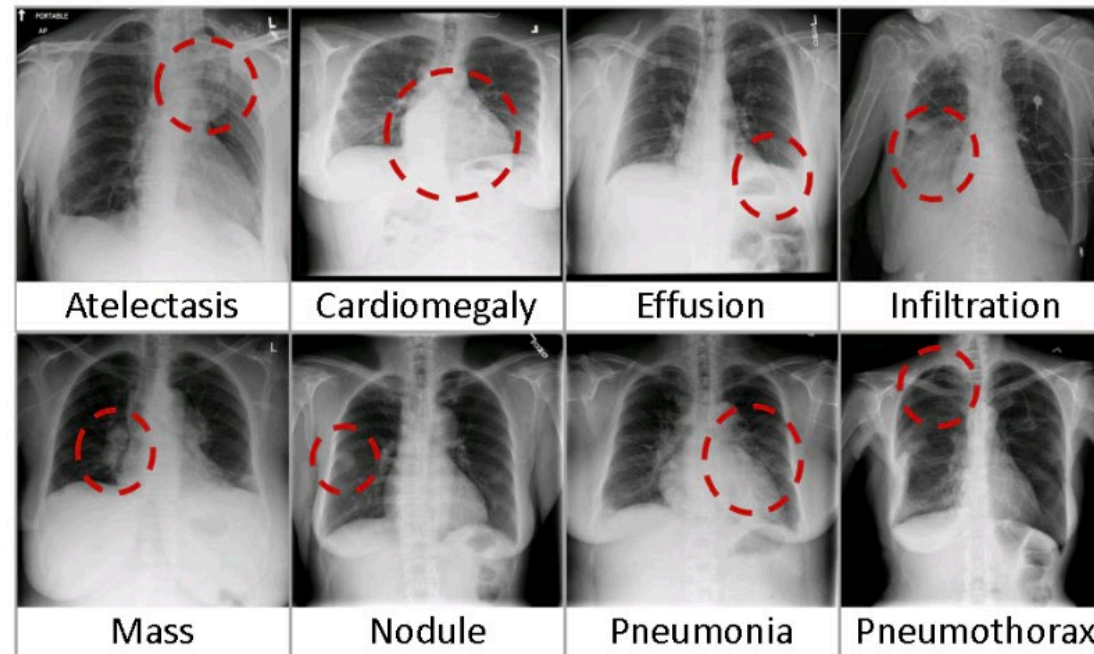


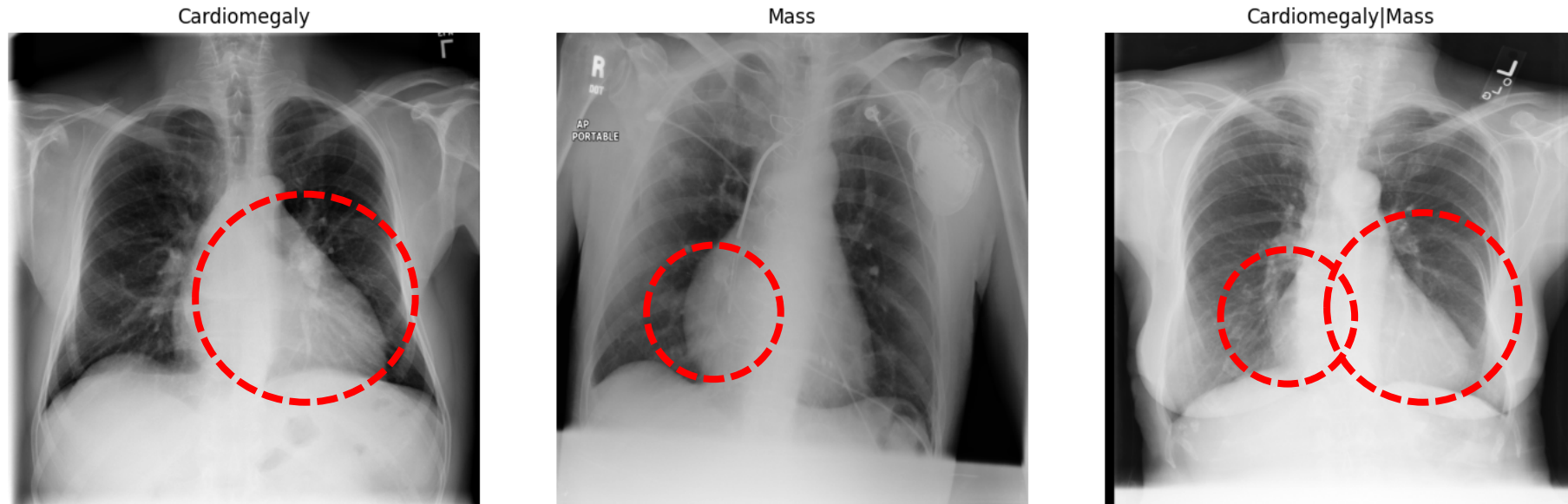
Deep neural networks with latent class modeling for multi-label image classification

Cheng-Chun Wu and Guan-Hua Huang
Institute of Statistics
National Yang Ming Chiao Tung University, Taiwan



Source: adapted from Wang et al., 2017, ChestX-ray14.

