Automated Diagnosis of Thorax Diseases using Chest X-ray Images

ML group 8

Yuting(Eva) Gu

Sijun(Sunny) Zhang

Saifullah(Saif) Soliman

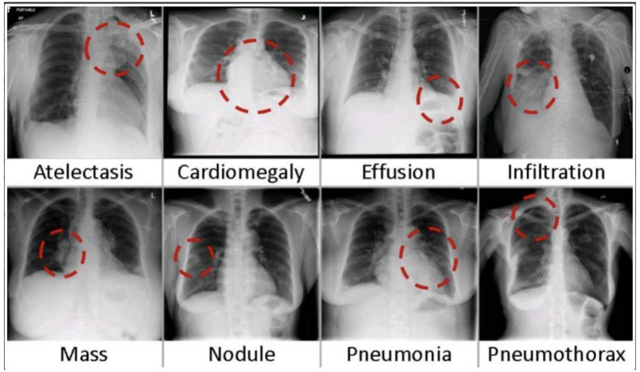
Outline

1. **Introduction**
2. **Method**
    - Data Acquisition & Processing
    - Unsupervised
        - Autoencoder & Clustering
    - Supervised
        - ResNet50
        - ViT
3. **Result**
4. **Discussion**

Introduction

Background

- ChestX-ray14 database contain 112,120 frontal-view X-ray images annotated with 14 thoracic disease labels
- Valuable resource for training and testing multiclass classification models for thoracic disease diagnostics
- Multiclass classification tasks were performed using different CNN models, with ResNet50 emerging as the best-performing model
- Vision transformers (ViTs) have recently emerged as a superior approach, surpassing CNNs in performance for image classification tasks



Objectives

1. Conduct a comparative analysis between the ResNet50 and ViT models for supervised multi-class classification tasks using ChestX-ray14 database
2. Explore an unsupervised approach by applying an autoencoder method to extract latent representations from the ChestX-ray14 images.
3. Employ autoencoder and clustering techniques on ChestX-ray14 images to unveil latent patterns and groupings within the dataset
4. Gain a comprehensive understanding of the strengths and limitations of each approach in thoracic disease identification and image analysis tasks.

Methodology

Data Source & Data Acquisition

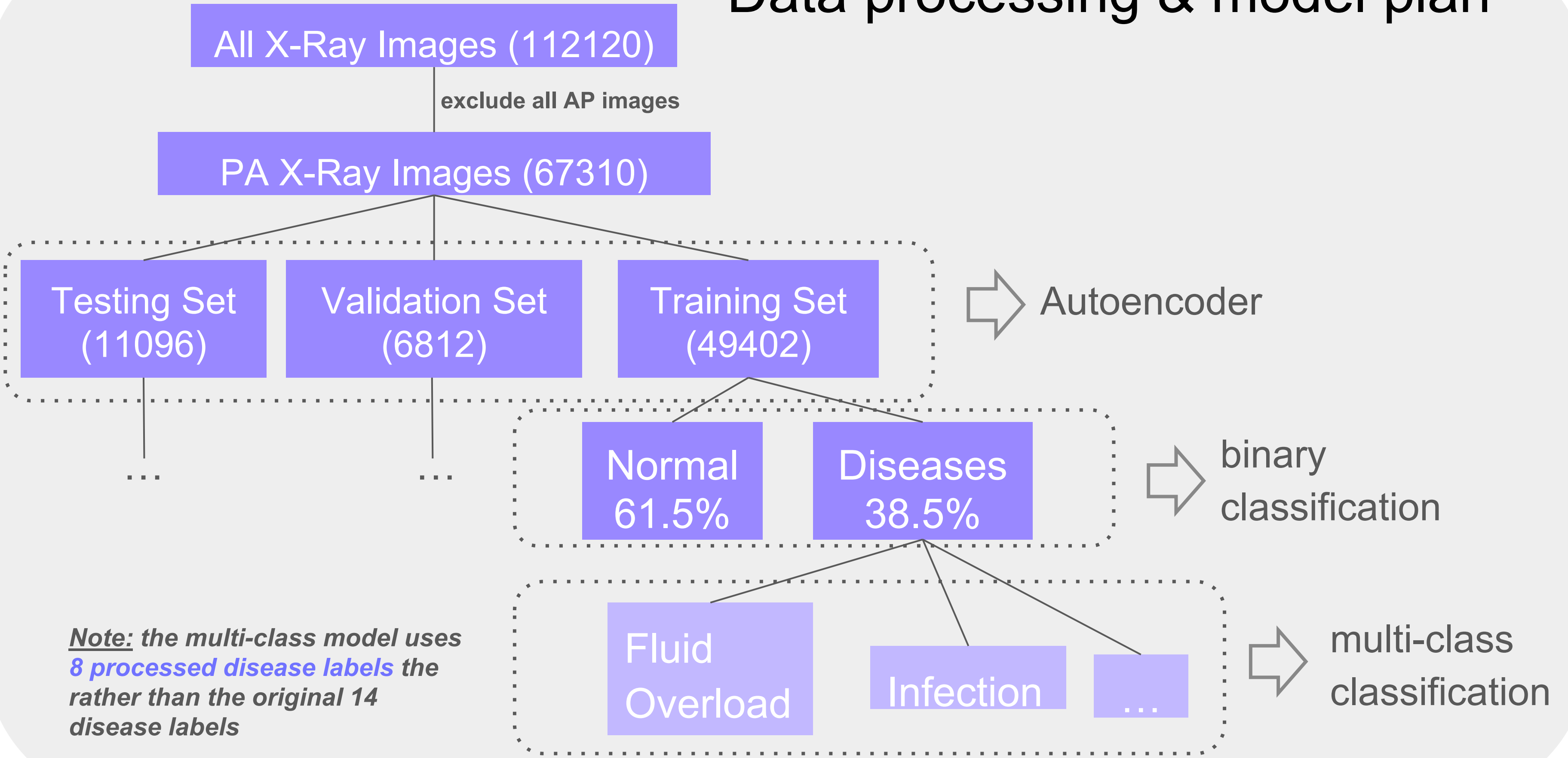
Data Source:

