胸部X光影像之多標籤深度學習分類

MULTI-LABEL DEEP LEARNING CLASSIFICATION OF CHEST X-RAYS

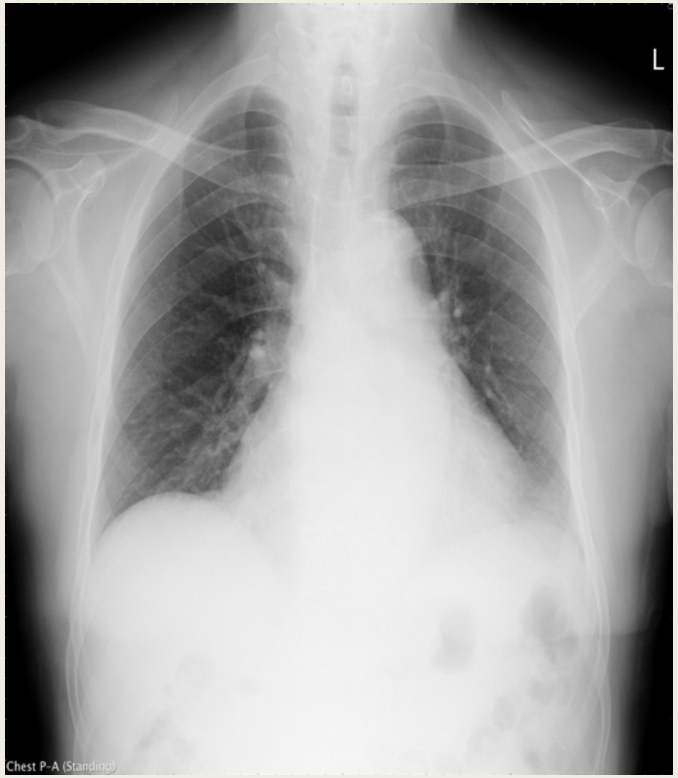
古明章、傅琦佳、黃冠華
國立陽明交通大學 統計學研究所
2021/10/30

胸部X光影像之多標籤遷移學習分類

MULTI-LABEL DEEP LEARNING CLASSIFICATION OF CHEST X-RAYS

古明章、傅琦佳、黃冠華
國立陽明交通大學 統計學研究所
2021/10/30

Image classification

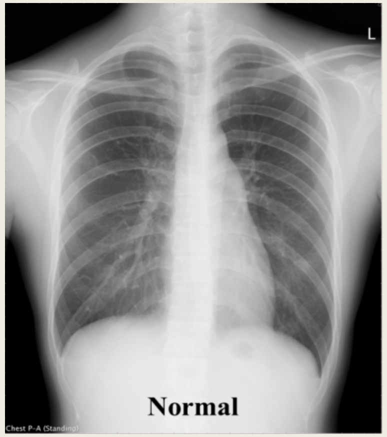


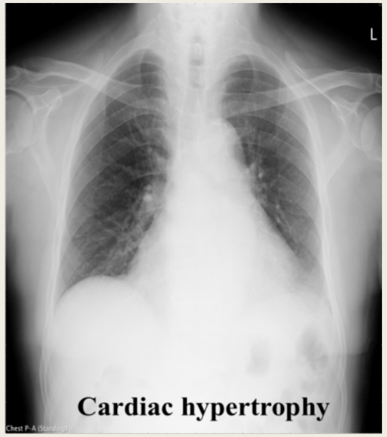
Assume given set of pathology categories (labels):

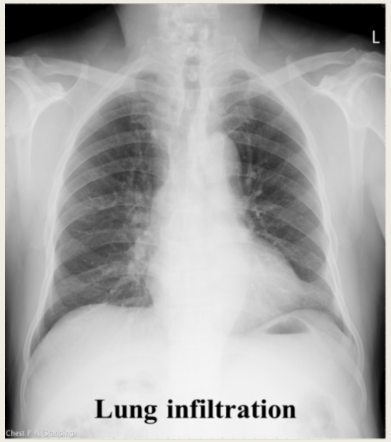
{normal, cardiac hypertrophy, aortic sclerosis, lung infiltration, ...}

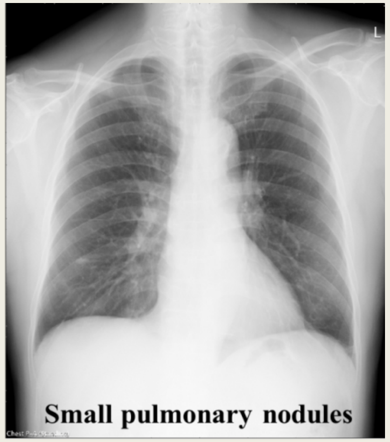
→ cardiac hypertrophy

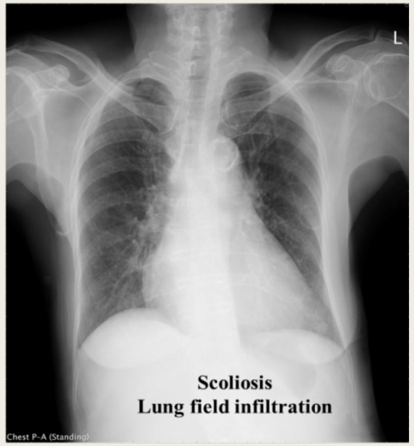
Chest X-rays pathology categories





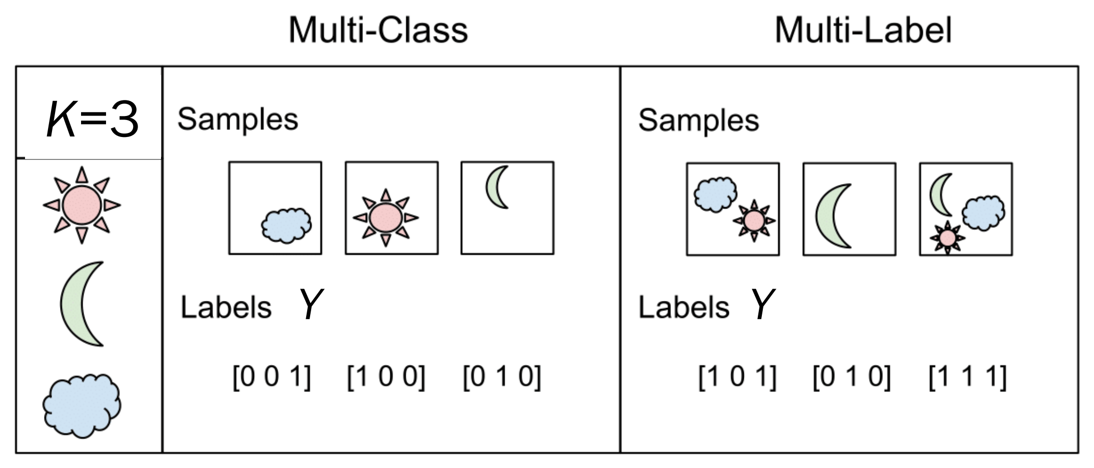






Multiple pathologies per patient

Multi-label classification



- Each image can belong to more than one pathology category (class).
- The outcome vector $\boldsymbol{Y} = (Y_1, \cdots, Y_K)$ will be a one-hot vector (i.e., $Y_k = 1 \text{ or } 0,\ \forall k$) with more than a positive class (i.e., $Y_p = 1$), so it will be a vector of 0's and 1's with K dimensionality.
- This task is treated as K different binary and independent classification problems.