1. With the rapid development of deep learning, architectures such as CNNs and Transformers have been widely applied in various clinical tasks, including disease detection and diagnosis.
2. Chest X-rays (CXR) are currently the most commonly used diagnostic tool due to their low cost and high efficiency, especially for initial assessments of lung and heart diseases.
3. This study focuses on NIH ChestX-ray14 to improve multi-label classification performance and interpretability, and the method can also generalize to other tasks.



Mainstream approach: multi-label $\rightarrow$ independent binary tasks.

BCE:

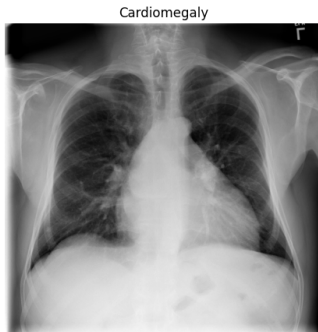
$$\Pr(Y_{i1} = y_1, \dots, Y_{iM} = y_m) \\ = \prod_{m=1}^M \Pr(Y_{iM})$$

RLCA:

$$\Pr(Y_{i1} = y_1, \dots, Y_{iM} = y_m) \\ = \sum_{j=1}^J \{ \Pr(S_i = j) \prod_{m=1}^M \Pr(Y_{iM} | S_i = j) \}$$

Integrate RLCA to model latent label structures, label dependency, and improve interpretability.

Chest X-ray Images



Deep Neural Network
(DenseNet121 & ViT-B/32)
+
RLCA



Multi-label
Classification



1. Treats latent classes as hidden expert groups.
2. Uses MLE-based objective for joint learning.

Backbone Networks

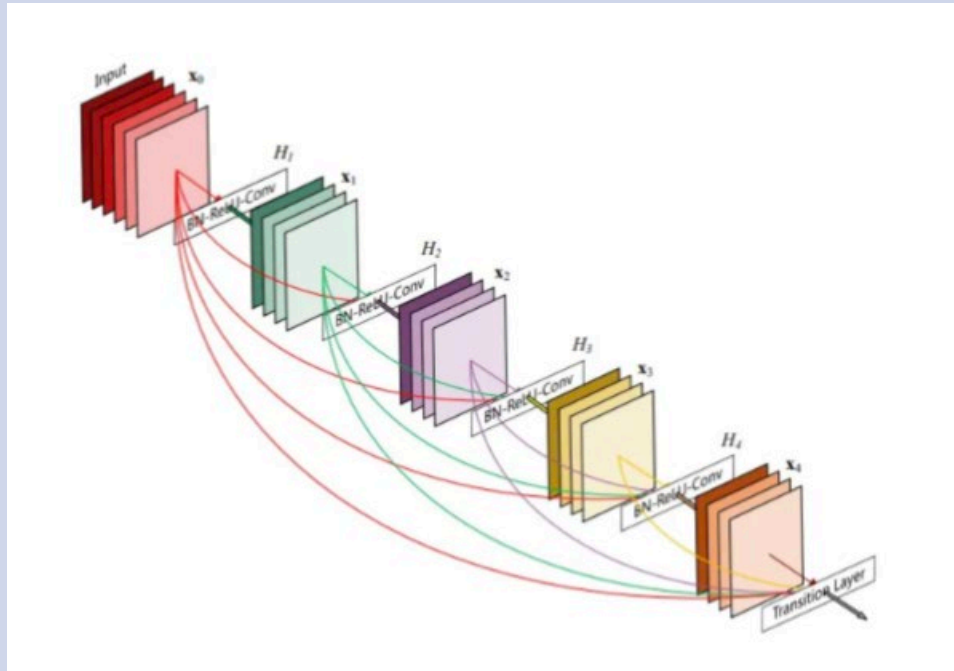
- └ DenseNet121 (CNN-based)
- └ Vision Transformer (Transformer-based)

Loss

- └ WBCE
- └ RLCA

DenseNet121

- Densely connected mechanism.
- Feature reuse & enhanced information flow.
- L layers result in $L(L+1)/2$ connections.
- Growth rate k controls new feature maps.



A dense block with $L = 5$ and $k = 4$. (Image source: Huang et al.)

ViT-B/32

- Image patching & sequence processing.
- Multi-head self-attention mechanism.
- Global context modeling.