ChestX-ray14, which contains frontal-view X-ray images of 30,805 patients with normal chest condition or with one or multiple of 14 thoracic issues (1992-2015)

Data Acquisition:

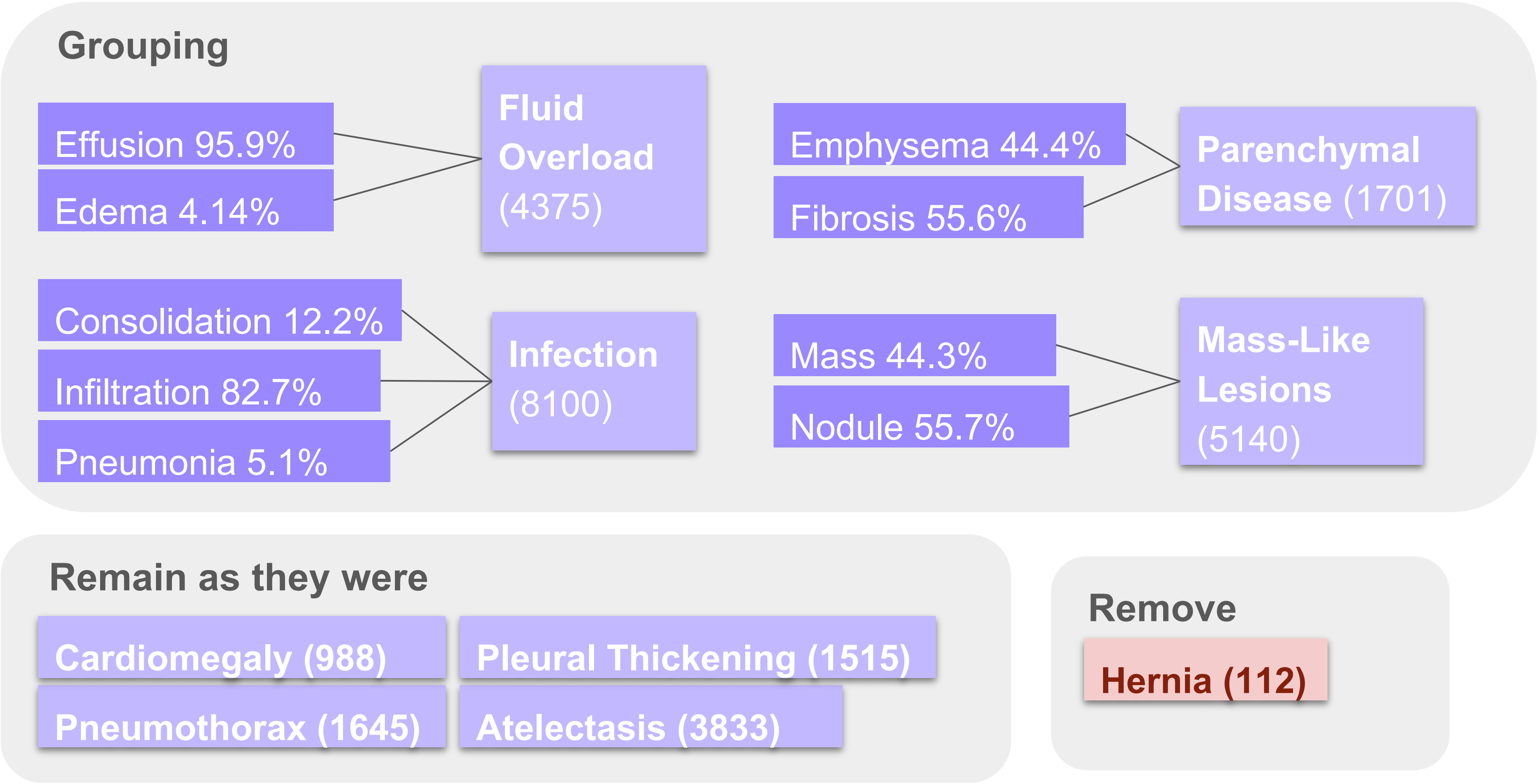
Images and labels could be downloaded from https://paperswithcode.com/dataset/chestx-ray14