Multi-label classification: example

- A dataset containing chest x-ray images with disease labels
- Each image can belong to more than one of the following 4 abnormalities：心臟肥大、主動脈硬化、肺紋增加、脊椎病變
- Set $\boldsymbol{Y} = (Y_{心臟肥大}, Y_{主動脈硬化}, Y_{肺紋增加}, Y_{脊椎病變})$
- For心臟肥大+主動脈硬化, $\boldsymbol{Y} = (1,1,0,0)$
- For脊椎病變only, $\boldsymbol{Y} = (0,0,0,1)$
- For正常, $\boldsymbol{Y} = (0,0,0,0)$

Multi-label classification

- May have more than one class to be assigned and the label vector may be 0 or 1 in each element.
- Activation function : sigmoid $\sigma(z_k) = \frac{e^{z_k}}{1+e^{z_k}}, k = 1, \cdots, K$
- Loss function : batch weighted binary cross-entropy (bW-BCE)

$$L_{\mathrm{bW-BCE}} = \sum_{j=1}^{J} \left(\sum_{m=1}^{M} \left\{ \beta_{Pj} \sum_{k:\, y_{jmk}=1} \left[-\ln\left(\sigma\left(f_k(\boldsymbol{x}_{jm}) \right) \right) \right] + \right.\right.$$

Transfer learning

When

- Training data extremely limited in some professional fields
- Training data and testing data may follow different distributions

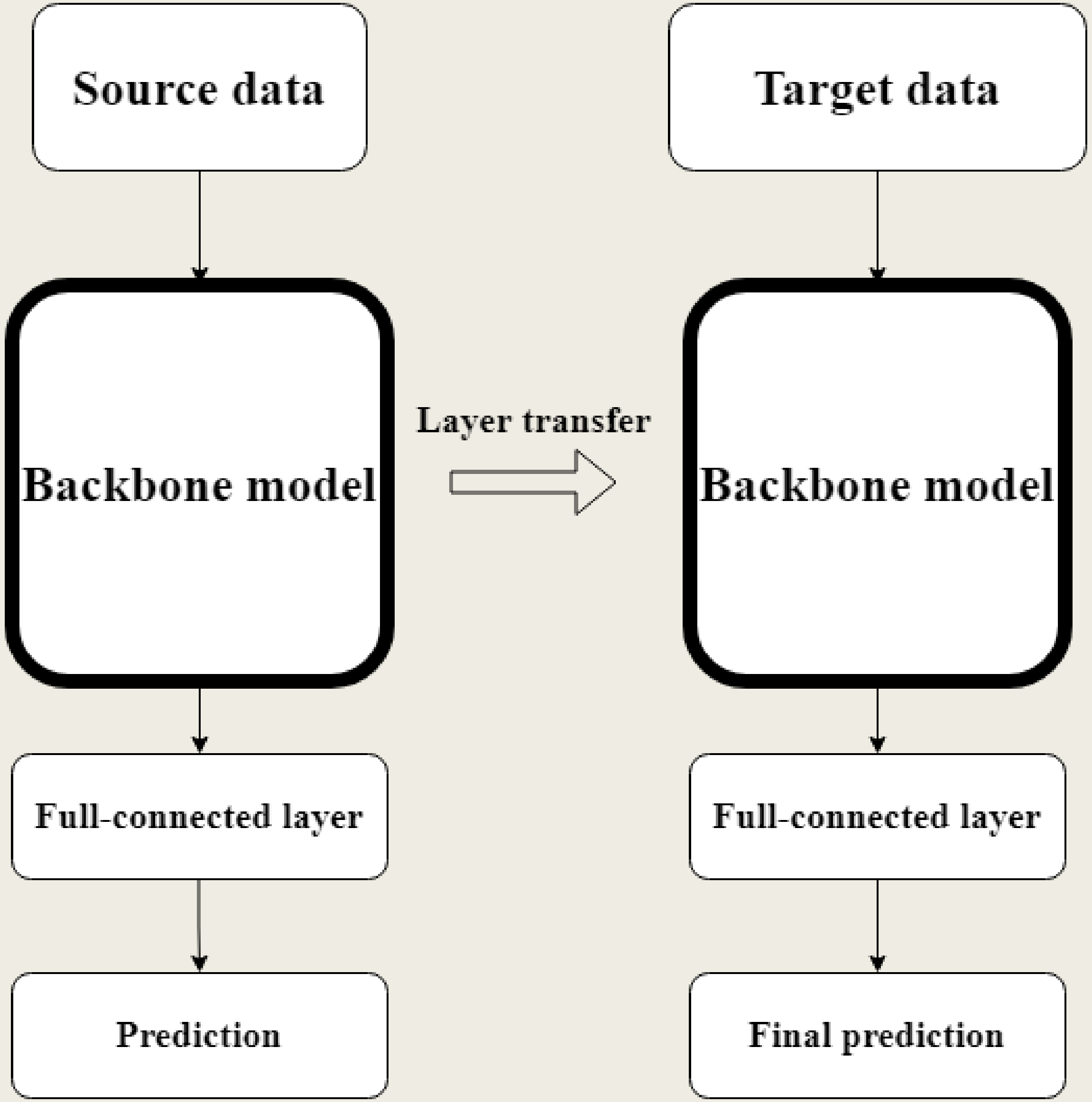
What

- Transfer the trained parameters to a new model in order to accelerate and optimize the process of training
- Inherit the existing neural network and adjust it for new data

Why

- Standing on the shoulders of giants
- Training cost can be very low
- Suitable for learning tasks in small datasets

Transfer learning



Target dataset (E-DA Chest X-ray)

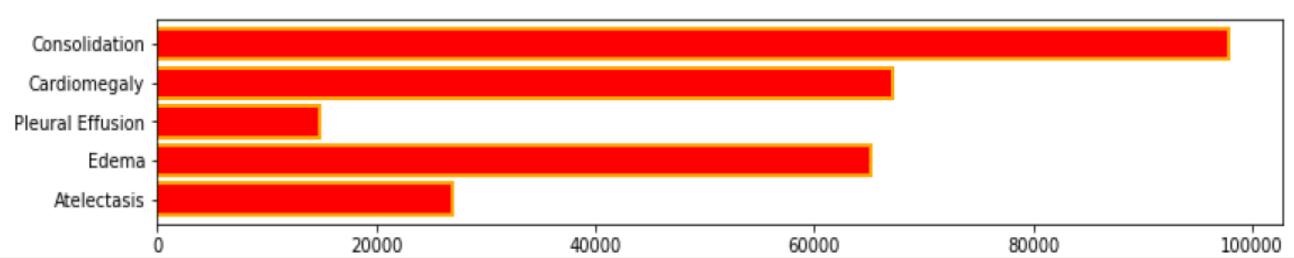
Categories	Sample Size	Subcategory	Sample Size
normal	1314	normal	1314
aortic sclerosis/calcification	91	aortic arch atherosclerotic plaque	28
		aortic arch calcification	16
		aortic atherosclerosis	25
		aortic wall calcification	22
arterial curvature	96	Aortic curvature	67
		Thoracic vertebral artery curvature	29
abnormal lung fields	33	small pulmonary nodules	5
		shadows of pulmonary nodules	8
		tuberculosis	5
		pulmonary fibrosis	15
increased lung patterns	154	increased lung streak	24
		lung field infiltration	85
		obvious hilar	45
spinal lesions	151	degenerative joint disease of the thoracic spine	76
		scoliosis	75
intercostal pleural thickening	36	intercostal pleural thickening	36
cardiac hypertrophy	42	cardiac hypertrophy	42
heart pacemaker placement	7	heart pacemaker placement	7