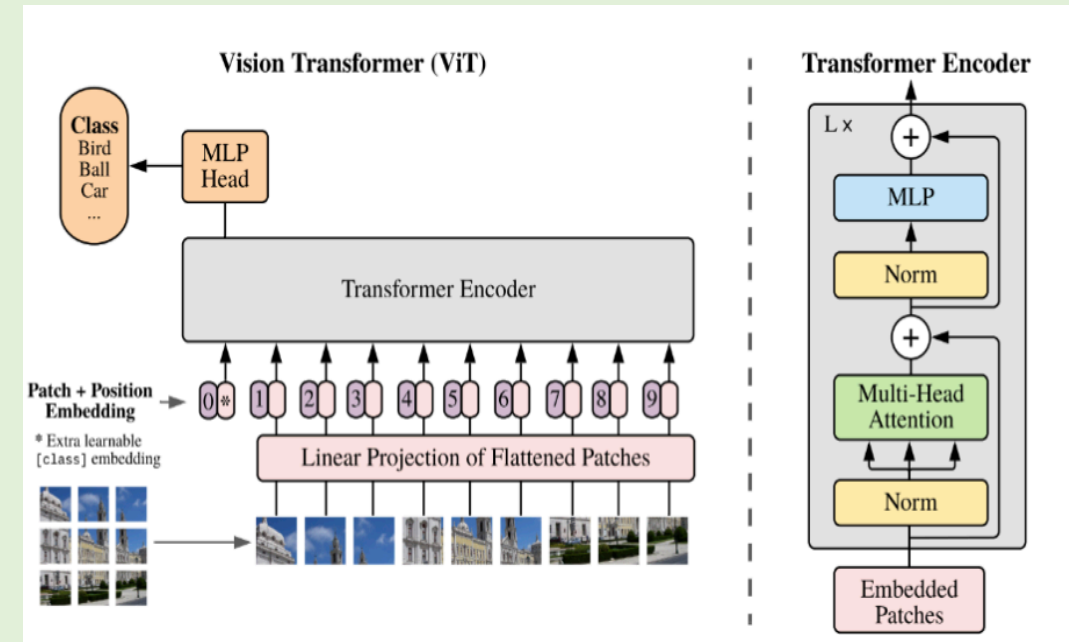


Image source: Dosovitskiy et al.

1. Two-level design with covariates:

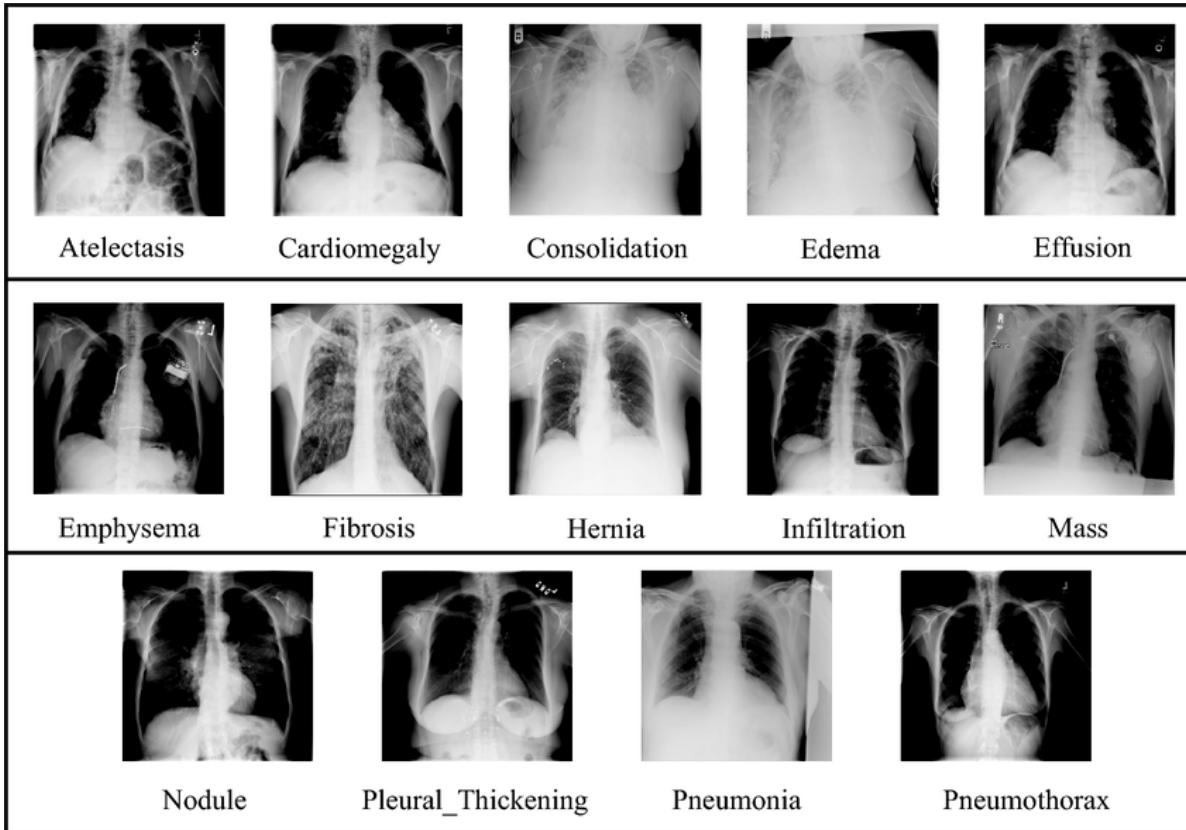
- **Primary covariates $\mathbf{u}$** $\rightarrow$ predict latent class membership probabilities