Data processing & model plan

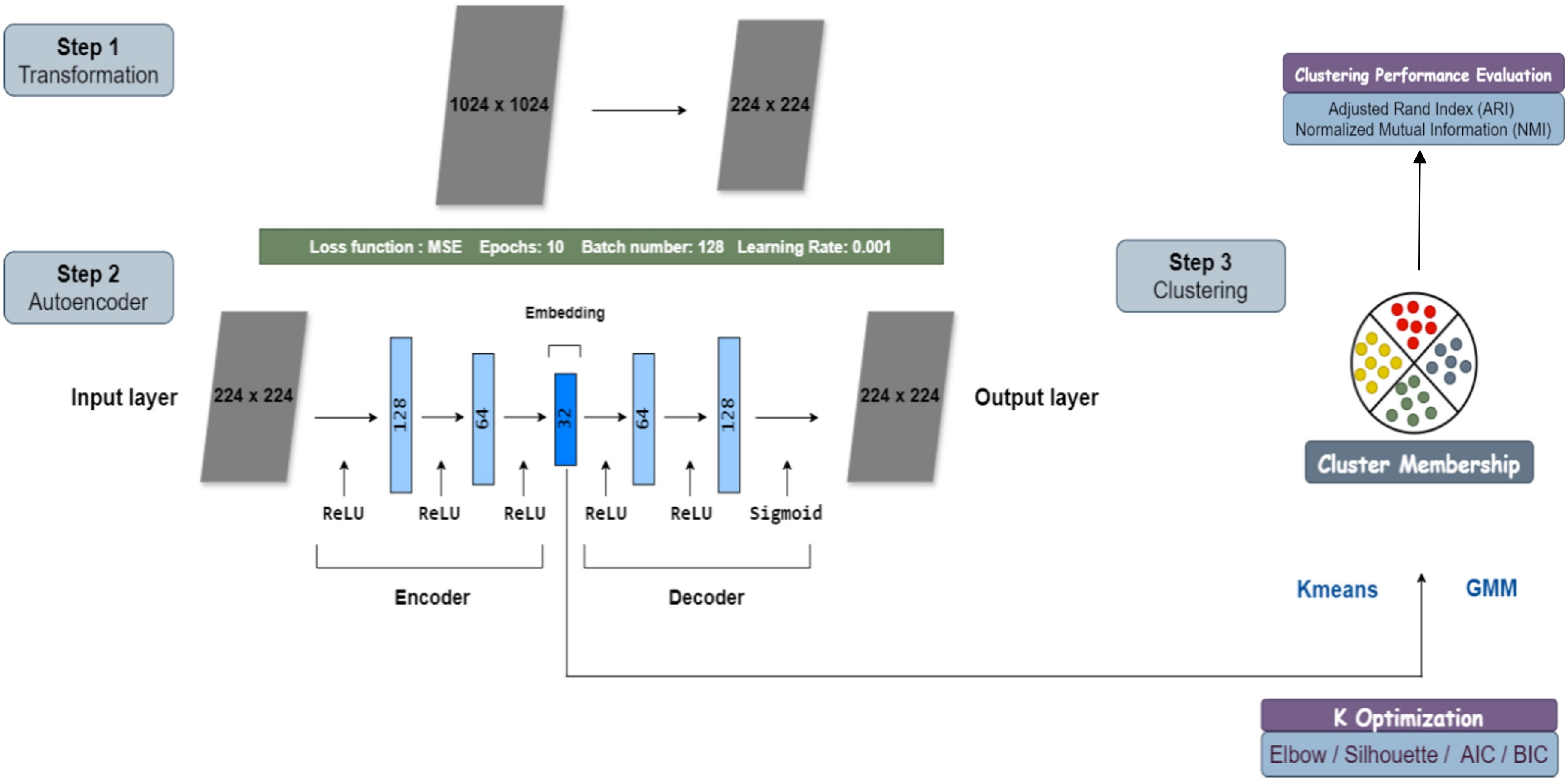


14 original disease labels ⇨ 8 new disease labels

*__Note:__ the numbers are for the **training set***



Unsupervised - Autoencoder & Clustering



Supervised - ResNet 50

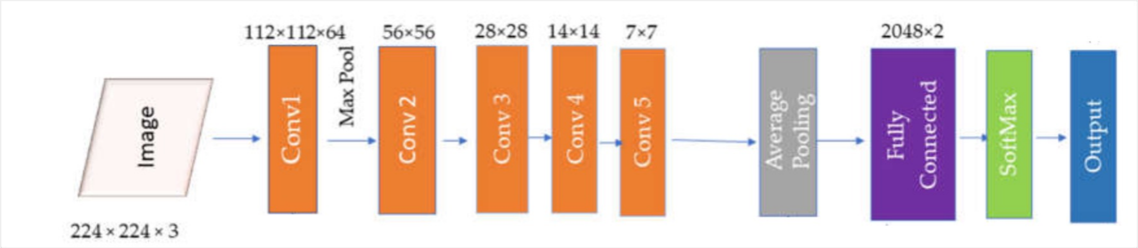
ResNet50 - a 50-layer convolutional network for effective image classification, published in 2015.

Based on the general CNN structure - convolutional layers, pooling layers etc.

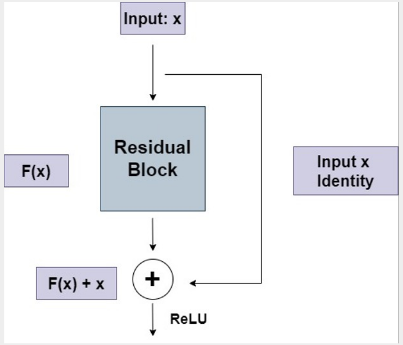
What's special -

- ResNet50 incorporates **residual blocks** where layers **learn additional features** from the input.
- These enhancements are **added back to the block's original input** to generate the block's output.
- While preserving the deep network structure, it facilitates a stable gradient flow and **prevents gradients from vanishing**.