- Source : E-DA hospital
- Size : 1924
- Category : 19
- Data type : DICOM
- Image size :
  1824~2688 pixels in length
  1536~2680 pixels in width

Source datasets

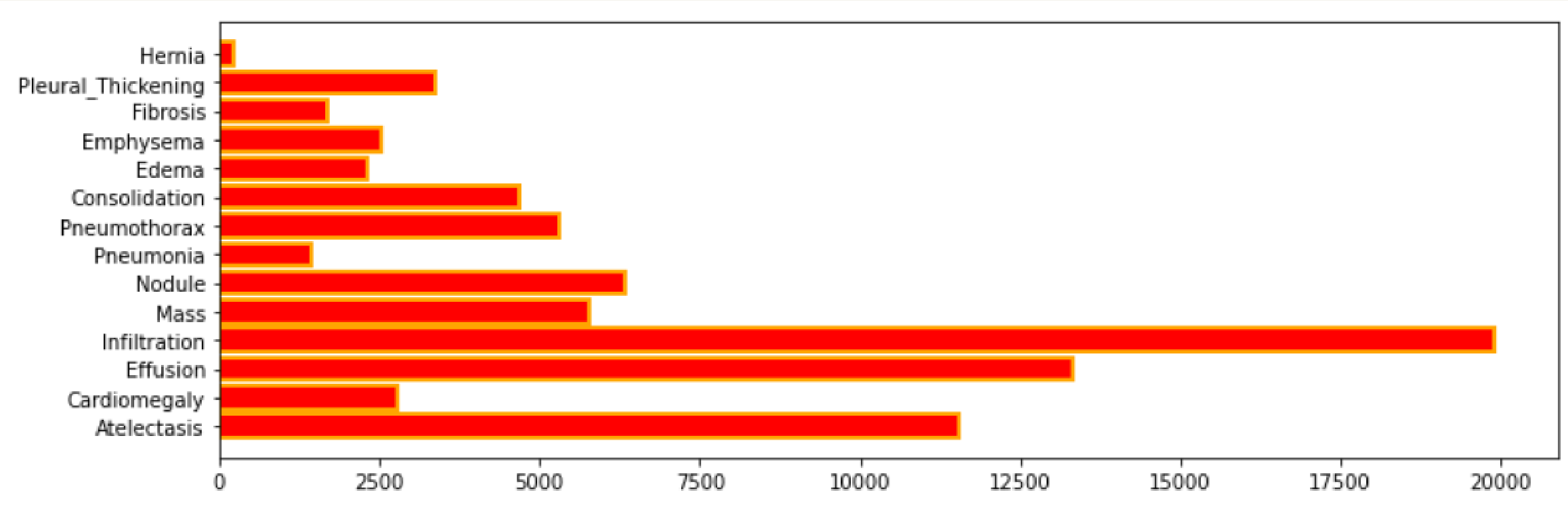
Category	Name	Class	Size	Similarity with the target data
Source data	ImageNet	22,000 +	15 million+	Not closed
Source data	CheXpert	14	224,316	Closed
Source data	NIH Chest X-ray	14	112,120	Closed

CheXpert



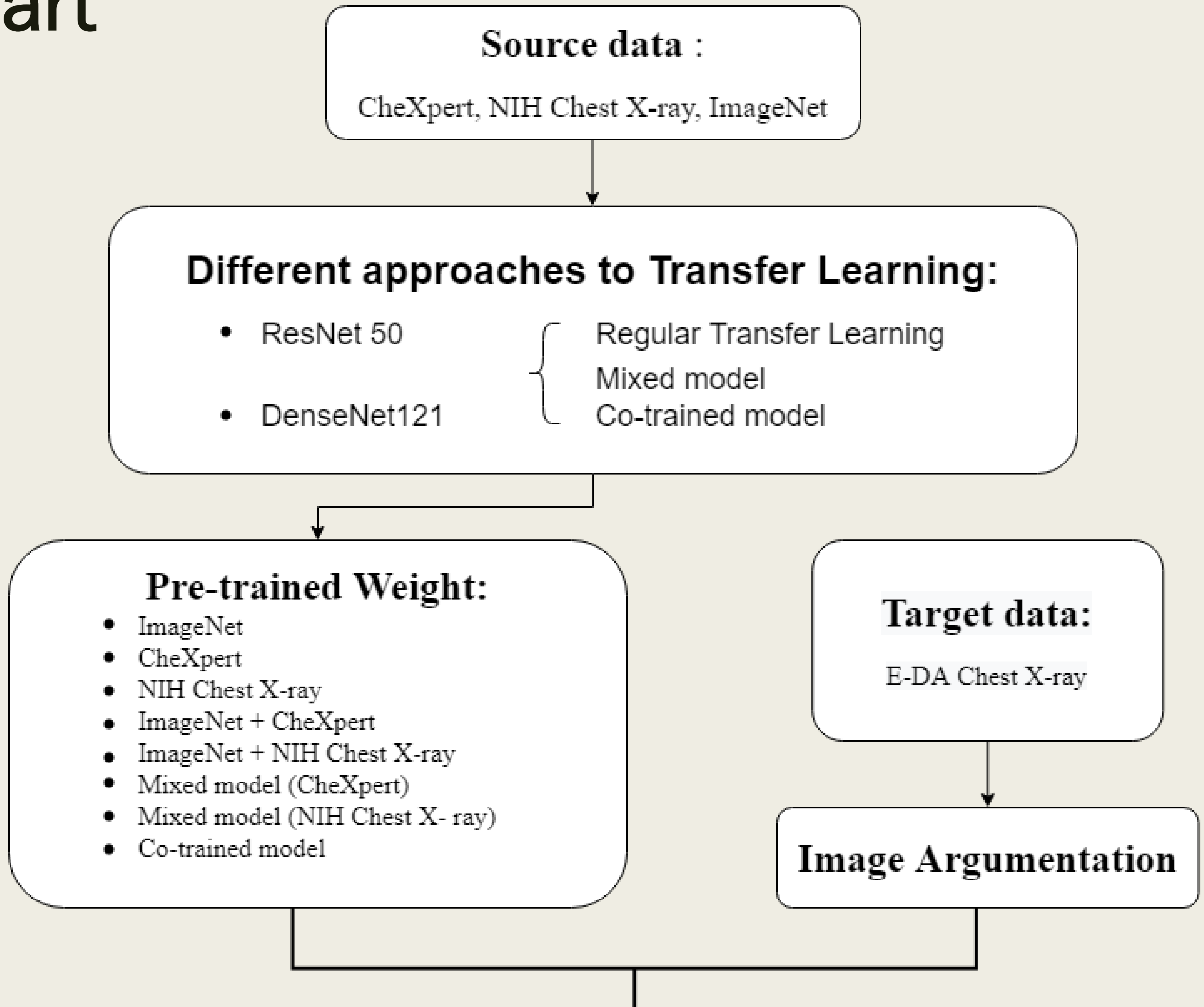
- Size : 224,316
- Category : 14
- Data type : PNG
- Image size : 1024*1024
- Source : https://stanfordmlgroup.github.io/competitions/chexpert/
- Characteristic : uncertain label u

