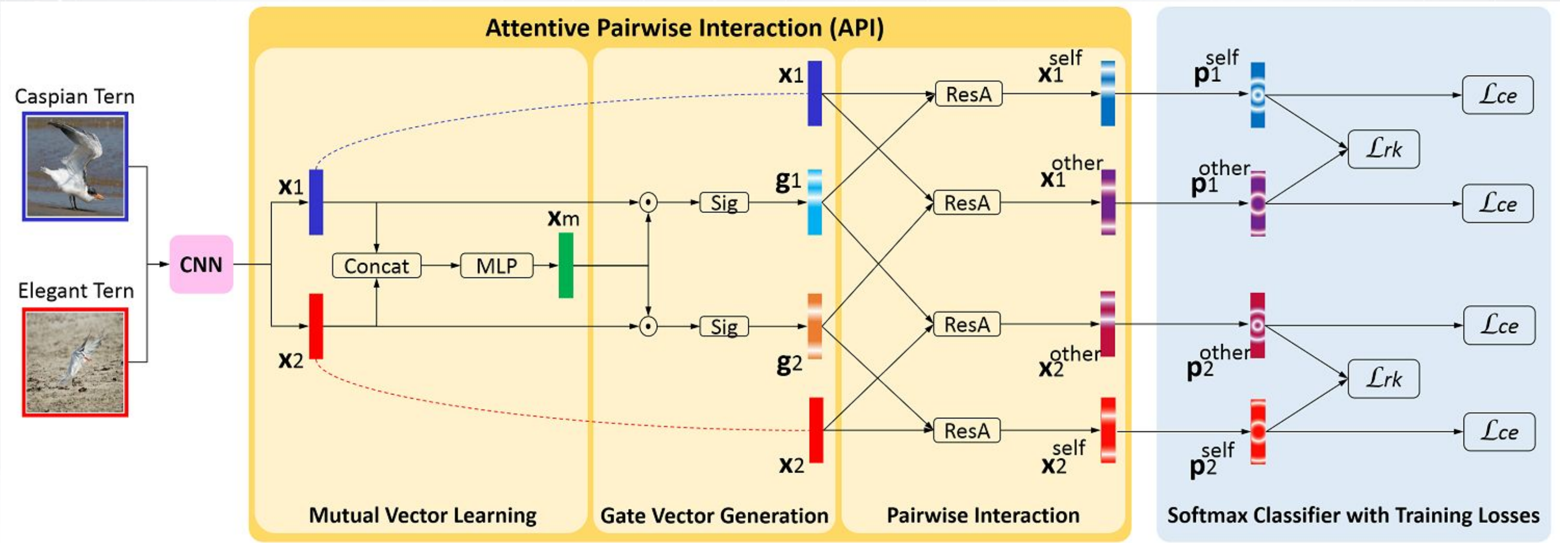
Augmentation Techniques

RandomRotation

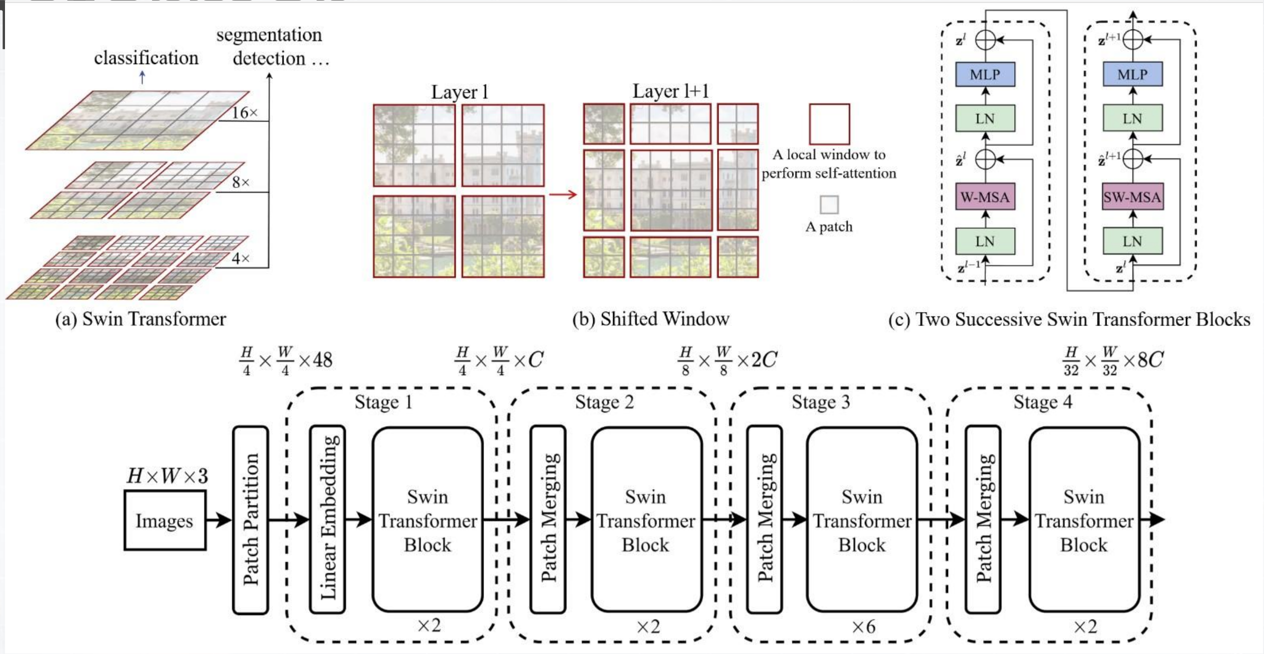
RandomHorizontalFlip

RandomVerticalFlip

Fine-Grain Model – API Net



Fine-Grain Model – Swin tran



Experiments

- **Architectures**
  - CNNs: VGG16, ResNet18, API-Net
  - Transformers: ViT, Swin-T
- **Optimizers**
  - Adam
  - SGD
- **Optimizer parameters**
  - Learning Rate
  - Betas (running averages of gradient)
  - Weight Decay
  - Momentum
- **Datasets**
  - MHIST
  - LC25000
- **Augmentation techniques**
  - Random rotation angle, Probability of random horizontal/vertical flip