General ResNet50 structure

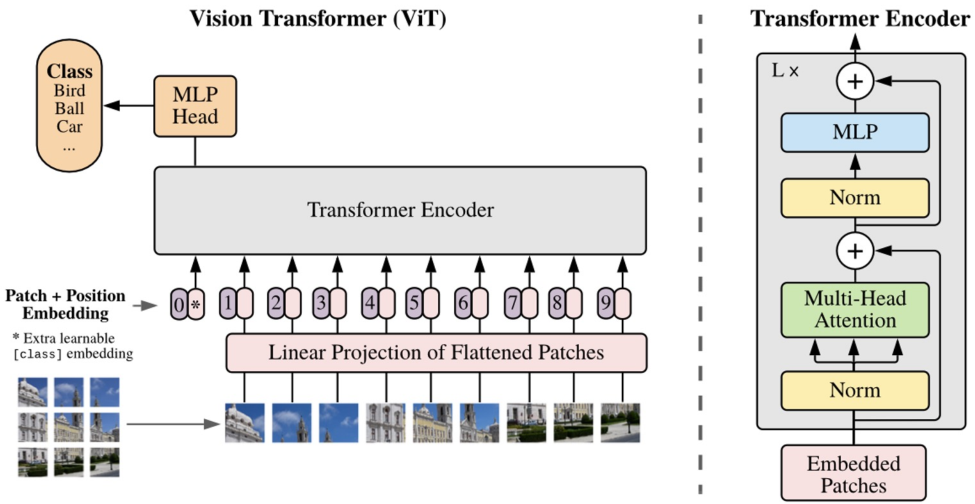


the Residual Block



Supervised - Vision Transformer (ViT)

- SOTA method for image classification task
- Transformer was originally designed for NLP task, then the idea was adapted for computer vision task in 2020 by **treating images as a sequence of patches**.
- **Not a RNN** model, use self-attention mechanism to capture spatial relationship between patches without sequential processing



Our Model

- base ViT with 16 patch size
- 1024 -> 224 (input size of ViT)
- number of classes:
  - binary (Normal vs Diseases)
  - grouped multiple classes

Supervised - Performance Metrics

- **Binary Task:**
  - Accuracy, Recall (Sensitivity), Precision, Specificity, AUC, F1 Score

- **Multi-Class Task:**
  - One vs Rest: AUC, Precision, Recall, F1 score

Supervised - Transfer Learning

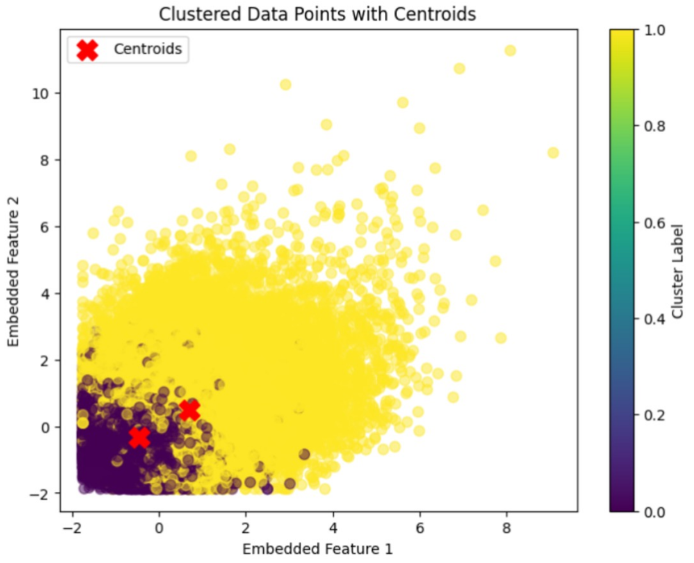
- The **pre-trained** ResNet50 / ViT model has learnt the **generic features** in image classification task. **Pre-trained Dataset:**
  - ResNet50 - on **ImageNet-1k**[1]
  - ViT - on **ImageNet-21k**[1]
- Therefore, we **re-use** them as a **starting point** and **modify** the **model parameters** with the **ChestX-ray14** data. Fine-tune our model for **30 epochs.**
  - Change the output layer for binary/multiclass outcome
- **Pre-trained parameters** are obtained from open source package **timm**[2,3].

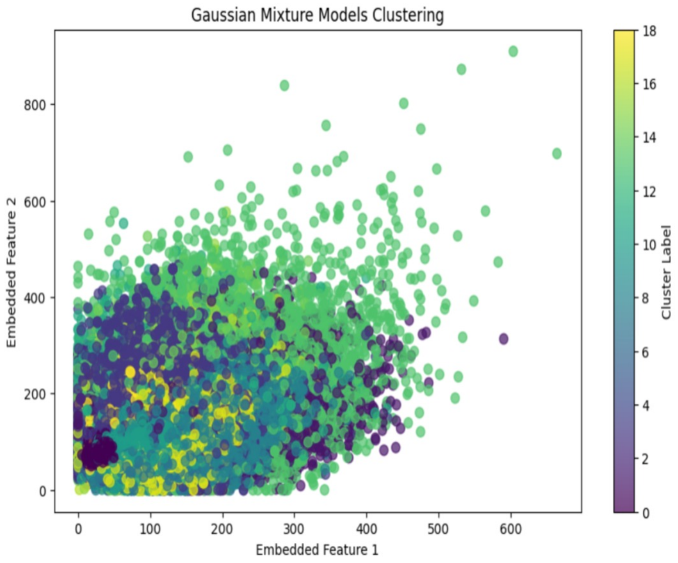
[1] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115: 211-252.
[2] https://huggingface.co/timm/resnet50.a1_in1k
[3] https://huggingface.co/timm/vit_base_patch16_224.orig_in21k