$$\log \left(\frac{\eta_j(\mathbf{u}_i)}{\eta_J(\mathbf{u}_i)} \right) = \beta_{0j} + \sum_{p=1}^P \beta_{pj} u_{ip}$$

- **Secondary covariates $\mathbf{z}$** $\rightarrow$ adjust the probability of each label given the latent class

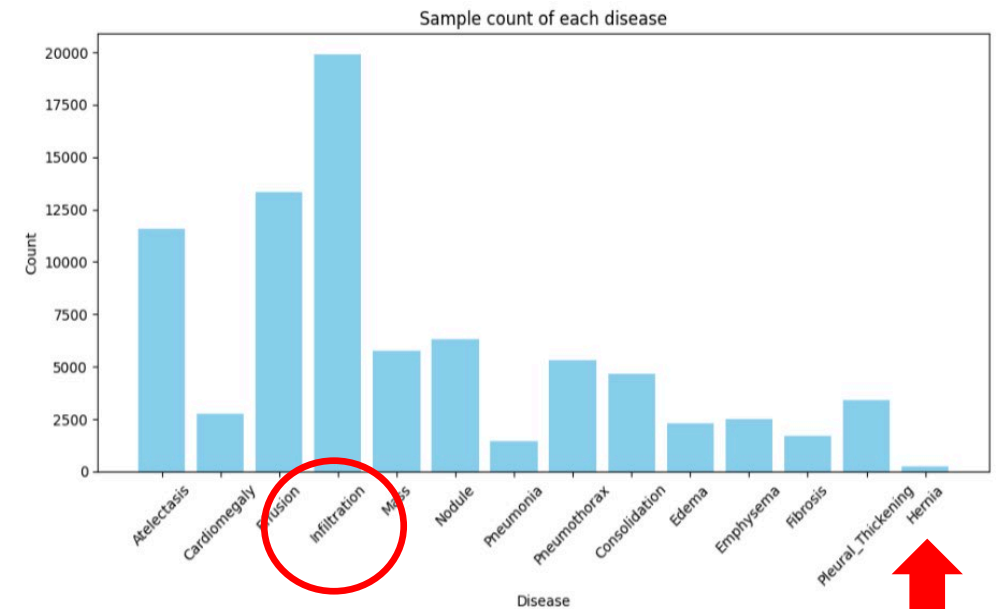
$$\log \left(\frac{p_{mj}(\mathbf{z}_{im})}{1 - p_{mj}(\mathbf{z}_{im})} \right) = \gamma_{mj} + \sum_{l=1}^L \alpha_{lm} z_{iml}$$

2. Helps reduce misclassification caused by other confounding factors or measurement errors, enabling the model to more accurately identify each sample's true latent structure.
3. Uses the EM algorithm for maximum likelihood estimation.