NIH Chest X-ray

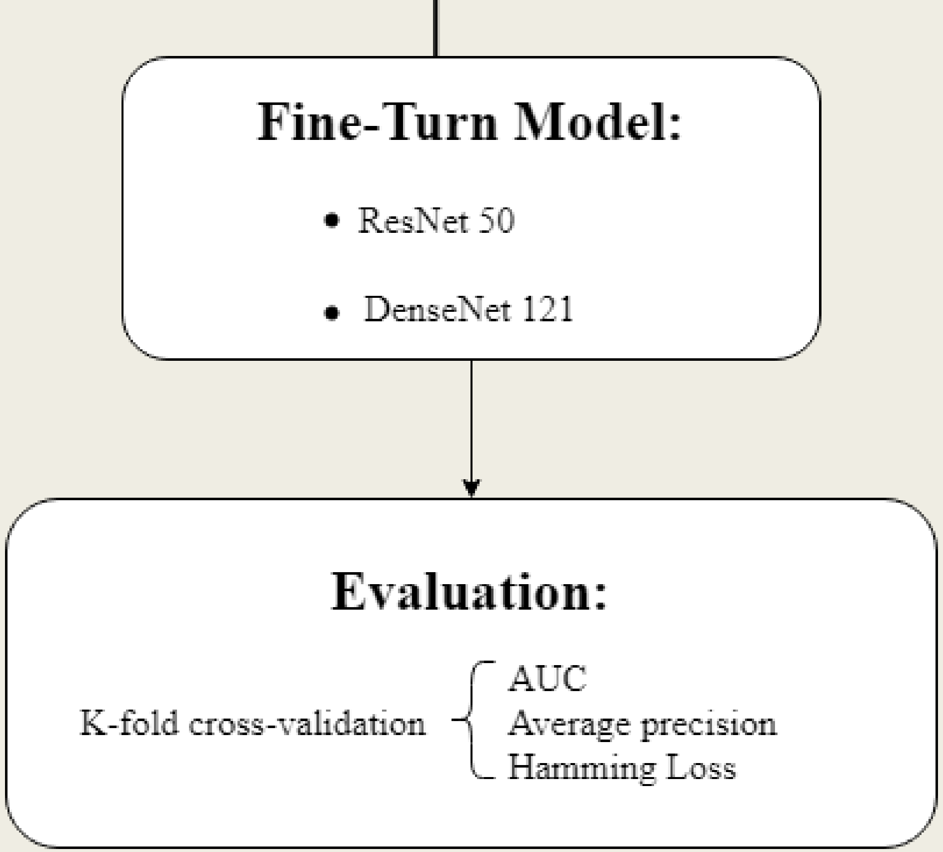


- Size : 112,120
- Category : 14
- Data type : PNG
- Image size : 1024*1024
- Source : https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345
- Characteristic : the labels are expected to be over 90% accurate and suitable for weakly-supervised learning.

Flow chart


Source data :
CheXpert, NIH Chest X-ray, ImageNet
Different approaches to Transfer Learning:
ResNet 50
DenseNet121
Regular Transfer Learning
Mixed model
Co-trained model
Pre-trained Weight:
ImageNet
CheXpert
NIH Chest X-ray
ImageNet + CheXpert
ImageNet + NIH Chest X-ray
Mixed model (CheXpert)
Mixed model (NIH Chest X- ray)
Co-trained model
Target data:
E-DA Chest X-ray
Image Argumentation

Flow chart


Fine-Turn Model:
ResNet 50
DenseNet 121
Evaluation:
K-fold cross-validation
AUC
Average precision
Hamming Loss

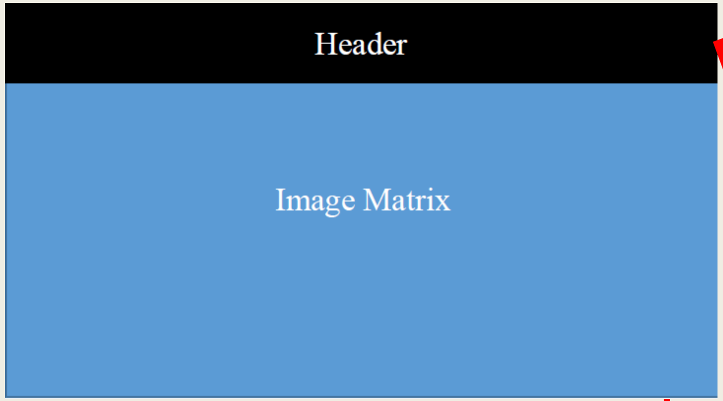
Pre-processing

Target data (**E-DA Chest X-ray**) :

- Removing replicate images .
- Merge original diseases and discard the class “heart pacemaker placement”
- Transform the DICOM format into PNG for saving memory.
- Resize the images into 512*512.
- Remove the fourth channel of these images.
- Use image Argumentation to randomly generate different images.