Results

Autoencoder & Clustering

Binary Classification:



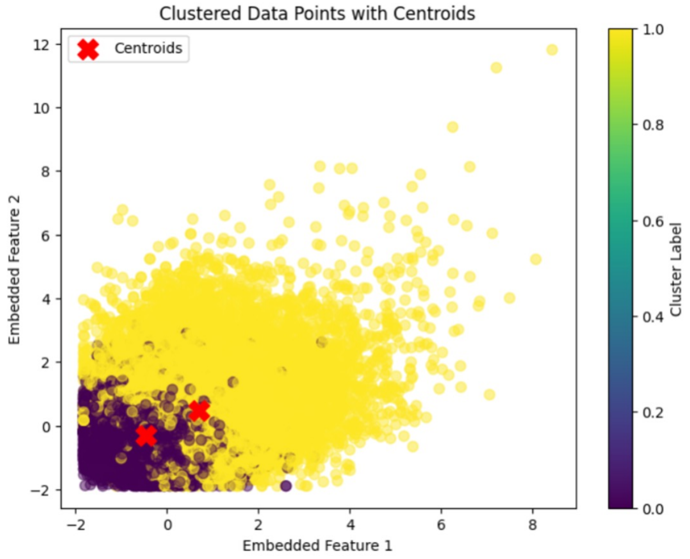


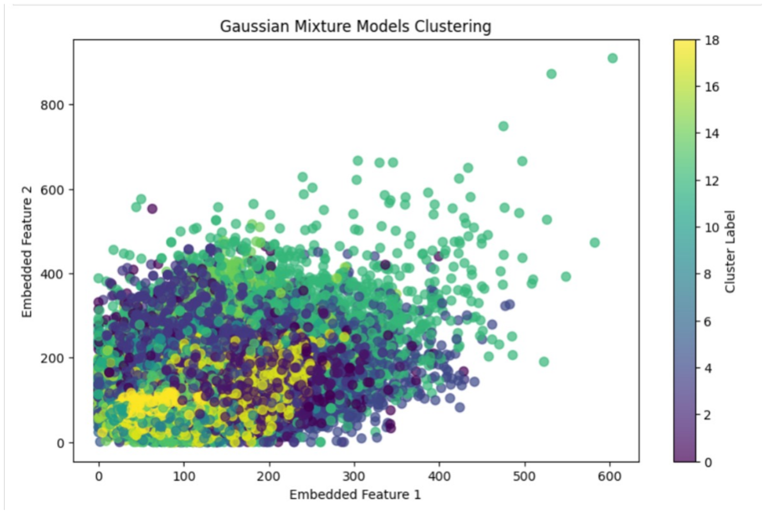
Optimal # of clusters = 2
Adjusted Rand Index (ARI) = 0.01
Normalized Mutual Information (NMI) = 0.00

Optimal # of clusters = 19
Adjusted Rand Index (ARI) = 0.01
Normalized Mutual Information (NMI) = 0.01

Autoencoder & Clustering

Multi-Class Classification:





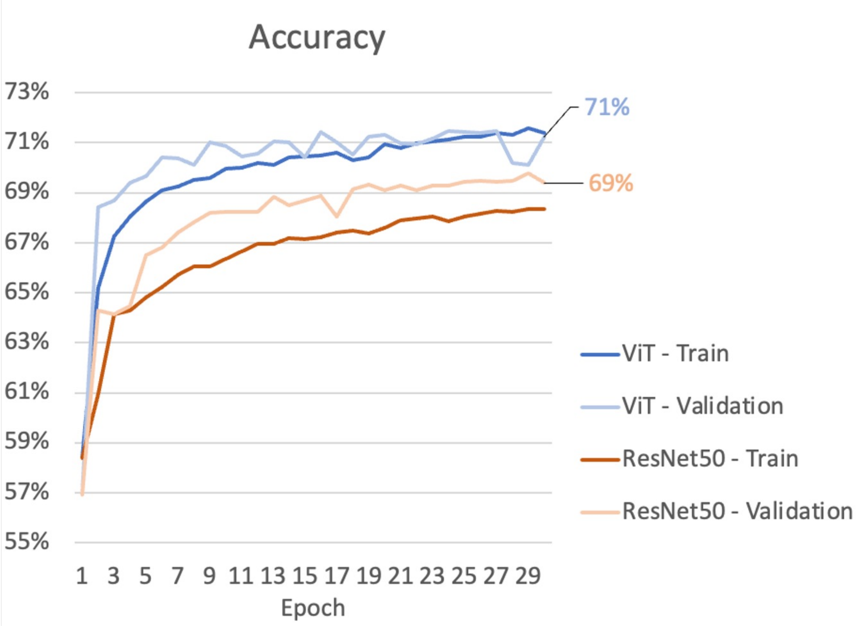
Optimal # of clusters = 2
Adjusted Rand Index (ARI) = 0.01
Normalized Mutual Information (NMI) = 0.002

