FPT Education

FPT UNIVERSITY

CAPSTONE PROJECT

MULTI-LABEL LONG-TAILED DISEASE RECOGNITION ON CHEST X-RAY IMAGES

HANOI, AUGUST 2023

OUR TEAM

ASSOC. PROF. PHAN DUY HÙNG
SUPERVISOR

ĐẶNG MINH ANH

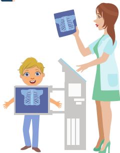
NGUYỄN THẾ TRUNG KIÊN

NGUYỄN MẠNH DŨNG

PRESENTATION OUTLINES

- **Introduction**
- **Data preparation**
- **Methodology**
- **Experiment**
- **Conclusion**

INTRODUCTION



Chest radiography

A common imaging modality used to assess the thorax and the world's most common medical imaging study.

2015

11 radiologists
12 million people
Rwanda

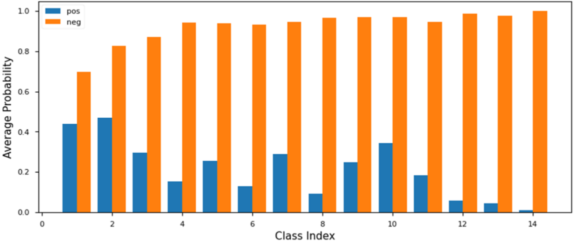
2 practicing radiologists
4 million people
Liberia

1.1 Problem & Motivation

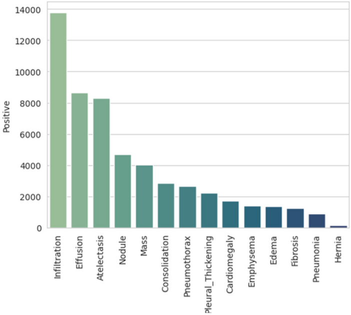
Accurate automated analysis of radiographs has the potential to improve the efficiency of radiologist workflow and extend expertise to under-served regions.

1.1 Problem & Motivation

- **Negative dominant**

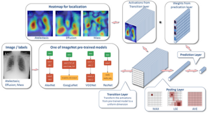


Long tailed distribution of positive samples.

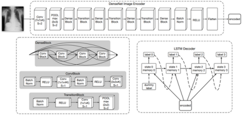


1.2 Related Work

Most of the top solutions for chest Xray diseases recognition are supervised learning.



Wang *et al.* took advantage of a large-scale chest X-ray dataset to formulate the disease diagnosis problem as multi-label classification, using RestNet architecture [1].



Yao et al. [2] employed DenseNet and LSTM to extract features and exploit the statistical label dependencies, respectively, and thus achieved improved diagnosis.

[1] Wang *et al.* Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

[2] Yao *et al.* "Learning to diagnose from scratch by exploiting dependencies among labels"

1.2 Related Work

Another study topic is to reduce the negative impact of imbalanced distribution.

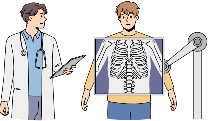
- **Re-sampling:** re-balanced classes by adjusting the number of samples per class. Many sampling strategies have been developed until now including over-sampling [3], under-sampling [4] and mixed sampling [5].
- **Re-weighting:** adjust the training loss values for different classes by multiplying them with different weights. Lin et al. [6] propose an instance-driven loss called Focal loss, which improves cross-entropy loss by down-weight losses assigned to well-classes examples. Asymmetric loss [7] is an improvement of focal loss for multi-label classification.

[3] Park *et al.* "Imbalanced Classification via Feature Dictionary-Based Minority Oversampling"
[4] Lee *et al.* "Framework for the Classification of Imbalanced Structured Data Using Under-Sampling and Convolutional Neural Network"
[5] Ding *et al.* "Ensemble: Towards imbalanced image classification ensembling under-sampling and over-sampling"
[6] Lin *et al.* "Focal loss for dense object detection"
[7] Ridnik *et al.* "Asymmetric loss for multi-label classification".

1.3 Contribution

- Fine-tuning model based on Swin Transformer B and Chest X-ray14 dataset.
- Propose a new loss to adapt with long tail distribution.
- Evaluate performance of model on Chest X-ray14 dataset.