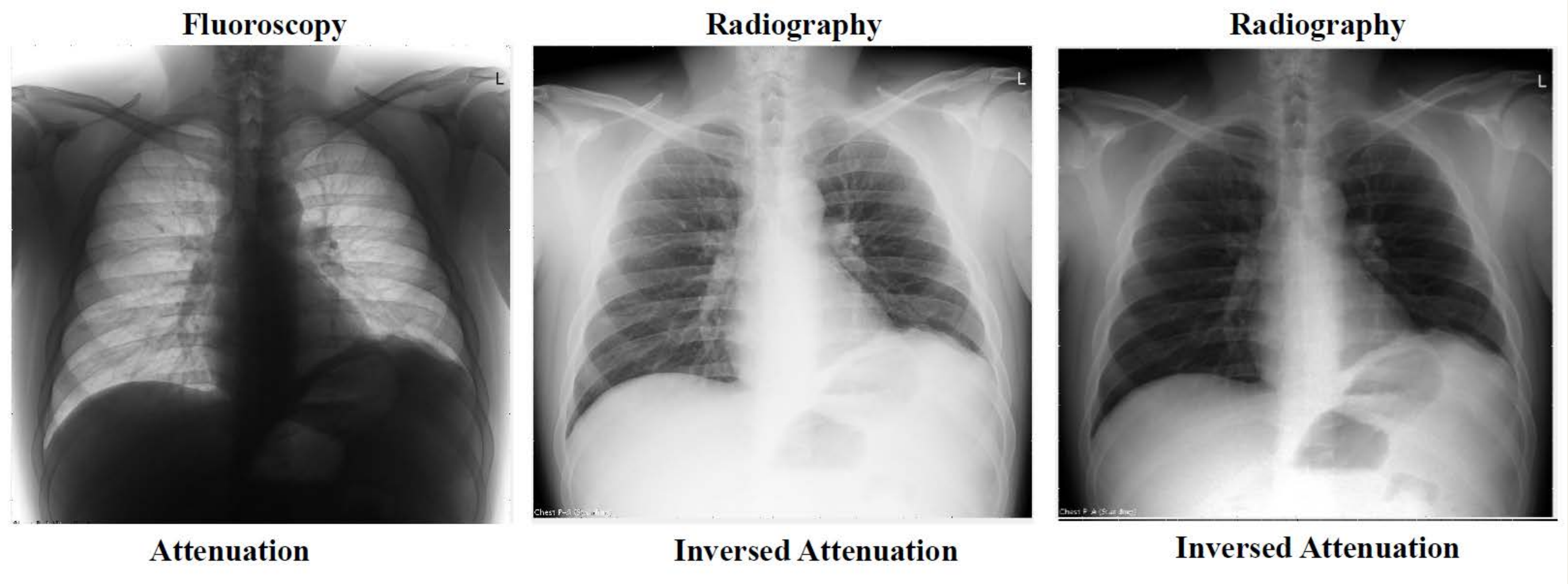
The format of x-ray image

- DICOM: Digital Imaging and Communications in Medicine

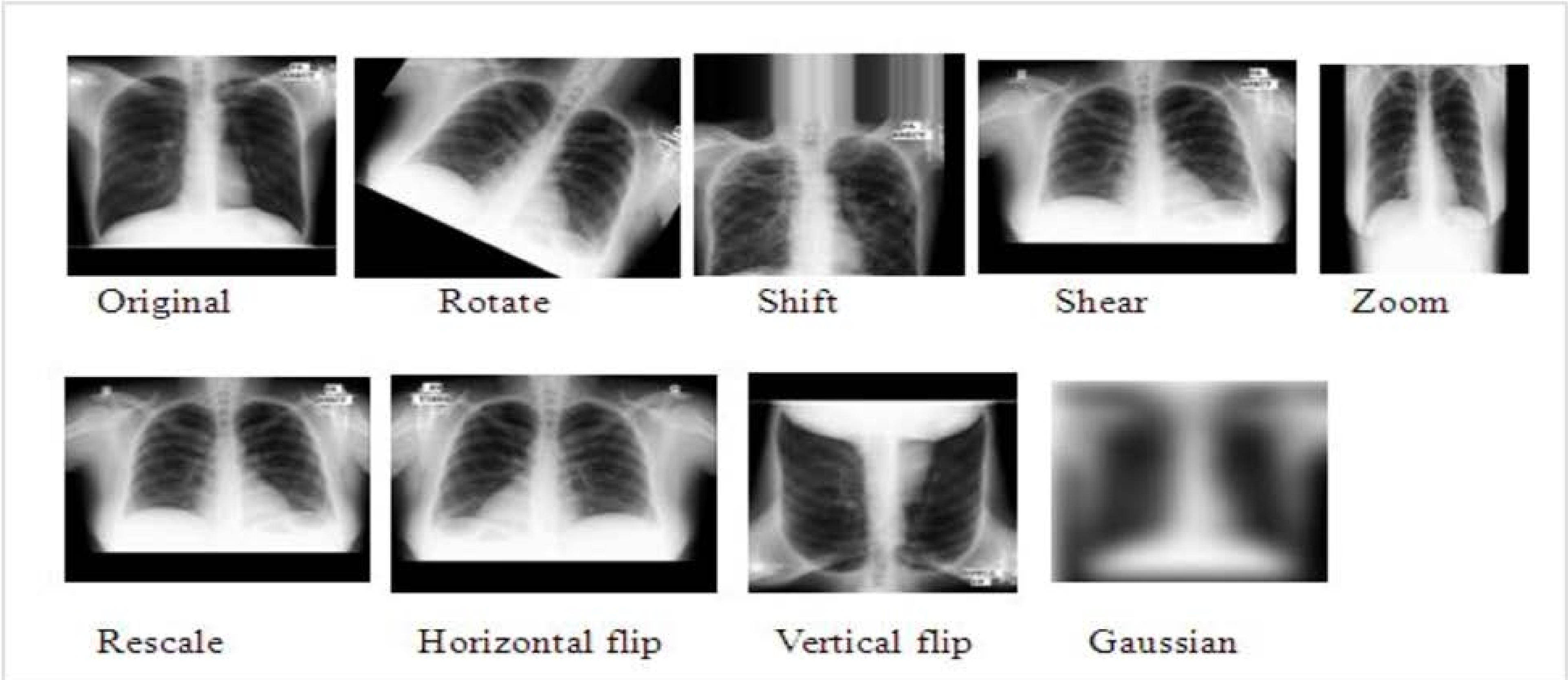


- **Image**
  - Width: 1944
  - Height: 2448
  - BitDepth: 12
  - ColorType: 'grayscale‘
  - ImagerPixelSpacing
- **Patient**
  - PatientName
  - PatientID
  - PatientBirthDate
  - PatientSex
- **Dose**
  - DistanceSourceToDetector: 1800
  - DistanceSourceToPatient: 1770
  - KVP: 100
  - ExposureTime: 11
  - XrayTubeCurrent: 400
  - Exposure: 4
- **Transformation**
  - PixelIntensityRelationshipSign: 1
  - PixelIntensityRelationship: 'LOG'
  - WindowCenter: 2048
  - WindowWidth: 4096
  - PresentationLUTShape: 'INVERSE'

Pre-processing: Inverse attenuation, contrasting



Pre-processing: Image augmentation



Pre-processing

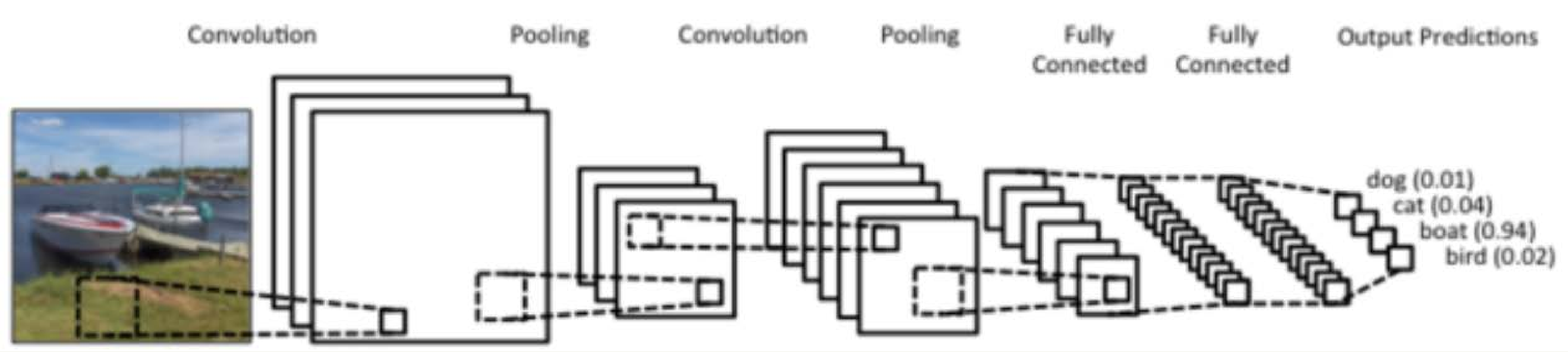
Source data (**CheXpert and NIH Chest X-ray**) :

- Resize the images into 512*512.
- Replicate the one-channel image three times and remove the fourth channel of the four-channel images.

