Lung Disease Prediction from X-ray Images Using Attention Guided Convolutional Neural Network

Parisa Foroozande Nejad[†], Ramtin Reyhani Kivi[‡]

Abstract—The diagnosis of lung diseases from chest X-ray images is a complex task due to the variability of imaging conditions and the subtlety of overlapping pathologies. To address these challenges and enhance diagnostic accuracy, we propose a multi-label classification framework leveraging a dual-branch architecture plus fusion mechanism with customized Convolutional Neural Networks (CNNs). The model integrates both global and local image information, with a global branch processing the entire image, a local branch focusing on identified lesion regions, and a fusion mechanism combining their outputs for final classification. This approach is applied to 10% of the ChestMNIST dataset, comprising 112,120 frontal X-rays across 14 disease classes, offering a robust benchmark for evaluating predictive performance.

The framework is assessed on three image resolutions (64x64, 128x128, 224x224), demonstrating consistent performance in classification accuracy and generalizability across varying resolutions, without significant improvements at higher resolutions. Comparative analysis highlights improvements in both micro and macro metrics introduced by the fusion strategy in multi-label scenarios. Future work may focus on addressing key challenges in our current approach, such as dataset imbalance and sparsity, improving computational efficiency, and exploring alternative masking strategies.

Index Terms—Chest X-ray, Deep Learning, Attention Mechanism, Convolutional Neural Network, Medical Image Classification

I. INTRODUCTION

The analysis of chest X-ray images has emerged as a critical tool in the detection and diagnosis of lung diseases, which affect millions of individuals globally. The use of advanced computational techniques to interpret these images promises to significantly enhance the accuracy and efficiency of medical diagnoses. This study focuses on the development of a robust algorithm for lung disease prediction using chest X-ray images, leveraging deep learning and convolutional neural networks. The work is built upon the foundational research presented in Wang et al. [1], which introduced the ChestX-ray8 database, a benchmark for weakly-supervised classification and localization of thoracic diseases.

The primary challenge in automated X-ray analysis is the variability of imaging conditions and the overlapping manifestations of multiple diseases, which often result in diagnostic uncertainty. While previous works have proposed CNN-based models, limitations such as inadequate generalizability and

insufficient handling of multi-label scenarios persist. Recent datasets, such as the ChestMNIST dataset, provide an opportunity to address these issues with more diverse and granular imaging data. With over 112,000 frontal X-ray images from 30,805 unique patients and 14 annotated disease classes, ChestMNIST is well-suited for developing and benchmarking sophisticated multi-label classification models.

This paper introduces an approach using CNNs architecture for multi-label classification of lung diseases using the ChestMNIST dataset [2]. The algorithm utilizes a dual-branch design: a local branch for fine-grained feature extraction and a global branch for broader contextual understanding, and a fusion mechanism to combine the outputs and enhance predictive performance. Experiments are conducted across three image resolutions (64x64, 128x128, 224x224) to evaluate the model's robustness and scalability.

The main contributions of this paper are as follows:

- **Comprehensive Dataset Utilization:** Leveraging the ChestMNIST dataset's extensive and diverse imaging data to train and evaluate multi-label classification models.
- **Introduction of AG-CNN Model:** We propose an attention-guided convolutional neural network (AG-CNN) that combines global and local image analysis for more accurate lung disease classification.
- **Robust Performance Analysis:** Conducting experiments across multiple image resolutions (64x64, 128x128, 224x224) to assess the model's generalizability and efficiency.

This paper is structured as follows. Section II reviews the state of the art in chest X-ray analysis and disease prediction. Section III details the model architecture and training strategies. Section IV describes the experimental results and analysis. Section V concludes with a discussion on implications and future directions.

II. RELATED WORK

The use of chest X-ray images for automated lung disease diagnosis has been a significant focus of research in medical imaging and artificial intelligence. Several studies have proposed models to classify thoracic diseases using various datasets and methodologies, highlighting both advancements and challenges in this domain.