Evaluation Metrics

Precision: measures accuracy of positive predictions
Recall/Sensitivity: ability of model to identify relevant instances
Test Accuracy: overall correctness of unseen data
F1 scores: mean of precision and recall (1 = perfect precision and recall
0 = worst)

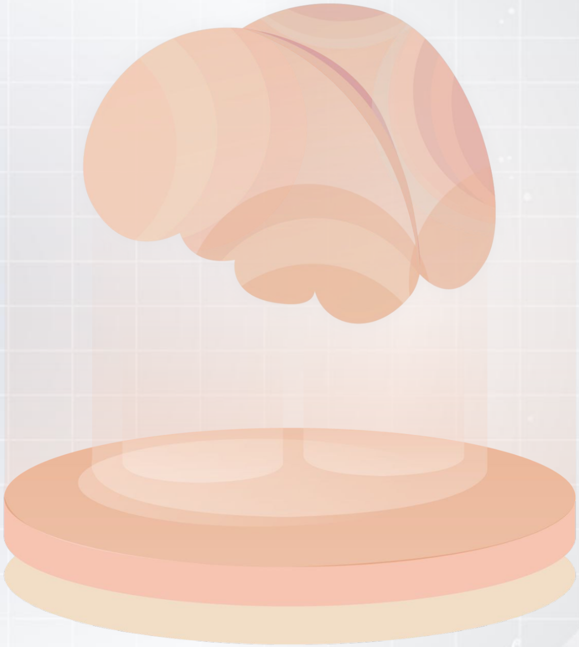
ACTUAL VALUES	POSITIVE	NEGATIVE
POSITIVE	TP	FN
NEGATIVE	FP	TN

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Results



CNN Architectures on MHIST

Model	VGG16		ResNet18		VGG16		ResNet18		API-Net	
Pretrained	X		X		O		O		O	
Augmentation	X	O	X	O	X	O	X	O	X	O
Accuracy	71.14	73.39	67.35	77.07	84.44	85.36	83.52	86.18	82.19	**86.90**
Precision	56.75	62.43	47.96	60.00	78.47	78.94	75.27	79.58	76.49	**81.87**
F1-score	51.39	60.00	40.83	46.67	76.94	74.44	68.06	73.06	78.61	**80.28**
Recall	63.36	65.06	58.10	84.00	80.06	84.01	84.19	87.38	74.47	**83.53**

# of samples: 3125

CNN Architectures on LC25000

Model	VGG16		ResNet18		VGG16		ResNet18		API-Net	
Pretrained	X		X		O		O		O	
Augmentation	X	O	X	O	X	O	X	O	X	O
Accuracy	96.94	97.02	96.92	96.98	97.90	99.42	99.60	99.76	99.71	**99.80**
Precision	96.94	97.02	96.92	96.98	97.90	99.42	99.60	99.76	99.71	**99.80**
F1-score	96.94	97.02	96.92	96.98	97.90	99.42	99.60	99.76	99.71	**99.80**
Recall	96.94	97.02	96.92	96.98	97.90	99.42	99.60	99.76	99.71	**99.80**

# of samples: 25,000

Transformer Architectures on MHIST

Model	ViT		ViT		Swin-T		Swin-T	
Pretrained	X		O		X		O	
Augmentation	X	O	X	O	X	O	X	O
Accuracy	62.2	63.13	81.88	79.30	62.20	62.20	**83.08**	80.08
Precision	48.08	49.89	73.89	74.07	76.73	76.73	**82.98**	80.01
F1-score	64.94	66.32	76.18	70.61	48.89	48.89	**82.94**	79.37
Recall	100.0	62.41	78.61	67.46	63.15	63.15	**82.91**	80.04

# of samples: 3125

Transformer Architectures on LC25000

Model	ViT		ViT		Swin-T		Swin-T	
Pretrained	X		O		X		O	
Augmentation	X	O	X	O	X	O	X	O
Accuracy	65.48	70.36	99.00	97.34	92.75	92.22	**99.94**	99.28
Precision	66.50	68.54	99.80	99.90	90.83	92.48	**99.94**	99.28
F1-score	68.12	73.62	99.9	99.60	90.46	92.34	**99.94**	99.28
Recall	69.82	79.52	1.00	99.30	90.10	92.36	**99.94**	99.28

# of samples: 25000

Conclusion

- **Impact of Augmentation**
  - Significantly improved CNNs
  - Minimal improvement in Transformers
- **Transfer Learning Benefits**
  - Improved model performance across both architectures
- **Fine Grained Models**
  - The implementation only resulted in slight improvements in performance which shows that it can distinguished small difference in classes, but effectiveness is affected by size and quality of dataset

Future work

- **Expanding Dataset**
  - Implement same experiments with a larger medical data set (>100,000 samples)
- **Explore Meta-Learning**
  - Enables a model's ability to learn from a limited number of examples
- **Leveraging Domain-Specific Pretrained Models**
  - Models pretrained on similar tasks can significantly improve performance due to relevance of similar features.
- **Hybrid Architecture**
  - Develop U-Net structure that combines CNN + Transformer to utilize their strengths of spatial hierarchies and long range dependencies

THANKS!