To deal with the uncertainty label of CheXpert, we reconstruct a five-dimension label vector according to the original paper, then take it as a five class multi-label

	Atelectasis	Cardiomegaly	Consolidation	Edema	Pleural Effusion
U-Ignore	0.818 (0.759,0.877)	0.828 (0.769,0.888)	0.938 (0.905,0.970)	0.934 (0.893,0.975)	0.928 (0.894,0.962)
U-Zeros	*0.811 (0.751,0.872)*	0.840 (0.783,0.897)	0.932 (0.898,0.966)	0.929 (0.888,0.970)	0.931 (0.897,0.965)
U-Ones	**0.858 (0.806,0.910)**	0.832 (0.773,0.890)	0.899 (0.854,0.944)	0.941 (0.903,0.980)	0.934 (0.901,0.967)

CNN architectures



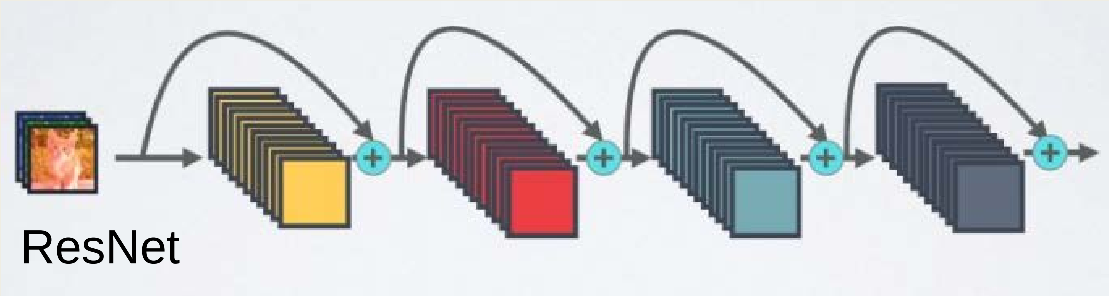
Input

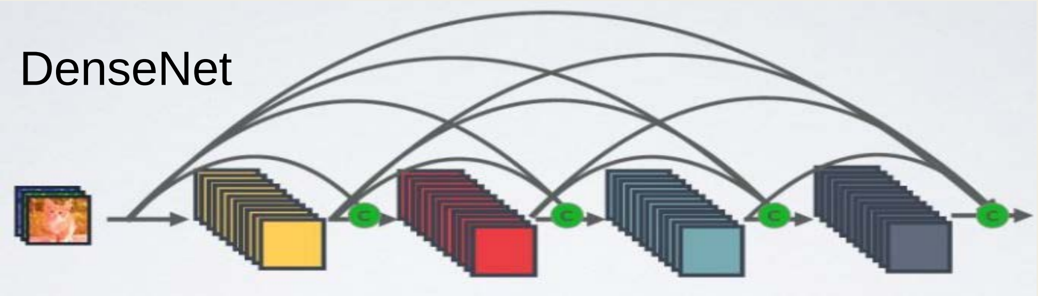
Feature extraction

Classification

Back-bone CNN model

	ResNet	**DenseNet**
Innovation	residual learning	dense shortcuts
	shortcuts connection	feature reuse
	no degradation	transition layer
Output in L layer	$X_L=H_L(x_{L-1})+ x_{L-1}$	$x_{L-1}= H_L([x_0, x_1, \ldots, x_{L-1},])$
Splicing method	element-wise add	concatenate
training speed	fast	slow
Number of parameters	big	small

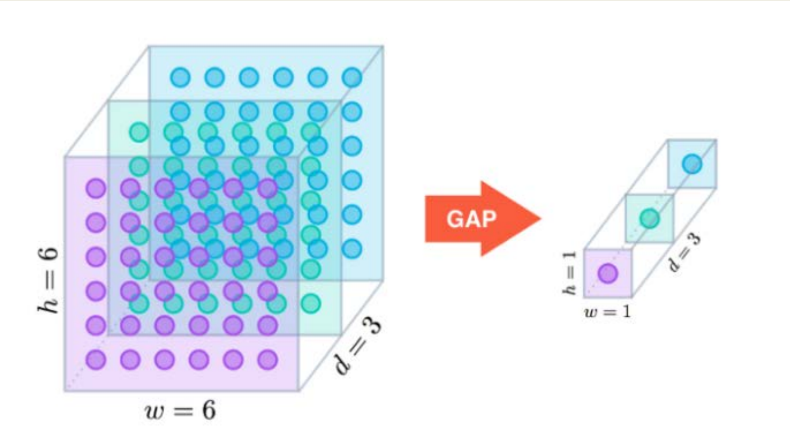




Parameter settings

Parameters	Settings
Input size	512*512*3
Optimizer	Adam
Learning rate	0.0001
Batch size	16
Epochs	30
Loss	Weighted binary cross-entropy
Metrics	Binary accuracy, Mae

- Dropout layer when training pre-trained weights
- GAP (global average pooling layer) to reduce dimension instead of flatten.
- Learning rate scheduler to speed up convergence