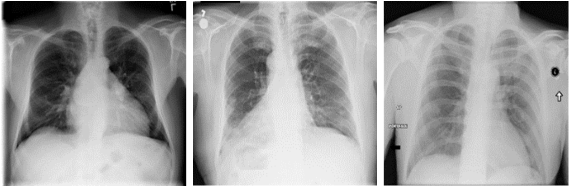
DATA PREPARATION



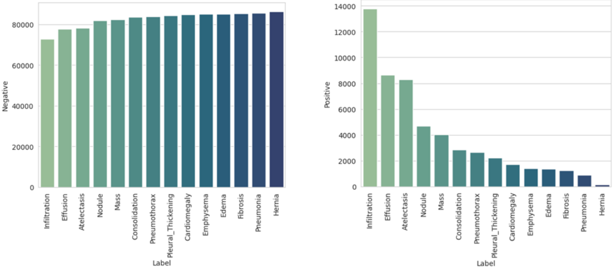
Chest-Xray14 has 112,120 frontal X-ray images with disease labels from 30,805 unique patients.

Image is grayscale and size of 1024x1024.

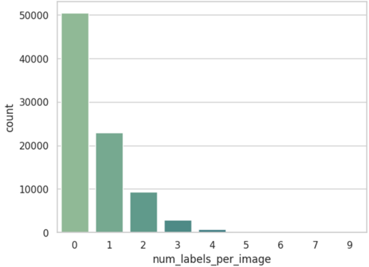
Dataset	# samples
Train	69219
Validation	17305
Test	25596



Sample images of Chest X-ray14 dataset



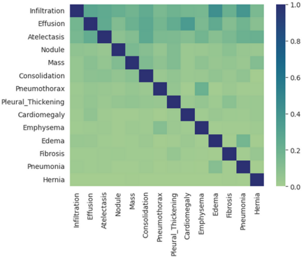
Distribution of negative samples (left) and positive sample (right) of each class.



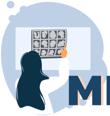
70% of samples have no positive labels

30% have positive labels

In nature, there are correlations between different diseases, that mean one disease can lead to the appearance of other diseases.

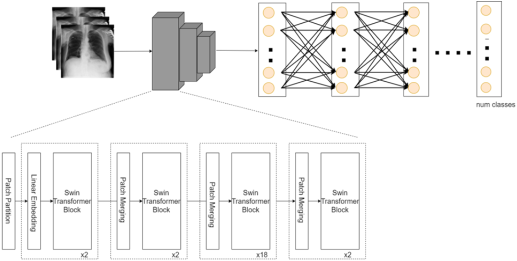


Pathology belonging to head classes usually have high co-occurrence to these in tail classes.



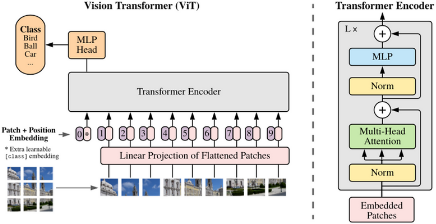
METHODOLOGY

3.1 Swin Transformer



Pipeline of model

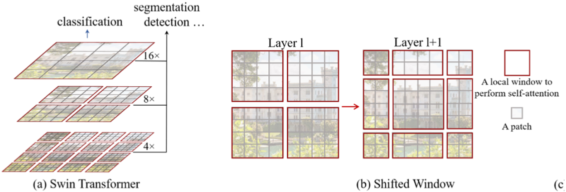
3.1 Swin Transformer



ViT overview [8]

[8] Dosovitskiy *et al.* "An image is worth 16x16 words: Transformers for image recognition at scale"

3.1 Swin Transformer



Patches merging (a) and shifted window (b) in Swin Transformer architecture [9]

[9] Liu *et al.* "Swin transformer: Hierarchical vision transformer using shifted windows."