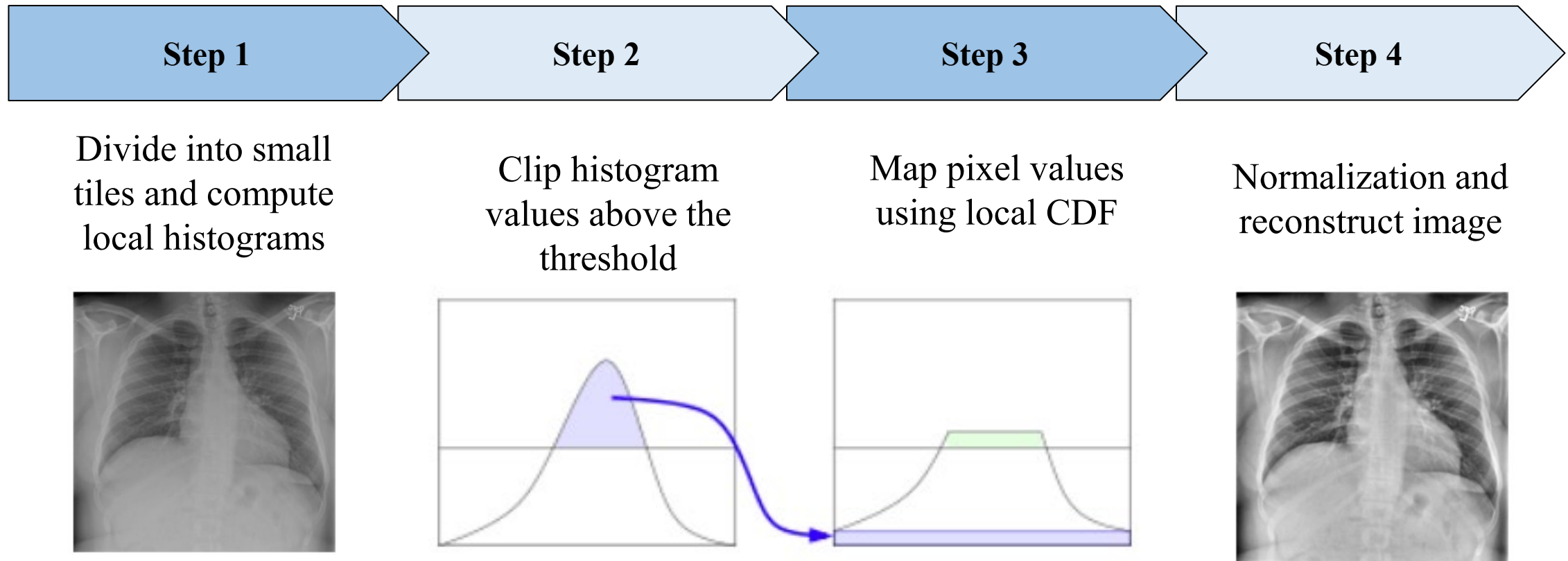
Challenges

- Severe class imbalance.
- Patients may have multiple disease simultaneously.
- "No Finding" treated as a complementary state rather than an independent class.



0.2%₈

The original CXR images often have insufficient contrast; therefore, CLAHE is applied to enhance local contrast and improve structural clarity.



Original



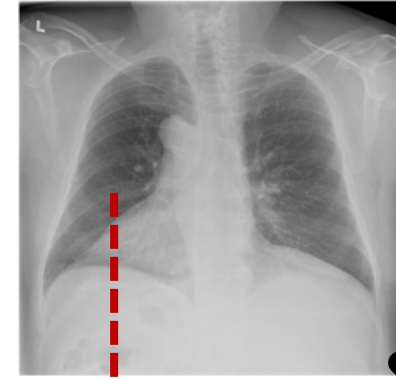
Rotation



Color Jitter



Horizontal flip



Heart shifted right

- Simulate possible natural variations in clinical settings to increase data diversity
- Apply slight random rotations ($\pm 10^\circ$) and brightness/contrast perturbations
- Avoid flipping to maintain anatomical plausibility
- Enhance model generalization and reduce overfitting

Classification

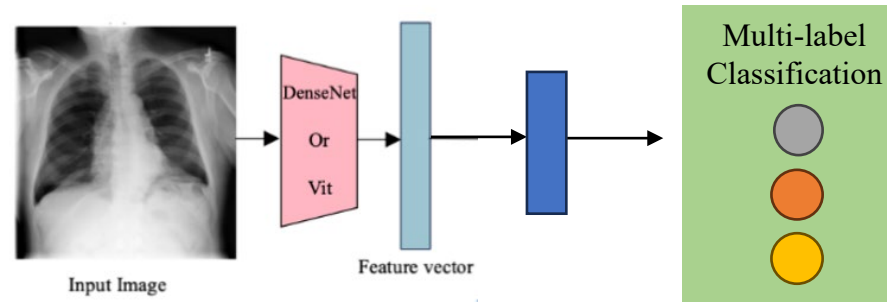
- AUC
- mAP
- Macro F1 score

		Actual	
		1	0
Prediction	1	True Positive (TP)	False Positive (FP)
	0	False Negative (FN)	True Negative (TN)

Metrics	Formula	Definition
Accuracy	$\frac{\sum_i TP_i + TF_i}{N}$	It measures the overall proportion of correct predictions across all classes.
Precision for class i	$\frac{TP_i}{TP_i + FP_i} = \frac{TP_i}{\text{Predicted } i}$	It measures the accuracy of positive predictions for class i .
Recall for class i	$\frac{TP_i}{TP_i + FN_i} = \frac{TP_i}{\text{Actual } i}$	It measures the model's ability to identify all actual instances of class i .
F1 score for class i	$\frac{2 * \text{Precision}_i * \text{Recall}_i}{(\text{Precision}_i + \text{Recall}_i)}$	The harmonic mean, balancing Precision and Recall for class i .
Macro F1 score	$\frac{\sum_i F1 \text{ score}_i}{3}$	This metric provides a single score summarizing the F1 performance across all classes, treating each class equally.
AUC for class i	$\int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR})$	It measures the model's ability to distinguish between positive and negative classes across all thresholds.
AP for class i	$\frac{1}{C} \sum_{i=1}^C AP_i$	It calculates the mean of the Average Precision (AP) over all classes, providing a comprehensive measure of overall performance, especially under class imbalance.

