

# Integrating Spatial and Channel Features for Multi-Disease Classification Using Attention-Based CNN

Thudum Venkatesh<sup>1,\*</sup>, Dantam Ramesh<sup>2</sup>

<sup>1</sup>Research Scholar, Department of CSE, JNTUH

<sup>2</sup>Professor of CSE, JNTUH University College of Engineering Jagtial

\*Corresponding Author: [venkatesh.thudum@gmail.com](mailto:venkatesh.thudum@gmail.com)

## ARTICLE INFO

## ABSTRACT

Received: 31 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

The increasing adoption of deep learning in medical image analysis has enabled significant improvements in disease detection and classification. This paper presents a multiple disease classification model that integrates spatial and channel feature extraction with a Convolutional Neural Network (CNN) enhanced by an attention mechanism. The model is trained on a chest X-ray images to classify four different diseases, and its performance is evaluated using key metrics. Experimental results demonstrate that the proposed model achieves an accuracy of 94.0%, with an average precision of 94.4%, recall of 94.6%, and F1-score of 93.4%, outperforming conventional CNN architectures. An ablation study was conducted to show the effectiveness of the spatial and channel feature extraction components in improving classification performance.

**Keywords:** Multiple disease, CNN, X-ray, Spatial attention, Channel attention

## INTRODUCTION

The fast advancements in deep learning have significantly transformed medical image analysis, enabling automated and efficient disease diagnosis. Traditional machine learning methods uses handcrafted features, which may not capture the complex variations present in medical images. CNNs have emerged as a powerful tool for extracting meaningful spatial patterns from medical images, demonstrating superior performance in disease classification tasks. However, conventional CNN models may fail to effectively leverage both spatial and channel-wise information, which are crucial for accurate classification. To address this limitation, attention mechanisms have been integrated into CNN architectures [26] [27], to capture most complex patterns from medical image.

Multi-disease classification remains a challenging problem due to inter-class similarities, variations in image acquisition, and imbalanced datasets. The ability to classify multiple diseases from medical images is essential for efficient and accurate diagnostics, reducing the workload on radiologists and improving patient outcomes. Various studies have explored deep learning-based techniques for disease classification, highlighting the importance of feature extraction and attention mechanisms in enhancing model performance.

The effectiveness of attention-based deep learning models has been demonstrated in various domains, including natural language processing and computer vision. In medical image analysis, attention mechanisms will capture more complex features to focus on disease-specific patterns, improving interpretability and robustness. The integration of both spatial and channel attention mechanisms allows deep learning models to learn hierarchical representations, ensuring better discrimination between disease classes. Unlike conventional CNN [41] architectures that treat all features equally, attention-based approaches dynamically reweight feature maps, enhancing the model's ability to differentiate between subtle variations in pathological conditions.

Chest X-ray (CXR) [9] [11] [24] imaging is widely used in the diagnosis of pulmonary diseases due to its complex features and accessibility. However, distinguishing between different respiratory diseases using CXRs is challenging due to overlapping radiological features and subtle differences in disease manifestations. Diseases such as pneumonia, COVID-19 [7] [28] [30], tuberculosis, and lung opacity often exhibit similar patterns in X-ray images,

making accurate classification a difficult task. Deep learning models [29] incorporating spatial and channel-wise feature extraction have shown promising results in improving disease classification performance, enabling faster and more reliable diagnoses.

Recent advancements in deep learning [14] [18] have demonstrated the potential of CNN-based models combined with attention mechanisms for multi-disease classification in medical imaging. These models enhance feature representation by selectively emphasizing critical regions in medical images, leading to improved classification performance. Furthermore, studies have highlighted the role of ablation experiments in assessing the impact of different model components, and the effectiveness of various feature extraction techniques.