The seminal work by Wang et al. [1] introduced the ChestX-ray8 dataset, which served as a benchmark for weakly-supervised classification and localization of common thoracic diseases. This dataset, containing over 100,000 frontal X-ray

[†]Department of Mathematics, University of Padova, email: parisa.foroozandenejad@studenti.unipd.it

[‡]Department of Mathematics, University of Padova, email: ramtin.reyhanikivi@studenti.unipd.it
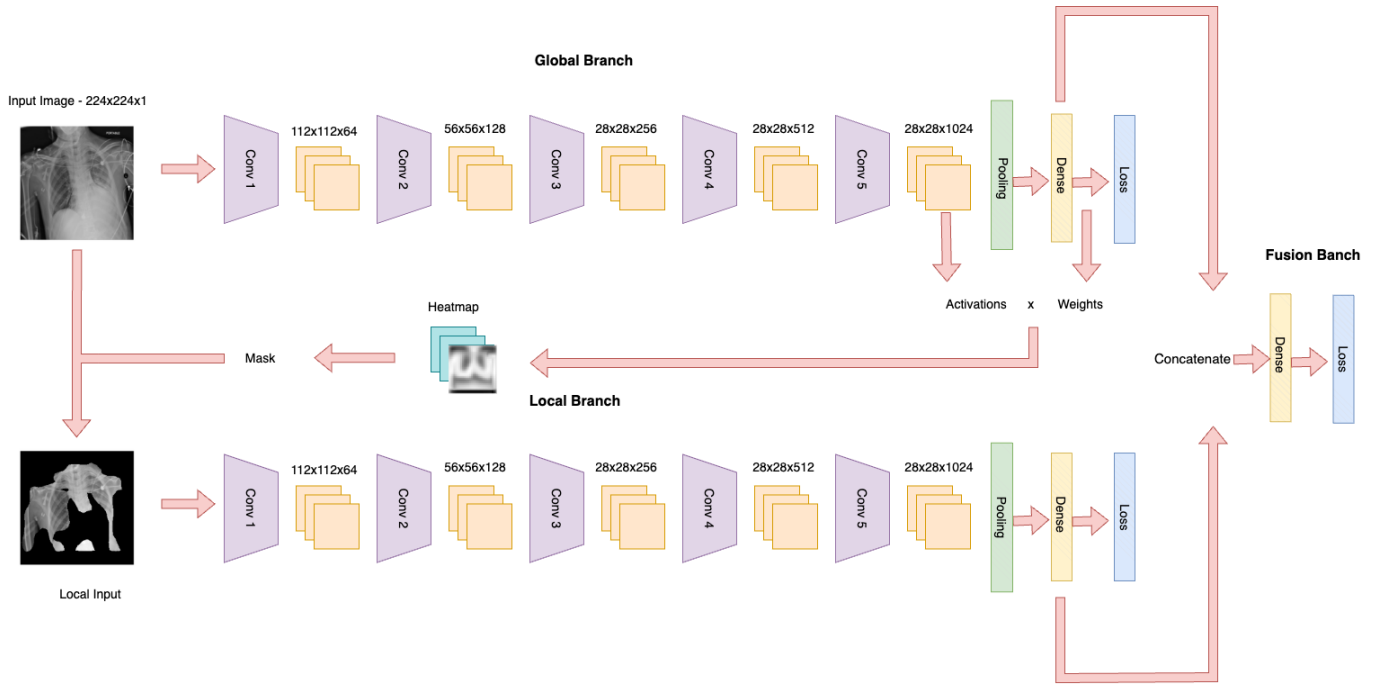
Figure 1: The architecture of the attention-guided convolutional neural network.

images, enabled researchers to develop and test deep learning models. The study demonstrated the potential of convolutional neural networks (CNNs) in handling medical image data but highlighted challenges such as class imbalance and multi-label classification complexities.

Rajpurkar et al. [3] extended this line of work with the development of CheXNet, a 121-layer DenseNet architecture trained to detect pneumonia from chest X-ray images. Their model achieved radiologist-level performance on the pneumonia detection task, setting a high benchmark for accuracy. However, CheXNet primarily focused on binary classification and did not fully address multi-label scenarios involving overlapping disease features.

In a similar vein, Irvin et al. [4] introduced the CheXpert dataset, which provided a more comprehensive labeling approach, including uncertainty labels for ambiguous cases. The study emphasized the need for robust handling of uncertain labels and demonstrated improvements in model performance through curriculum learning. Nevertheless, their work primarily focused on binary or single-label classification tasks.