3.2 Class-Aware loss

3.2.1 Binary Cross Entropy Loss

The most popular way to consider multi-label classification is a series of binary classifications, so binary cross entropy is usually used as loss function.

$$BCE = -yL_{+} - (1-y)L_{-}$$

$$Where \quad \begin{cases} L_{+} = log(p) \\ L_{-} = log(1-p) \end{cases} and \quad L_{+}, L_{-} \quad are \quad respectively \quad positive \quad and \quad negative \quad loss \quad parts.$$

$$L_{total} = \sum_{i=1}^{C} L(p_i, y_i)$$

3.2 Class-Aware loss

3.2.2 Focal Loss

Another way to solve imbalance is focusing on positive labels, which called focal loss [6]. Focal loss weighted sample according to its prediction probability.

$$\begin{cases} L_+ = (1-p)^\gamma log(p) \\ L_- = p^\gamma log(1-p) \end{cases}$$

[6] Lin *et al.* "Focal loss for dense object detection"

3.2 Class-Aware loss

3.2.3 Asymmetric Loss

Asymmetric loss (ASL) [7] is an improvement of FL which introduces two separate gamma coefficients: $\gamma-$, $\gamma+$and negative probability shift.

$$p_m = max(p - m, 0)$$

$$\begin{cases} L_+ = (1-p)^{\gamma+} log(p) \\ L_- = (p_m)^{\gamma-} log(1-p_m) \end{cases}$$

[7] Ridnik *et al.* "Asymmetric loss for multi-label classification".

3.2 Class-Aware loss

3.2.4 Class-Aware Loss

In the long-tail distribution, serious imbalance of positive sample of head classes compared to tail class lead to the domination of positive samples of head classes, thus reduce the performance of model on tail classes.

$$\begin{cases} L_+ = w^+ log(p) \\ L_- = w^- log(1-p) \end{cases}$$

$$w^+ = (1 - \alpha_i^{1-p})$$

$$w^- = 1 - (1 - \alpha_i)^p$$

$$\alpha_i = \frac{P_i}{P_i + N_i}$$

3.2 Class-Aware loss

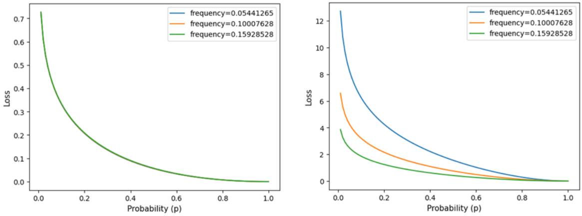
3.2.4 Class-Aware Loss

Total loss is the weighted sum of each class loss, wi is the weighted factor and base on the ratio of positive samples between different classes. Therefore, positive, and negative labels as well as positive labels of different classes are treated differently.

$$L_{total} = \sum_{i=1}^{C} w_i L(p_i, y_i)$$

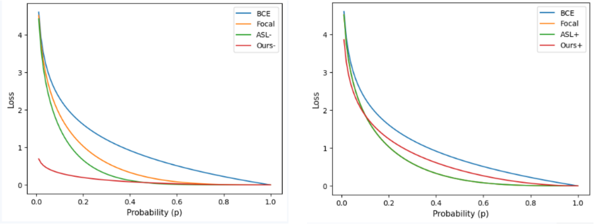
$$\begin{cases} L_+ = w^+ log(p) \\ L_- = w^- log(1 - p_m) \end{cases}$$

3.2 Class-Aware loss



Negative (left) and positive (right) loss of different class.

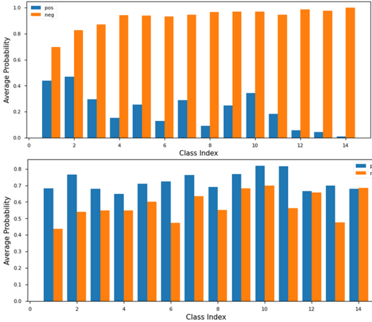
3.2 Class-Aware loss



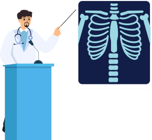
Negative (left) and positive (right) loss of different loss functions.