Optimal # of clusters = 19
Adjusted Rand Index (ARI) = 0.003
Normalized Mutual Information (NMI) = 0.006

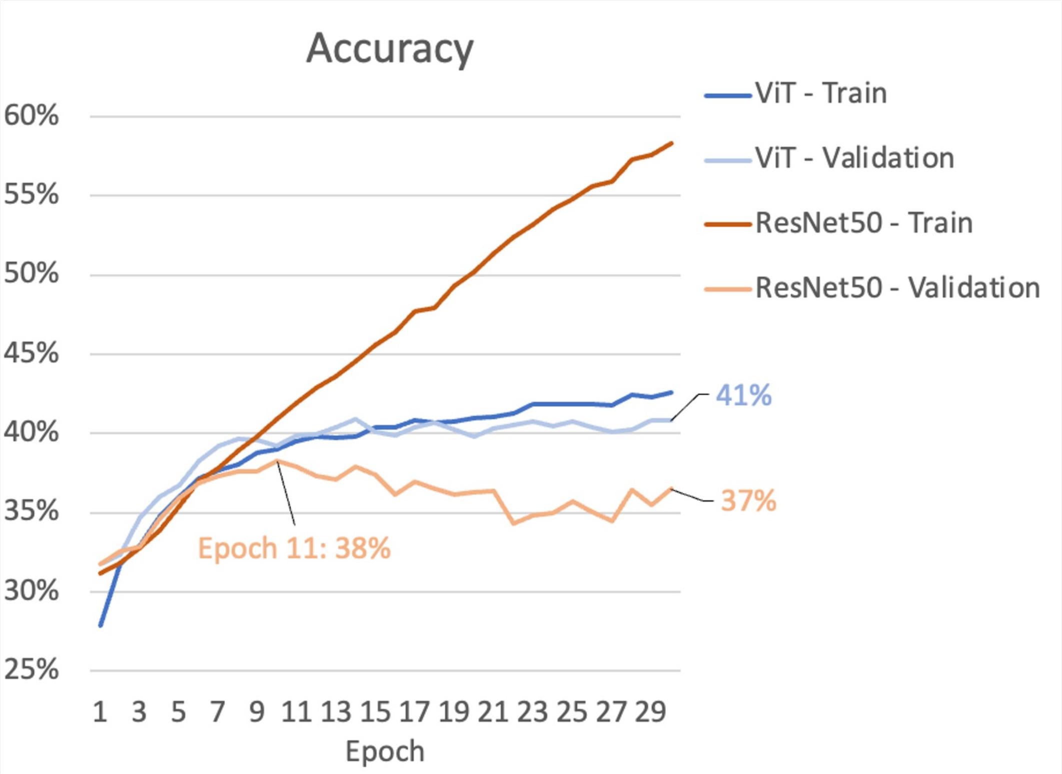
Binary Classification - ResNet50 & ViT

Test Set Performance

	ViT	ResNet50
Accuracy	**0.70**	0.68
AUC	**0.67**	0.66
Recall/Sensitivity	**0.77**	0.77
Precision	**0.74**	0.73
F1-score	**0.76**	0.75
Specificity	**0.58**	0.54



Multiclass Classification - ResNet50 & ViT



Multiclass Classification - ResNet50 & ViT

	ViT				ResNet50			
Accuracy	0.38				0.34			
	Precision	Recall	F1-score	AUC	precision	recall	F1-score	AUC
Fluid overload	0.35	0.38	0.37	0.60	0.35	0.23	0.28	0.56
Infection	**0.41**	**0.59**	**0.48**	0.61	**0.38**	**0.62**	**0.47**	0.59
Mass like Lesions	0.39	0.29	0.33	0.61	0.26	0.42	0.32	0.63
Parenchymal Disease	0.30	0.28	0.29	0.62	0.18	0.08	0.11	0.53
Atelectasis	0.32	0.26	0.29	0.59	0.32	0.18	0.23	0.56
Cardiomegaly	0.39	0.37	0.38	**0.67**	0.35	0.29	0.32	**0.63**
Pneumothorax	0.39	0.18	0.25	0.57	0.32	0.14	0.2	0.55
Pleural Thickening	**0.21**	**0.04**	**0.06**	**0.52**	**0.13**	**0.01**	**0.02**	**0.50**

Discussions & Conclusions

Discussion & Conclusions

- **Unsupervised** Autoencoder and clustering **did not achieve a desirable result**
  - Encoder architecture is not performing well in feature selection
- **ViT** has **better performance** compared to ResNET50 in both binary and multiclass classification of thoracic diseases - inline with current literature
- Both models are **underperforming** in detecting **true negative cases**, yet **ViT** can **better** detect them
- Overall performance of both models in multiclass classification **needs a lot of improvement**
  - can be owed to that images have multiple diagnoses
- **Cardiomegaly** - though it has <1000 training images, it has the best AUC among all disease groups.
  - Possible reason: “an enlarged heart” is relatively easier to detect compared to other diseases

Limitations & Future Direction

Limitation:

- Dataset is not large enough.
  - did not apply **data augmentation** since X-ray is **not rotatable or shiftable**

Future Work:

1. **Enhance encoder architecture** to better perform feature extraction
2. Include a **weighted loss function** to handle the difference in sample size in the multiclass groupings
3. **Compile multiple X-ray datasets** to overcome the **low sample numbers** in some of the classes
4. Try **no-pretrained model** if have larger dataset

Thank you!