Recent studies have leveraged the ChestMNIST dataset [5], derived from the NIH ChestX-ray14 dataset, to explore multi-label classification with different image resolutions. This dataset, designed explicitly for machine learning applications, includes 112,120 frontal X-ray images annotated with 14 disease classes. Researchers have utilized architectures like ResNet and attention mechanisms to address challenges in multi-label classification. For instance, Zhang et al. [6] proposed an attention-guided convolutional neural network (AG-CNN) to improve classification accuracy by focusing on disease-relevant regions of the image.

Deep learning has shown success in various computer vision tasks [7]–[9], and recent surveys [10]–[12] have demonstrated its extensive application in medical image analysis [13], [14], especially in chest X-ray image classification [15], [16]. Yao et al. [17] and Kumar et al. [18] addressed multi-label classification of chest X-ray images by exploring potential dependencies among labels. Rajpurkar et al. [16] trained a convolutional neural network for multi-label classification.

Despite these advancements, existing methods often suffer from limited generalizability across diverse imaging conditions and suboptimal performance on high-resolution images. Furthermore, the fusion of global and local feature representations remains an underexplored area in multi-label classification.

In this paper, we extend the existing literature by introducing a dual-branch CNN architecture, specifically designed for multi-label classification using the ChestMNIST dataset. Our approach combines local and global feature extraction, enabling enhanced predictive accuracy and robust performance across three image resolutions (64x64, 128x128, 224x224). Compared to prior work, our model addresses the limitations of single-branch architectures and demonstrates improved scalability and generalizability.

III. Model Architecture

A. Model

The architecture of AG-CNN is presented in Figure 1. It comprises three major branches: the global branch, the local branch, and the fusion branch. Each branch serves a distinct purpose in the classification process.

1) Global Branch: The global branch is designed to analyze the entire chest X-ray image. It consists of downsampling blocks followed by global average pooling and a fully connected layer with sigmoid activation for classification.
2) Local Branch: The local branch focuses on specific lesion areas identified through attention-guided mask inference. It has the same convolutional network structure as the global branch. This branch processes extracted regions of the image to enhance the detection of localized pathologies.
3) Fusion Branch: The fusion branch concatenates the outputs of the dense layers from both the global and local branches. This concatenated layer is then connected to another fully connected layer for final classification, integrating both global and local information to improve diagnostic accuracy.

The simple CNN model used in this study consists of five convolutional blocks, each followed by batch normalization. The first three blocks also contain a max-pooling layer. We did not use the max-pooling for the last two blocks, since it was neccessary for our masking process. The first block starts with a Conv2D layer with 64 filters, followed by 128, 256, 512, and finally 1024 filters in subsequent blocks, before transitioning to a global average pooling layer and a fully connected layer for classification into fourteen categories, resulting in a total of approximately 6.29 million parameters. Both models leverage batch normalization and ReLU activation and Sigmiod activation for output layer because of multi-label classification to enhance training stability and performance.

B. Attention-Guided Mask Inference

To locate discriminative regions in the global image, we construct a binary mask through an attention process. By performing a global pooling after the last convolution layer, the weights learned in the prediction layer can function as the weights of spatial maps from the last convolution layer. Therefore, we can produce weighted spatial activation maps for each disease class (with a size of $S \times S \times C$ by multiplying the activation from the last convolution layer (with a size of $S \times S \times D$) and the weights of prediction layer (with a size of $D \times C$), in which $S \times S$ is the size of the spatial map from the last convolution, C is the number of classes for the classification and D is the dimension of features. Now that we have a heatmap for each class, we apply weights to each class (here we used uniform weights) and combine them to create a single binary mask. The binary mask is then normalized to $[0, 1]$ and is used to generate masked images (Figure 2) by applying a threshold of 0.5.

C. Training Strategy

The AG-CNN model is trained in a three-stage process:

- **Stage I**: Train/fine-tune the global branch using the entire chest X-ray images.
- **Stage II**: Obtain local images through mask inference and use them to train/fine-tune the local branch.