Backbone Loss	CNN-based	Transformer-based
WBCE	DenseNet121 + WBCE	ViT-B/32 + WBCE
RLCA	DenseNet121 + RLCA	ViT-B/32 + RLCA

- CNN uses DenseNet121, Transformer uses ViT-B/32.
- Feature dimension reduced to 512 to alleviate overfitting.



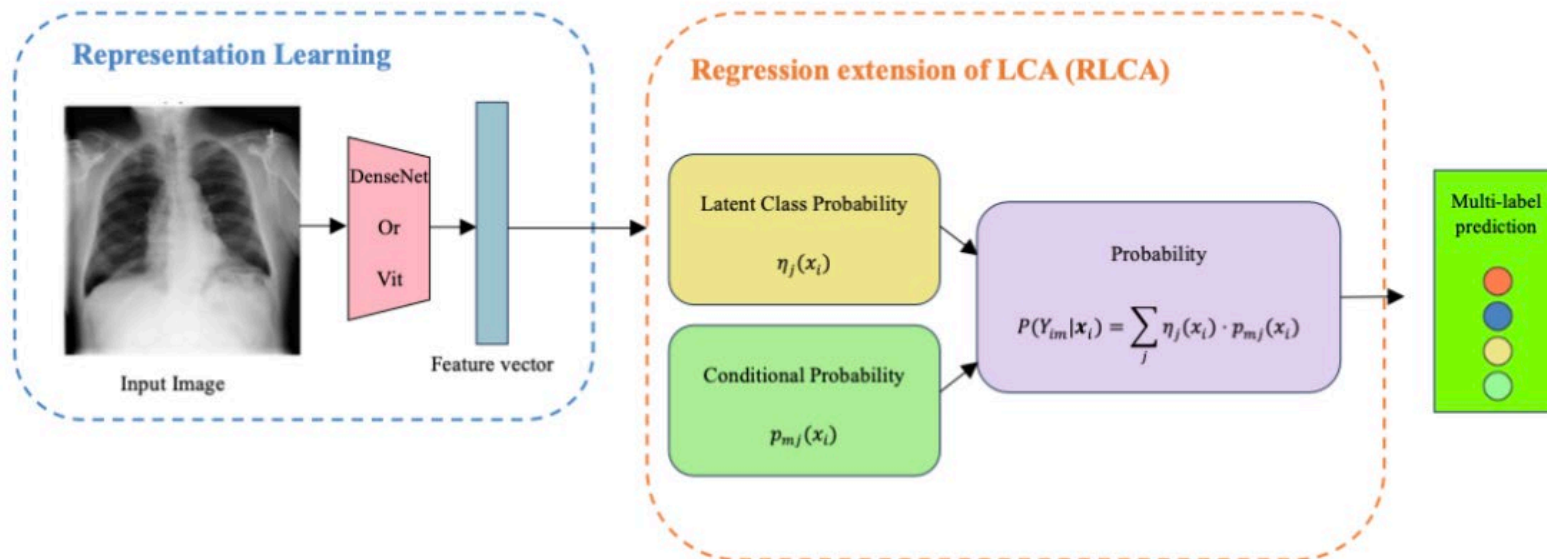
- Employs sigmoid activation and weighted binary cross-entropy (WBCE) loss.
- WBCE designed to handle severe label imbalance by assigning higher weights to rare classes.

$$L_{WBCE} = - \sum_i \sum_m [w_m^{(1)} \cdot y_{im} \log(\hat{y}_{im}) + w_m^{(0)} \cdot (1 - y_{im}) \log(1 - \hat{y}_{im})]$$

$$w_m^{(1)} = \frac{n_{\text{total}}}{n_{\text{positive}} + \varepsilon} \quad \text{and} \quad w_m^{(0)} = \frac{n_{\text{total}}}{n_{\text{negative}} + \varepsilon}$$

- Backbone shared across all comparison schemes to ensure fairness.

- Introduces a latent class structure to replace independent sigmoid outputs.
- Models **latent class probabilities** (η_j) and **conditional label probabilities** (p_{mj}) jointly.
- Captures label co-occurrence patterns and sample-level heterogeneity.
- Provides explicit probabilistic structure, enhancing interpretability.