3.2 Class-Aware loss

Average prediction probability when using BCE loss(top) and our proposed loss(bottom)



EXPERIMENT



4.1 Experiment setting

- Model : Swin-B pretrained weights on ImageNet.
- Dataset: Chest Xray14 dataset.
- Data augmentation: resizing to 256x256, center crop with size of 224, horizontal flip and random rotation.
- Evaluation metrics: AUC, mAP

Hyper parameters	Value
Batch size	32
Num epoch	30
Initial learning rate	1e-4
Optimizer	Adam
Weight decay	0.0005
p_margin	0.2

The parameters of model

4.2 Experiment results

Method	Wang et al.	Yao et al.	DNet	Our
Atelectasis	0.7003	0.733	0.767	0.8395
Cardiomegaly	0.81	0.856	0.883	0.9365
Effusion	0.77	0.806	0.828	0.8721
Infiltration	0.6614	0.673	0.709	0.7642
Mass	0.6933	0.718	0.821	0.9131
Nodule	0.6687	0.777	0.758	0.8669
Pneumonia	0.6580	0.684	0.731	0.7999

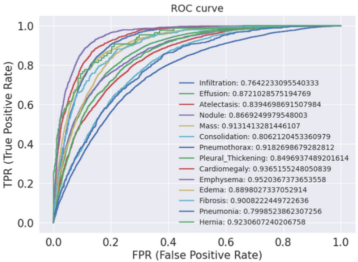
AUC score using our method on the official Chest X-ray14 test set.

4.2 Experiment results

Pneumothorax	0.7993	0.805	0.846	0.9183
Consolidation	0.7032	0.711	0.745	0.8062
Edema	0.8052	0.806	0.835	0.8898
Emphysema	0.8330	0.842	0.895	0.9520
Fibrosis	0.7859	0.743	0.818	0.9008
Pleural Thick	0.6835	0.724	0.761	0.8497
Hernia	0.8717	0.775	0.896	0.9231
Mean	0.7451	0.761	0.807	0.874

AUC score using our method on the official Chest X-ray14 test set.

4.2 Experiment results



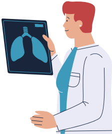
ROC curve of each class

4.2 Experiment results

Loss function	AUC score	mAP
BCE	0.804	0.268
Focal	0.821	0.272
Asymmetric	0.835	0.293
Our w/o decoupling	0.870	0.319
Our w/o margin	0.868	0.312
Our with decoupling and margin	**0.874**	**0.338**

Comparison of different loss function

CONCLUSION



5.1 Conclusion

In this study, we proposed a new loss function to address the problem of class imbalance in image classification. The proposed loss function was applied on the Swin Transformer model with pretrained weights on the ImageNet dataset, using the base version of Swin Transformer, Swin-B. Our experimental results showed that our proposed loss function significantly improve the performance of image classification tasks.

5.2 Limitation and future works

Due to the equipment limitation, we only used the Swin-B version; future research can explore a bigger backbone. Additionally, we only evaluated our proposed loss function on Chest-Xray14 datasets and did not examine its performance on other long-tailed datasets, exploring new loss performance is another possible study.

REFERENCES

[1] Wang *et al.* Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

[2] Yao et al. "Learning to diagnose from scratch by exploiting dependencies among labels"

[3] Park *et al.* "Imbalanced Classification via Feature Dictionary-Based Minority Oversampling"

[4] Lee *et al.* "Framework for the Classification of Imbalanced Structured Data Using Under-Sampling and Convolutional Neural Network"

[5] Ding *et al.* "Ensemble: Towards imbalanced image classification ensembling under-sampling and over-sampling"

[6] Lin *et al.* "Focal loss for dense object detection"

[7] Ridnik *et al.* "Asymmetric loss for multi-label classification".

[8] Dosovitskiy et al. "An image is worth 16x16 words: Transformers for image recognition at scale"

[9] Liu et al. "Swin transformer: Hierarchical vision transformer using shifted windows."

[10] Guendel, S., Grbic, S., Georgescu, B., Liu, S., Maier, A. and Comaniciu, D., 2019. Learning to recognize abnormalities in chest x-rays with location-aware dense networks.

Thanks for your attention