Approaches for transfer learning

1. Regular transfer learning
2. Mixed model
3. Co-trained model

Regular transfer learning

ResNet 50
DenseNet 121

+

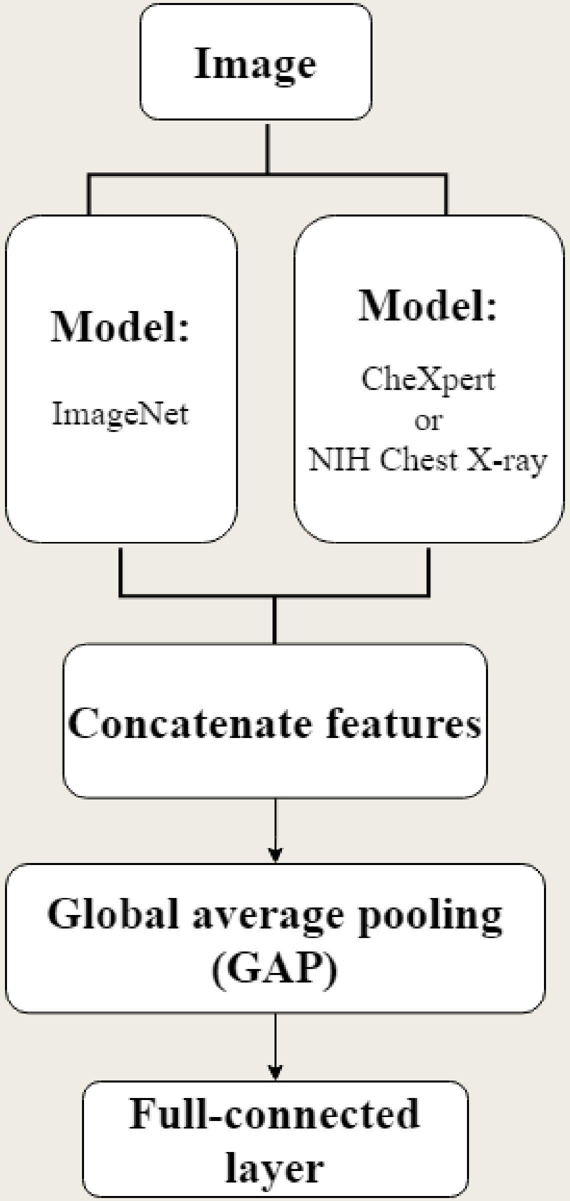
Combined Dataset

ImageNet + CheXpert
ImageNet + NIH Chest-ray

Single Dataset

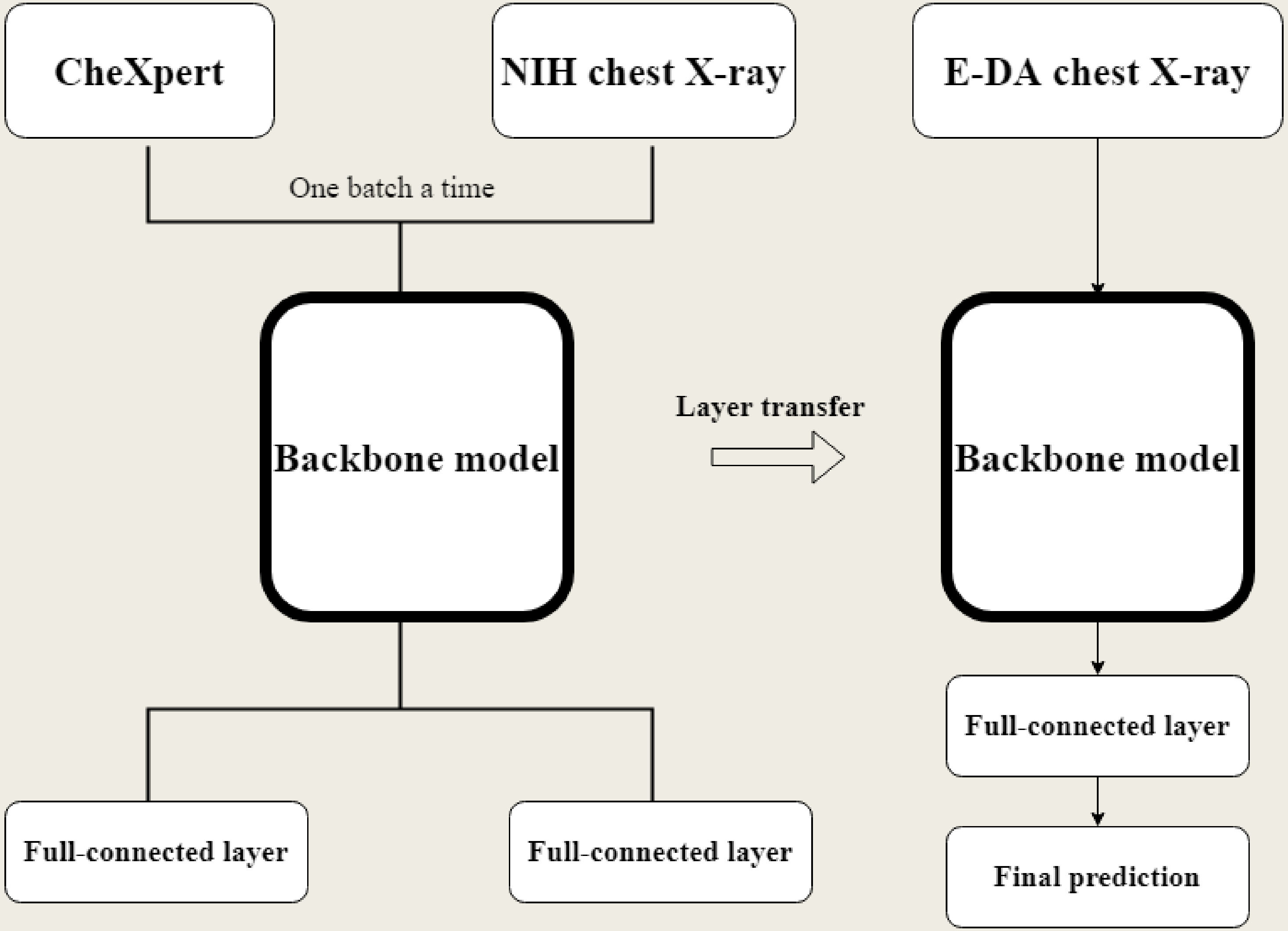
ImageNet
CheXpert
NIH Chest X-ray

Mixed model

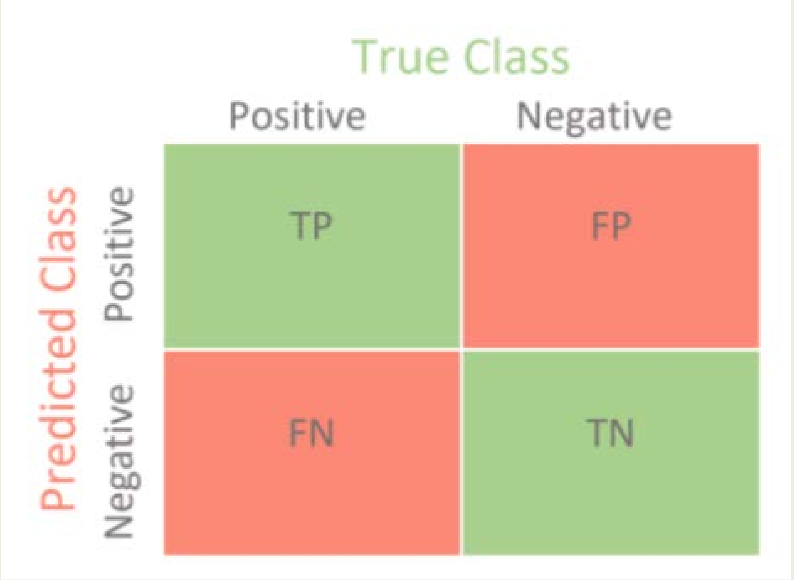


- Pros: Expand features from two different domains
- Cons: Cost twice the memory and time to store and upgrade the parameters

Co-trained model



Evaluation metrics



- **Stratified *K*-fold cross-validation :**

  with package “MultilabelStratifiedKFold”

- **Metrics :**

1. **ROC curve and AUC :** $\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}$, $\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$
2. **Precision curve and Average precision :** Recall = $\frac{\mathrm{TP}}{\mathrm{TP+FN}} = \mathrm{TPR}$,

   Precision = $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$
3. **Hamming Loss :**

$$h_{loss} = \frac{1}{N}\sum_{i=1}^{N} \frac{XOR(Y_{ij}, P_{ij})}{L}, where\ \ XOR(x,y): \begin{cases} 1, & if\ x \neq y \\ 0, & if\ x = y \end{cases}$$

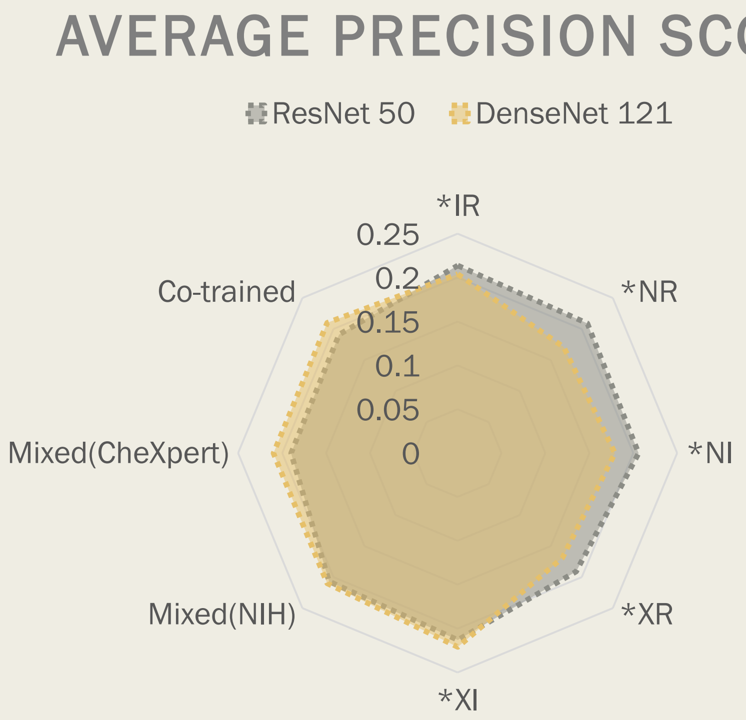
$$threshold = \left(1 - P_{ij}\right) \frac{\sum_{i=1}^{N} Y_{ij}}{\sum_{i=1}^{N}(1 - Y_{ij})}$$