- **Stage III**: Concatenate the global average pooling outputs of the global and local branches, and fine-tune the fusion branch while keeping the previous branches' weights fixed.
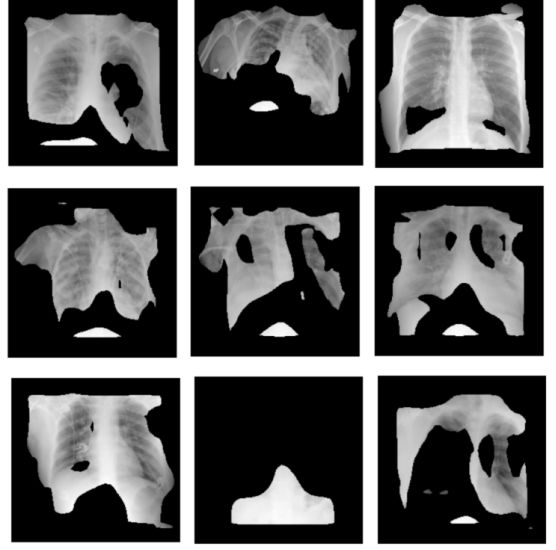

Figure 2: Samples from the inferred masks

D. Multi-label Classification Loss Layer

The Adam optimizer is used for training, with a binary cross entropy loss (BCE) function and a weighted cross entropy (WCE) for the three image resolutions. We also experimented with hinge loss, but since it had worse results, we kept training only with BCE and WCE loss.

The WCE loss is supposed to work better than the BCE, because the model has difficulty learning positive instances and the image labels are rather sparse, meaning there are extensively more 0's than 1's. This is due to our one-hot-like image labeling strategy and the unbalanced numbers of pathology and "Normal" classes. Therefore, as mentioned in Wang et al. [1] we consider a positive/negative balancing factor β_P and β_N to enforce the learning of positive examples. the weighted cross entropy (WCE) is defined as follows,

$$L_{W-CEL}(f(\vec{x}), \vec{y}) =$$
$$\beta_P \sum_{y_c=1} -\ln\left(f\left(x_c\right)\right) + \beta_N \sum_{y_c=0} -\ln\left(1 - f\left(x_c\right)\right), \quad (1)$$

where β_P is set to $\frac{|P|+|N|}{|P|}$ while β_N is set to $\frac{|P|+|N|}{|N|}$. $|P|$ and $|N|$ are the total number of 1's and 0's in a batch of image labels.

E. Algorithm

The overall procedure for our CNN is summarized in Algorithm 1.

Algorithm 1 Attention Guided CNN

Require: Input image I; Label vector L for 14 classes
Ensure: Probability score $p_f(c|[I, I_c])$ for each class c

1: Initialize global branch weights W_g and local branch weights W_l.
2: **Stage I: Global Branch Training**
3:     Train global branch on I to learn W_g.
4:     Compute global feature map F_g and global probability $\tilde{p}_g(c|I)$.
5:     Optimize global loss function $\mathcal{L}_g$ using weighted cross-entropy and binary cross-entropy.
6: **Stage II: Local Branch Training**
7:     Generate attention mask M using dense layer weights and activations from the last convolutional layer, then extract region I_c from I using M.
8:     Train local branch on I_c to learn W_l.
9:     Compute local feature map F_l and local probability $\tilde{p}_l(c|I_c)$.
10:     Optimize local loss function $\mathcal{L}_l$ using weighted cross-entropy and binary cross-entropy.
11: **Stage III: Fusion Branch Training**
12:     Concatenate global and local features and apply a Dense layer: $F_f = \text{Dense}([\text{Dense}(F_g), \text{Dense}(F_l)])$.
13:     Train fusion branch to learn W_f.
14:     Compute fused probability $\hat{p}_f(c|[I, I_c])$.
15:     Optimize fusion loss function $\mathcal{L}_f$ using weighted cross-entropy and binary cross-entropy.
16: **Output:** Final probability scores $\hat{p}_f(c|[I, I_c])$ for all classes c.

IV. RESULTS

In this section, we present the performance evaluation of our CNN model across different configurations, focusing on the global, local, and fusion branches for three image resolutions: 64×64, 128×128, and 224×224. The evaluation metrics used are Area Under the Curve (AUC), Binary Accuracy and weighted Precision, Recall and F1 Score, with two loss functions: Binary Cross-Entropy (BCE) and Weighted Cross-Entropy (WCE).