Appendix 1

More background information

- Computer vision - field of AI - enables computers to interpret and analyze visual information
- In healthcare, it involves the use of algorithms to analyze medical images, aiding in clinical diagnosis & decision-making
    - Interpretation of X-rays, CT scans & MRIs
    - Detection and localization of abnormalities
- Multiple advantages in medical diagnosis
    - Enhanced accuracy
    - Efficiency
    - Scalability

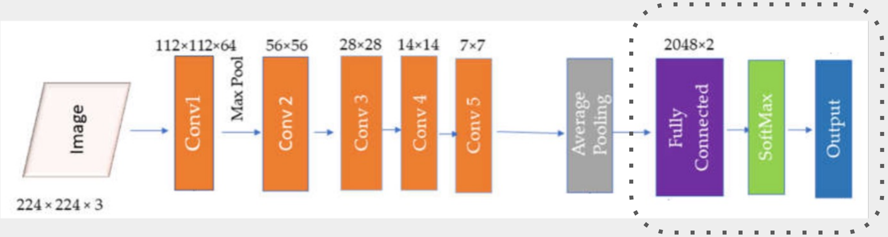
Appendix 2

Transfer Learning for ResNet50

Modified ResNet50

- Based on ResNet50 pre-trained on the **ImageNet-1k** dataset.
- From python package **`timm`**, including enhancements that boost accuracy and efficiency.