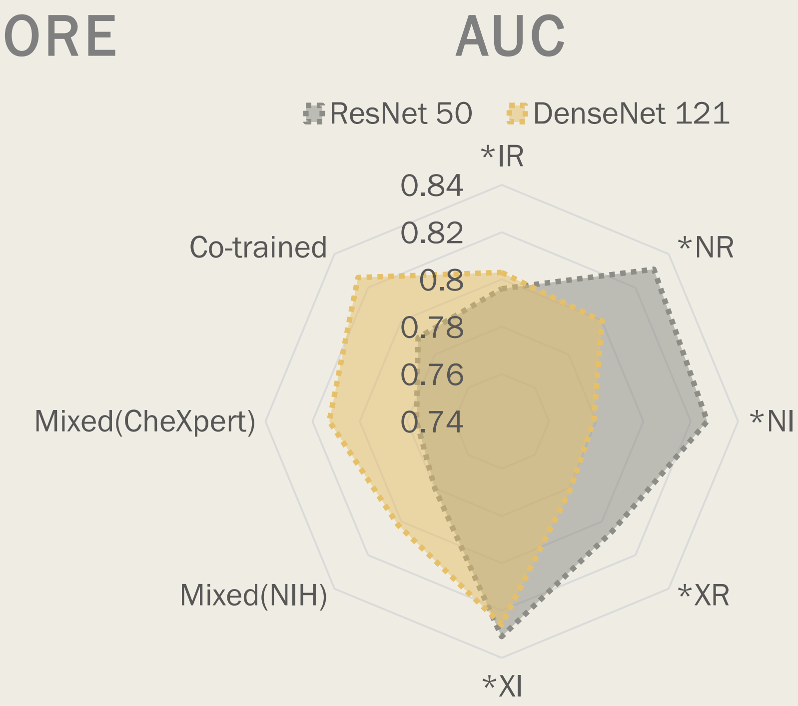
Three different perspectives

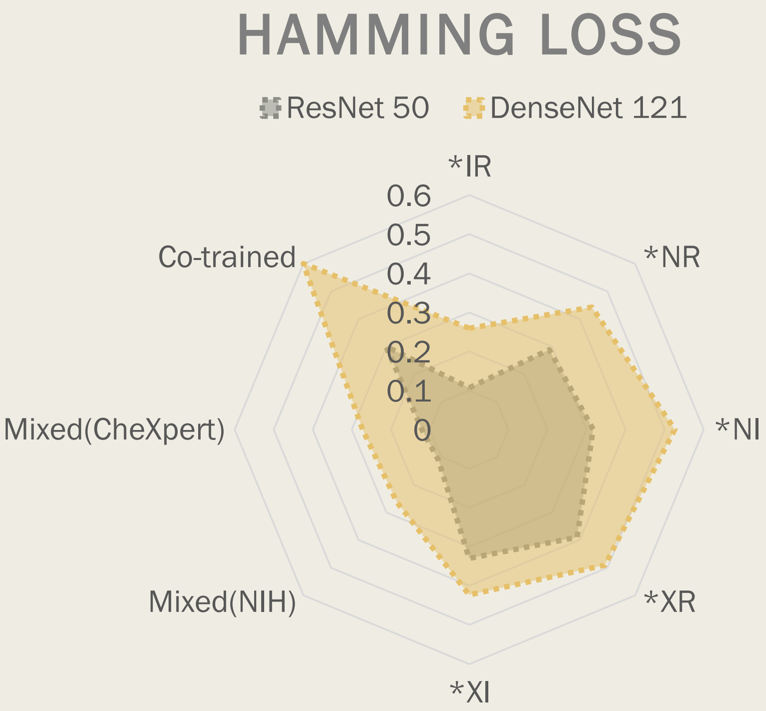
1. Backbone model selection : ResNet 50 vs. DenseNet 121.
2. Source data selection : ImageNet vs. CheXpert vs. NIH chest X-ray
3. Combination method selection :

   Regular transfer learning vs. Mixed model vs. Co-trained model

Backbone model selection







- For regular transfer learning the ResNet 50 performs better. On the contrary, the DenseNet 121 performs better in the mixed model and the co-trained method.
- DenseNet performs almost twice worse than ResNet in the hamming loss, which might be due to the dense-connection in DenseNet.

Source data selection

ResNet 50

	Training		Testing	
Dataset	**Binary accuracy**	**MAE**	**AP**	**AUC**
ImageNet	93.53% (+/-0.45%)	0.07 (+/-0.00)	0.214	0.796
CheXpert	81.85% (+/-7.80%)	0.20 (+/-0.08)	0.191	0.806
NIH	87.85% (+/-1.71%)	0.14 (+/-0.02)	0.209	0.831

DenseNet 121

	Training		Testing	
Dataset	**Binary accuracy**	**MAE**	**AP**	**AUC**
ImageNet	91.56% (+/-1.21%)	0.11(+/-0.01)	0.204	0.803
CheXpert	79.88% (+/-9.99%)	0.23(+/-0.09)	0.169	0.781
NIH	80.70% (+/-12.10%)	0.22(+/-0.11)	0.171	0.8

- The NIH dataset performs better than CheXpert, the reason may due to the uncertain labelling.
- ImageNet performs better than the NIH dataset in training process, but the performance becomes even worse in the testing process.

Combination method selection

ResNet 50

		Training		Testing	
Method	Dataset	Binary accuracy	MAE	AP	AUC
Regular Transfer learning	ImageNet + CheXpert	85.22% (+/-3.59%)	0.16 (+/-0.04)	0.213	0.831
	ImageNet + NIH	86.82% (+/-2.04%)	0.15 (+/-0.02)	0.206	0.827
Mixed model	ImageNet + CheXpert	92.97% (+/-0.68%)	0.08 (+/-0.01)	0.19	0.776
	ImageNet + NIH	93.48% (+/-0.50%)	0.07 (+/-0.01)	0.207	0.78
Co-trained	CheXpert + NIH	85.48% (+/-4.71%)	0.16 (+/-0.05)	0.191	0.79

DenseNet 121

		Training		Testing	
Method	Dataset	Binary accuracy	MAE	AP	AUC
Regular Transfer learning	ImageNet + CheXpert	86.44% (+/-1.47%)	0.17(+/-0.02)	0.221	0.826
	ImageNet + NIH	78.35% (+/-11.89%)	0.25(+/-0.11)	0.179	0.779
Mixed model	ImageNet + CheXpert	91.40% (+/-0.75%)	0.11(+/-0.01)	0.21	0.813
	ImageNet + NIH	93.48% (+/-0.50%)	0.07(+/-0.01)	0.21	0.802
Co-trained	CheXpert + NIH	77.45% (+/-6.91%)	0.26(+/-0.07)	0.21	0.826

Combination method selection

- In training process the method “mixed model” has the best performance, but it doesn’t reflect in the testing process.
- For ResNet 50 the regular transfer learning performs the best.
- For DenseNet 121 the performance of co-trained is tied for first with the regular transfer learning.
- The performance after combination is better, especially for DenseNet 121.

Conclusion

Subject	Contents	Results
Backbone model	ResNet50	Regular transfer learning
	DenseNet121	Mixed model and co-trained model
Source datasets	ImageNet, CheXpert , NIH chest X-ray	NIH chest X-ray
Combination method	Regular transfer-learning, Mixed model, Co-trained model	ResNet50 for Regular transfer-learning DenseNet121 for Co-trained model

- Single dataset is suitable for ResNet 50 and combined dataset is suitable for DenseNet 121.
- A clean label and closed domain to our target data performs better.
- No matter which way we choose to combine datasets, the result is better.

The End !

Thank you !