A. Performance Analysis

Table 1 summarizes the performance of the global, local, and fusion branches for each resolution. The results indicate that the fusion branch consistently outperforms the global and local branches in terms of AUC, particularly when using the BCE loss function. For instance, at a resolution of 64×64, the fusion branch achieves an AUC of 0.799 with BCE, compared to 0.779 and 0.729 for the global and local branches, respectively. Similarly, at 128×128 resolution, the fusion branch achieves an AUC of 0.777 with BCE, outperforming both the global (0.704) and local (0.667) branches.

When using the WCE loss function, the performance gap between the branches narrows, but the fusion branch still maintains a slight edge. For example, at 224×224 resolution, the fusion branch achieves an AUC of 0.741 with WCE, compared to 0.732 and 0.686 for the global and local branches, respectively.

Resolution	Branch	Loss	AUC	Binary Accuracy
64×64	Global	BCE	0.779	0.949
		WCE	0.727	0.619
	Local	BCE	0.729	0.949
		WCE	0.725	0.779
	Fusion	BCE	0.799	0.949
		WCE	0.745	0.823
128×128	Global	BCE	0.704	0.949
		WCE	0.734	0.854
	Local	BCE	0.667	0.839
		WCE	0.779	0.949
	Fusion	BCE	0.777	0.949
		WCE	0.750	0.812
224×224	Global	BCE	0.757	0.949
		WCE	0.732	0.807
	Local	BCE	0.691	0.949
		WCE	0.686	0.818
	Fusion	BCE	0.776	0.949
		WCE	0.741	0.789

Table 1: Performance comparison of the global, local, and fusion branches for CNN model for three image resolutions.

Resolution	Branch	Loss	Precision	Recall	F1
64×64	Global	BCE	0.18	0.45	0.24
		WCE	0.14	0.52	0.19
	Local	BCE	0.11	0.66	0.18
		WCE	0.13	0.50	0.20
	Fusion	BCE	0.19	0.40	0.25
		WCE	0.16	0.48	0.23
128×128	Global	BCE	0.15	0.46	0.21
		WCE	0.16	0.44	0.22
	Local	BCE	0.11	0.64	0.18
		WCE	0.18	0.45	0.24
	Fusion	BCE	0.15	0.48	0.22
		WCE	0.17	0.43	0.23
224×224	Global	BCE	0.13	0.53	0.21
		WCE	0.13	0.49	0.20
	Local	BCE	0.12	0.59	0.18
		WCE	0.13	0.53	0.20
	Fusion	BCE	0.15	0.49	0.22
		WCE	0.16	0.49	0.23

Table 2: Precision, Recall and F1 Score comparison of the global, local, and fusion branches for CNN model for three image resolutions.

B. Visual Representation

Figure 3 (based on Table 1) provides a visual comparison of the performance across the three branches and resolutions. The plot illustrates the superior performance of the fusion branch, particularly while using binary cross-entropy loss.

C. Time and Memory Usage Comparison

Table 3 presents a comparison of the training time and peak memory usage across different branches (Global, Local, and Fusion) of the simple CNN for three different image resolutions: 64×64, 128×128, and 224×224. For the Global and Local branches, the training time is measured over 50 epochs, while for the Fusion branch, the number of epochs is limited to 17 due to fewer trainable parameters. As the resolution increases, both training time and memory usage generally increase, with the Local branch consistently requiring more memory and time compared to the other branches. The Fusion branch, while requiring fewer epochs, still demands considerable memory, particularly with higher resolutions and when using weighted cross-entropy (WCE) loss.

Figure 3: Performance comparison of the global, local, and fusion branches for CNN model for three image resolutions.