Step 1

Estimate latent class probability

$$\log\left(\frac{\eta_j(\mathbf{u}_i)}{\eta_J(\mathbf{u}_i)}\right) = \beta_{0j} + \sum_{p=1}^P \beta_{pj} \cdot u_{ip}.$$

Step 2

Compute conditional probability

$$\log\left(\frac{p_{mj}(\mathbf{z}_{im})}{1 - p_{mj}(\mathbf{z}_{im})}\right) = \gamma_{mj} + \sum_{l=1}^L \alpha_{lm} z_{iml}$$

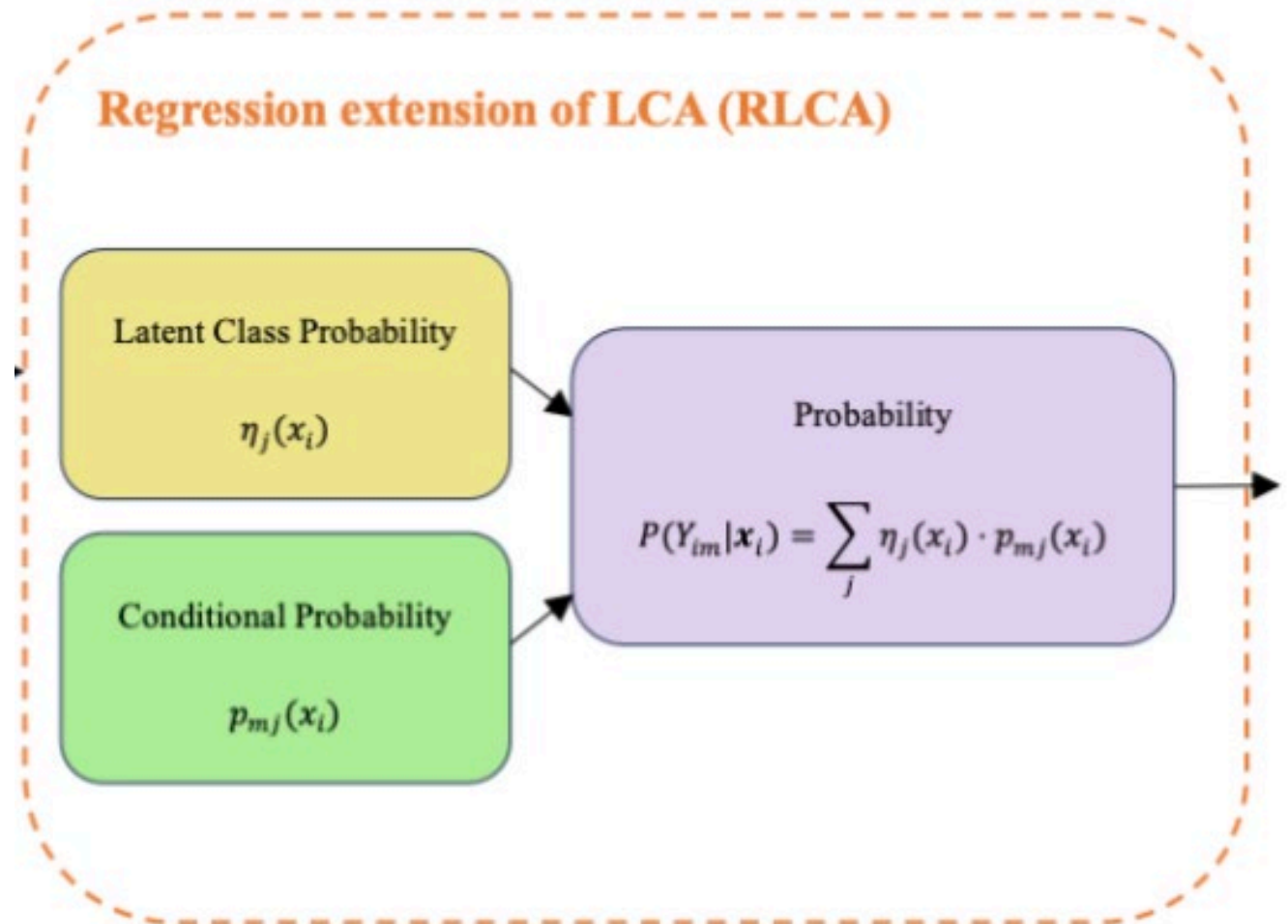
Step 3

Backbone features as covariates

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i \mid u_i, z_{i1}, \dots, z_{iM})$$

Step 4

Weighted negative log-likelihood.