modify to match the binary class/ 8 disease classes



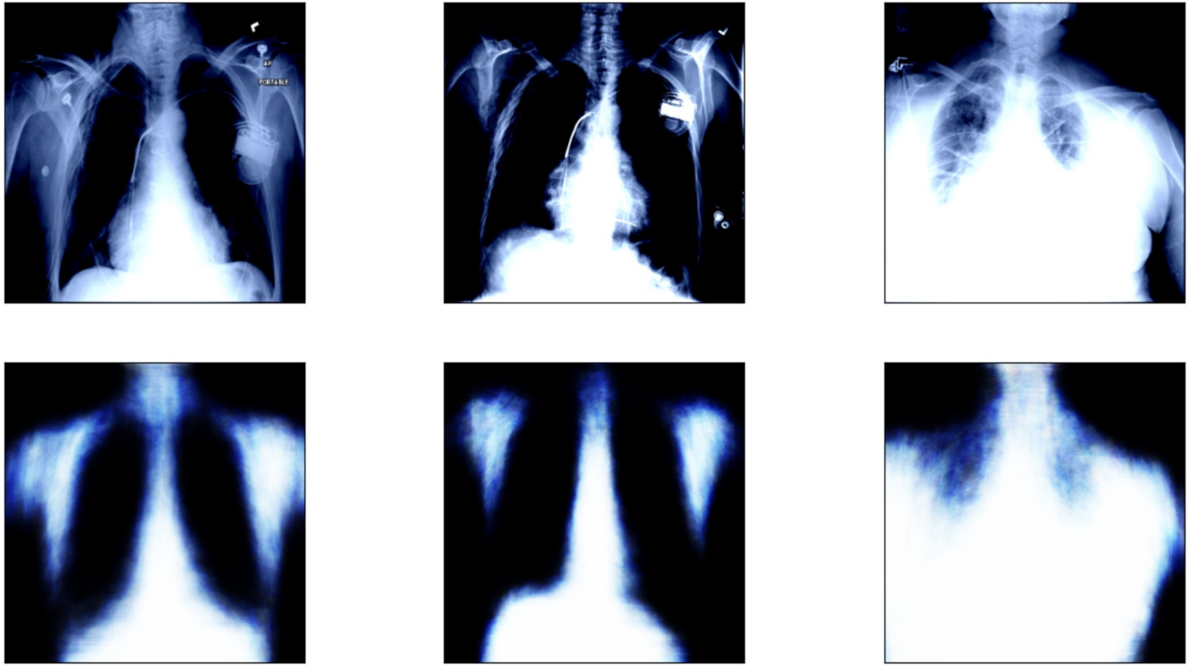
Appendix 3

ResNet50 & ViT Hyperparameters

Batch size: 128
Epoch: 30
Learning rate: 0.01

Appendix 4

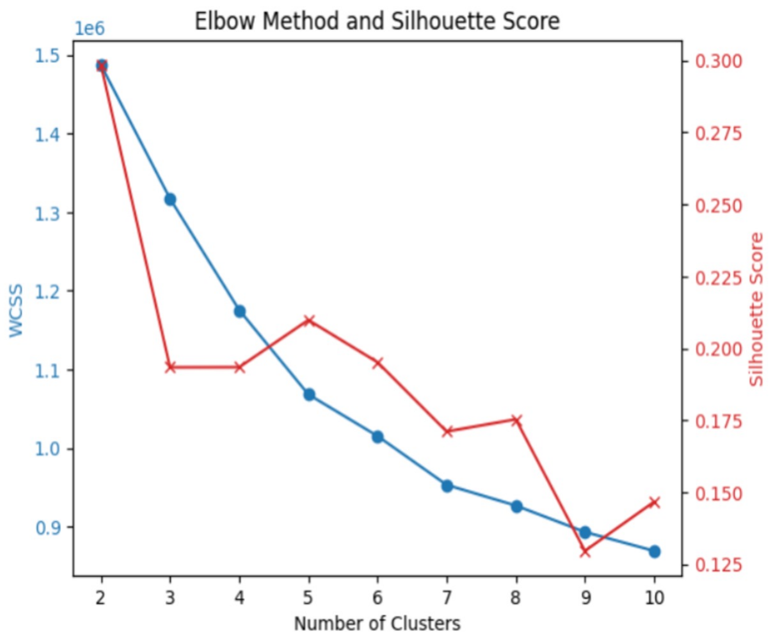
Result: Autoencoder Reconstruction



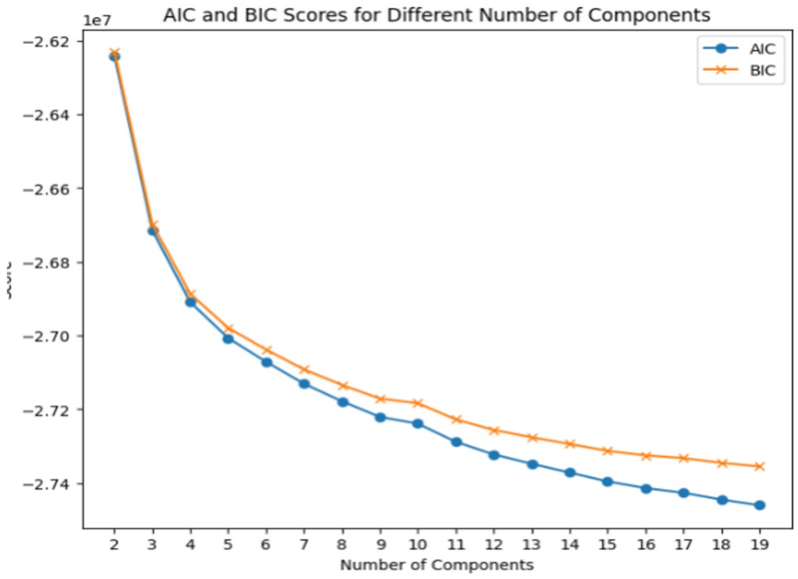
Appendix 5

Clustering Calibrition

Binary Classification: Kmeans



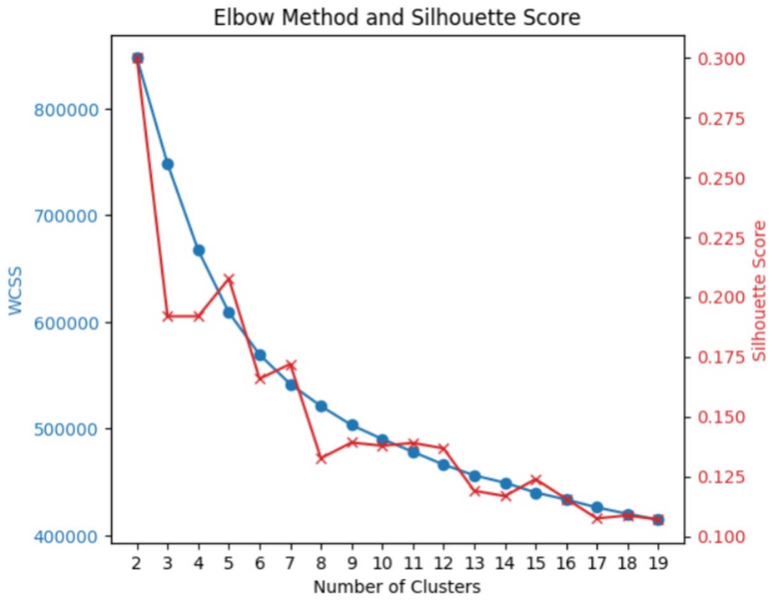
Binary Classification: GMM



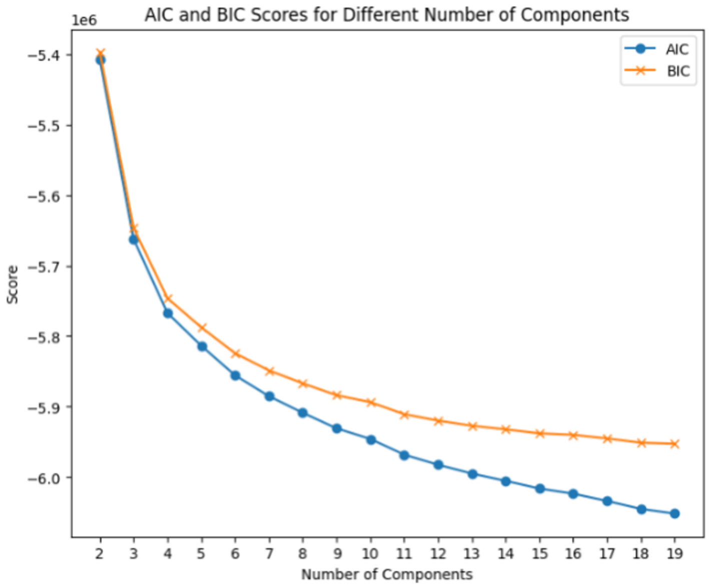
Appendix 5

Clustering Calibrition

Multi-Class Classification: Kmeans



Multi-Class Classification: GMM



Appendix 6

Multiple diagnosis situation for each image