Resolution	Branch	Loss	Training Time (s)	RAM Usage (MB)
64×64	Global	BCE	217.15	3060.16
		WCE	203.59	3468.77
	Local	BCE	1109.60	2166.50
		WCE	254.89	7618.06
	Fusion	BCE	132.01	2582.21
		WCE	144.81	9186.67
128×128	Global	BCE	662.65	2685.39
		WCE	535.46	2941.81
	Local	BCE	566.07	4276.23
		WCE	523.80	9521.59
	Fusion	BCE	384.35	4047.89
		WCE	305.95	9424.65
224×224	Global	BCE	1463.23	3184.02
		WCE	1815.86	2814.91
	Local	BCE	1766.10	7841.15
		WCE	1779.71	9375.21
	Fusion	BCE	744.48	7188.53
		WCE	738.57	9722.27

Table 3: Training time and peak memory usage comparison of the global, local, and fusion branches for the simple CNN for three image resolutions.

D. ROC Curves

All of the previously mentioned results above are based on micro AUC, and as we expect, the macro AUC performance follows the same pattern, as it gets better for the fusion branch. The ROC curves in Figure 4 illustrates an example of these outcomes. We chose the image size 128×128 for binary cross-entropy loss, since it has a significant improvement in AUC performance using the fusion branch over the global branch.

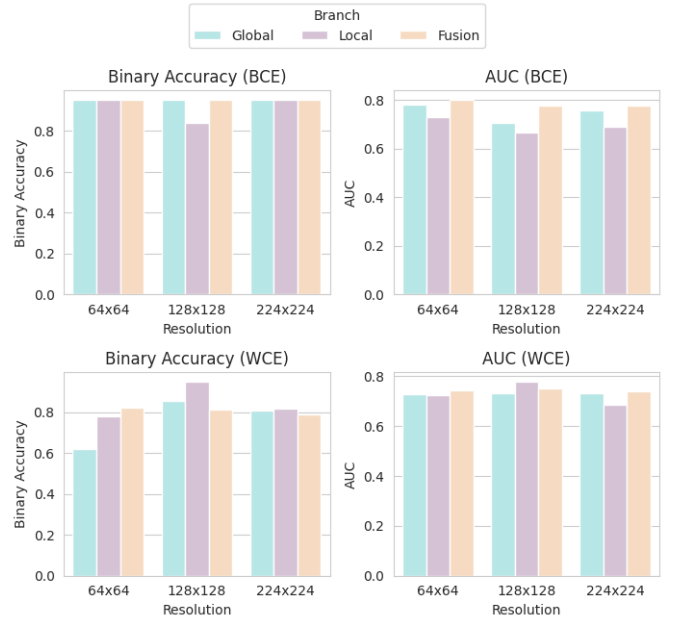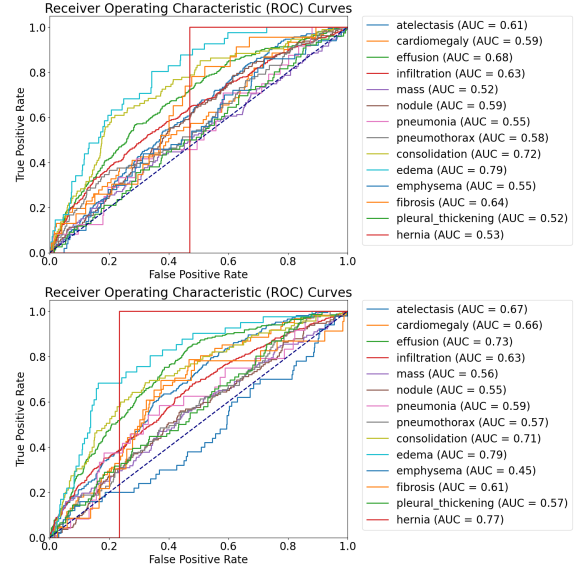Figure 4: ROC Curves for global brach (top) and fusion brach (bottom), 128×128 resolutions with BCE loss.

E. Model Comparison

Even though we did not have the computational power and time needed for training on different models, we confirmed that pre-trained models do not have higher performance on our data. We tested pretrained ResNet-50 and InceptionV3 on image size 224×224 with Weighted Cross-Entropy loss on global branch. Table 4 summarizes the performance across three different models.

Metric	CNN	ResNet-50	InceptionV3
Precision	0.13	0.12	0.10
Recall	0.49	0.58	0.79
F1 Score	0.20	0.19	0.17
Micro AUC	0.732	0.733	0.648
Binary Accuracy	0.807	0.835	0.836

Table 4: Performance comparison of customized CNN, ResNet-50, and InceptionV3 models. Note that Precision, Recall and F1 Score are averaged with weights equal to number of support (positive instances) for each class.