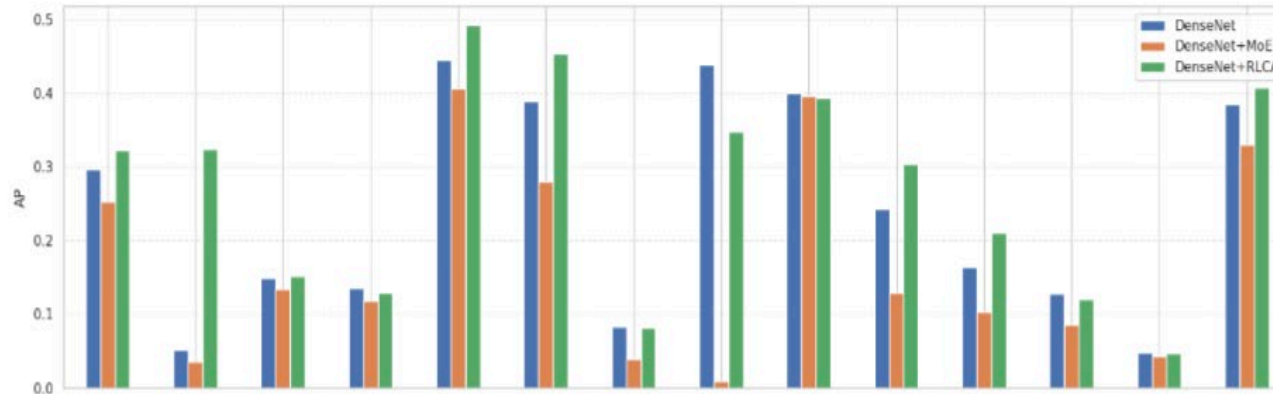
- The RLCA model achieves superior AUC and mAP, especially for rare and difficult cases, demonstrating robust overall performance.

	Model params	mAP	AUC	Recall	F1-score	Precision
WBCE (Scheme 1)	7,486,862	0.2397	0.7669	0.6964	0.5655	0.6103
RLCA (Scheme 2)	7,487,389	0.2705	0.8033	0.7067	0.5490	0.5725

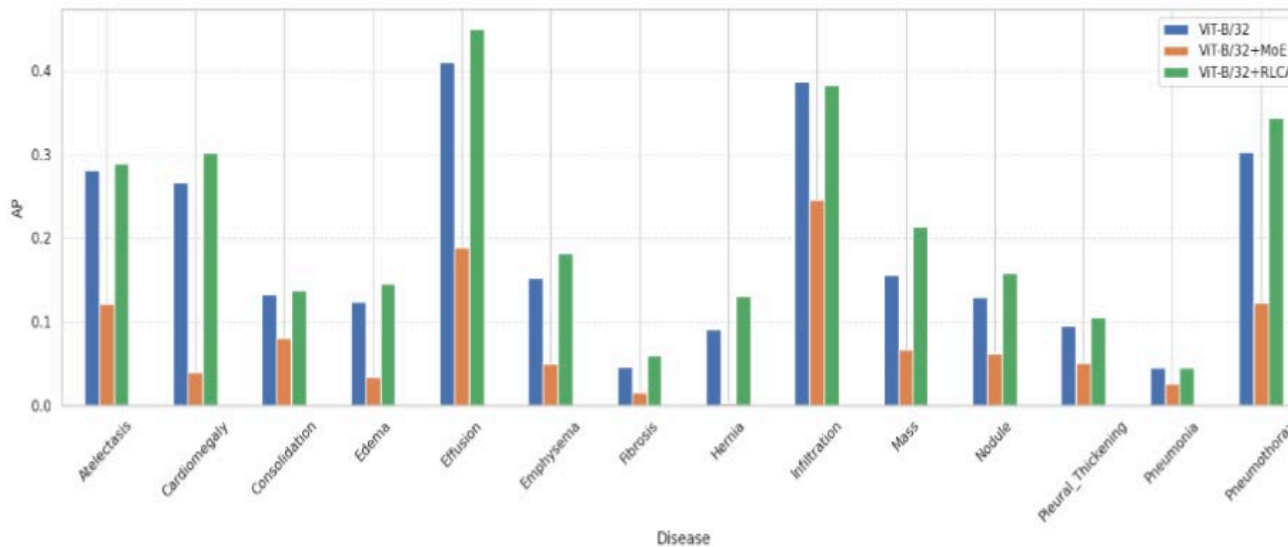
- CNN architecture outperforms Transformer, showing greater stability in imbalanced settings

	Model params	mAP	AUC	Recall	F1-score	Precision
WBCE (Scheme 1)	88,056,590	0.1874	0.7437	0.6793	0.5413	0.5615
RLCA (Scheme 2)	88,057,373	0.2108	0.7652	0.6813	0.4624	0.5504
 CNN-based RLCA (Scheme 2)	7,487,389	0.2705	0.8033	0.7067	0.5490	0.5725

Scheme 1~3 (DenseNet121)



Scheme 1~3 (ViT-B/32)



- RLCA achieves the best performance in most disease categories, effectively addressing class imbalance
- Overall, CNN backbone outperforms ViT, especially for rare disease recognition

Results

- Severe label imbalance severely degrades multi-label performance.
- Need for modeling latent label dependencies without heavy architectures.
- Proposed RLCA addresses both issues effectively.

Findings

- DenseNet121 outperformed ViT, especially under imbalance.
- RLCA improving rare category recognition.

Limitations & Future work

- Difficulties remain for extremely rare classes (e.g., Hernia).
- Possible integration with attention-based label modeling.
- Explore dynamic expert routing and other domains.