V. CONCLUDING REMARKS

A. Summary

In this work, we presented an attention-guided convolutional neural network for the multi-label classification of chest X-ray14 images. The model integrated both global and local information with the goal to improve diagnostic accuracy, using two different losses (BCE and WCE) and three different image resolutions. (64x64, 128x128, 224x224)

B. Key Findings

The Fusion branch demonstrated strong AUC values across different resolutions and loss functions, with its ROC curves confirming competitive performance across individual classes and macro AUC. These results indicate that the model effectively balances global and local feature representations, making it well-suited for multi-label classification tasks where class imbalances and threshold sensitivity play a significant role. The findings highlight the robustness of the Fusion model and emphasize the importance of evaluating both micro and macro AUC metrics for a comprehensive assessment of model performance.

C. Future Work

Future work will focus on addressing key challenges in the current approach, such as dataset imbalance and sparsity, by exploring techniques like synthetic data generation to improve class representation.

Additionally we aim to improve computational efficiency, and exploring alternative masking strategies to enhance lesion localization such as self-attention (transformer based models) or class-specific attention maps to better distinguish overlapping diseases. Furthermore, by using the whole dataset or different datasets we can gain a better generalization and robustness.

D. Challenges

- **Multilabel Classification:** One of the main challenges in this project was the nature of the classification task, which is multilabel rather than multiclass. This posed difficulties in selecting appropriate evaluation metrics, as traditional single-label metrics like accuracy do not fully capture the performance across multiple labels. Therefore, we opted for area under curve (AUC) and binary accuracy. Additionally, since each label has its own confusion matrix, visualizing the results became less interpretable and more complex. Furthermore, handling class imbalance in a multilabel setting is more challenging than in multiclass classification, as oversampling or undersampling cannot be applied as easily without disrupting label correlations, thus leading us to move on with the original, imbalanced data.

- **Runtime and GPU:** Another significant challenge was the computational cost and runtime constraints due to the size of the dataset. While we ran all experiments using our customized CNN model, we only tested ResNet-50 and InceptionV3 with weighted cross-entropy loss and an input size of 224 - solely on the global branch - to compare them with our model. Due to resource limitations, we could not afford to run these models for every input size, loss function, and across the local and fusion branches. Additionally, we used only 10% of the dataset to manage computational overhead. The actual challenge was not the training time but the process of masking and saving masked images as shown in Table 5, which was the main part of the overall runtime. Our access to GPU resources was limited, and we were unable to afford stronger or longer GPU usage, further constraining our ability to conduct more extensive experiments.

Resolution	Loss	Masking Time (minutes)
64 × 64	BCE	67
	WCE	78
128 × 128	BCE	74
	WCE	88
224 × 224	BCE	109
	WCE	140

Table 5: Masking times for different resolutions and loss functions.

- **Sparsity and Imbalance:** Another major challenge was dealing with both sparsity and class imbalance in our dataset. The dataset contained a high proportion of zeros and very few ones, making it highly sparse. Even after incorporating higher weights for positive instances, as suggested in Wang et al. [1], the models struggled to learn meaningful patterns. In some cases, these weights even worsened performance, highlighting the difficulty of addressing sparsity. Additionally, class imbalance further hindered training, as certain labels were significantly underrepresented, making it harder for the model to generalize across all classes. The Disease Co-occurrence Diagram (Figure 5) shows another complication in the model's ability to generalize and effectively learn from the data, since the model has difficulty learning the correlations between diseases and their respective features

Label	Count	Percentage (%)
Atelectasis	1140	10.17
Cardiomegaly	302	2.69
Effusion	1338	11.94
Infiltration	1999	17.83
Mass	582	5.19
Nodule	616	5.50
Pneumonia	146	1.30
Pneumothorax	512	4.57
Consolidation	474	4.23
Edema	219	1.95
Emphysema	269	2.40
Fibrosis	151	1.35
Pleural Thickening	341	3.04
Hernia	17	0.15
No Finding	**6040**	**53.88**

Table 6: Class-Wise Count of Positive Instances and Label Vectors with No Findings, Highlighting Imbalance in Label Distribution (10% of Data)
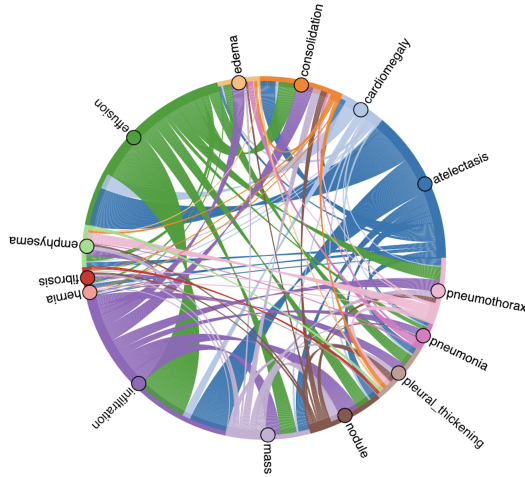

Figure 5: Disease co-occurrence diagram(10% of Data)

- **Transition Layer Modification:** Another challenge was related to the masking process described in the article [1], where a transition layer was used, but its functionality was not very clear. Instead of relying on the transition layer, we opted to directly use activations from the last convolutional layer for our masking process. To align the output with the input size, we resized the activations from the last layer, bypassing the transition layer entirely. This approach was chosen due to the lack of clarity on how the transition layer contributed to the masking process.

REFERENCES

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2097–2106, 2017.

[2] "ChestMNIST Dataset." https://zenodo.org/records/10519652.

[3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[4] J. Irvin, P. Rajpurkar, M. Ko, Y. Xu, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *arXiv preprint arXiv:1901.07031*, 2019.

[5] J. P. Cohen, M. Luck, and L. Muller, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.

[6] H. Zhang, X. Zhang, and Q. Huang, "Attention-guided convolutional neural network for thoracic disease classification," in *Proceedings of the International Conference on Medical Imaging with Deep Learning (MIDL)*, pp. 191–201, 2021.

[7] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian, and Y. Yang, "Progressive learning for person re-identification with one example," vol. 28, pp. 2872–2881, June 2019.

[8] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camstyle: A novel data augmentation method for person re-identification," vol. 28, pp. 1176–1190, Mar. 2019.

[9] L. Zhu, Z. Xu, Y. Yang, and A. G. Hauptmann, "Uncovering the temporal context for video question answering," *International Journal of Computer Vision*, vol. 124, no. 3, pp. 409–421, 2017.

[10] H. Fu, Y. Xu, S. Lin, X. Zhang, D. W. K. Wong, J. Liu, A. F. Frangi, M. Baskaran, and T. Aung, "Segmentation and quantification for angle-closure glaucoma assessment in anterior segment oct," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1930–1938, 2017.

[11] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[12] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: A review," *Journal of Medical Systems*, vol. 42, p. 226, Oct. 2018.

[13] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 31–40, 2017.

[14] R. Roy, T. Chakraborti, and A. S. Chowdhury, "A deep learning-shape driven level set synergism for pulmonary nodule segmentation," *Pattern Recognition Letters*, vol. 123, pp. 31–38, May 2019.

[15] Q. Guan and Y. Huang, "Multi-label chest x-ray image classification via category-wise residual attention learning," *Pattern Recognition Letters*, vol. 130, pp. 1–10, 2018.

[16] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, B. N. Patel, K. W. Yeom, K. Shpanskaya, F. G. Blankenberg, J. Seekins, T. J. Amrhein, D. A. Mong, S. S. Halabi, E. J. Zucker, A. Y. Ng, and M. P. Lungren, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists," *PLOS Medicine*, vol. 15, pp. 1–17, Nov. 2018.

[17] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *CoRR*, vol. abs/1710.10501, 2017.

[18] P. Kumar, M. Grewal, and M. M. Srivastava, "Boosted cascaded convnets for multilabel classification of thoracic diseases in chest radiographs," in *Image Analysis and Recognition* (A. Campilho, F. Karray, and B. ter Haar Romeny, eds.), (Cham), pp. 546–552, Springer International Publishing, 2018.