The integration of attention mechanisms [42] with CNNs has demonstrated significant improvements in multi-disease classification, paving the way for more accurate and interpretable AI-driven medical diagnosis systems. To overcome these challenges we conduct extensive experiments on a multi-disease classification dataset, evaluating the impact of spatial and channel attention mechanisms on model performance. Additionally, we perform an ablation study to analyze the contribution of each component in our model. Our results indicate that the proposed CNN with attention outperforms traditional CNN architectures, demonstrating higher accuracy and robustness in multi-disease classification tasks.

### Contributions:

- a) We propose a novel deep learning model that integrates spatial and channel attention mechanisms to improve feature extraction for multi-disease classification.
- b) Our model effectively captures both spatial and channel-wise features, enabling better differentiation between multiple diseases.
- c) We conduct a detailed ablation study to assess the impact of different components, ensuring the effectiveness of our proposed approach.
- d) Extensive experiments are conducted on a medical image dataset, demonstrating the superiority of our model in terms of classification accuracy and generalization.

### RELATED WORK

Alshmrani, G. M. M et al (2023) [1] worked on multiple disease detection like Pneumonia, Lung Cancer, tuberculosis (TB), Lung Opacity, and COVID-19 from chest X-ray, for this they used 20,00 lung images, trained on VGG+CNN hybrid models and achieved an accuracy of 0.96. In their pre-processing each image is converted to  $224 \times 224 \times 3$  a 3 dimensional matrix for training, and used many number of parameters, only extracted RGB based features. Yimer, F., et al (2021) also worked on multiple disease detection from X-ray of 11716 samples, for this fine tuned ImageNet and got an accuracy of 0.97, but extracted one dimensional features. Kabiraj, A., et al (2022) in [3] fine tuned EfficientNet model to classify thirteen thoracic lung diseases, and got an average accuracy of 0.88. Banerjee, S., et al (2020) [4] implemented federated learning model to classify viral and bacterial pneumonia and used various optimization methods like SGD, and also used Grad-CAM to identify pneumonia affected regions.

Kim, S., et al (2022) [5] used EfficientNet v2-M model to extract complex features from CXR samples to classify 3 classes like normal, pneumonia, and pneumothorax and got an accuracy of 0.82. But this model is getting over fitted due less number of samples. Allaouzi, I., and Ahmed, M. B. (2019) [6] extracted features with DenseNet model, on top this added classifier to detect the diseases. The data used in this approach is 134,327 CXRs samples, and also has different views level features. They detected 14 diseases from x-ray image, with an accuracy of 0.84. Baltruschat, I. M., et al (2019) [8] implemented multiple ResNet with different depths, used Chest X-ray 14 data set, and classified multiples disease. Average accuracy over all classes in 0.89, also used Grad-cam to find the disease affected regions.

Ibrahim, D. M., et al (2021) [10] used 4 different model like VGG19-CNN, ResNet152V2, ResNet152V2 + and Gated Recurrent Unit (GRU) to classify disease COVID-19, pneumonia, and lung cancer chest diseases. And added a fully connected layer on top of all these models to classify the disease. The average accuracy they for with hybrid model is 0.98 with ResNet152V2+Bi-GRU. Chen, J. I. Z. (2021). [12] Implemented customized CNN model trained on

COVID-19 thoracic x-rays, trained 10 fold cross validation method and achieved an accuracy of 0.82. Sanida, M. V., et al (2024) [13] worked multiple disease like opacity, tuberculosis, normal, viral pneumonia, and COVID-19 pneumonia from chest X-ray images. Implemented customized CNN model, got an accuracy of 0.98. Hussain, E., et al (2021) [15] implemented customized CNN model with 22 layers, and classified multiple diseases like COVID-19, Pneumonia-viral, Pneumonia-bacterial with an accuracy of 0.912.

Ozturk, T., et al (2020) [16] worked on binary and multiple disease detection, by training YOLO method, for binary class classification they got 0.98 and for multi class classification got an accuracy of 0.87. Pillai, A. S. (2022). [17] And [25] Worked on AI and ML method to detect 14 diseases, and the data set consist of x-ray views from different sides, and some statistical features like age, gender etc. Chen, K. C., et al (2020) [19] trained YOLOv3 model to detect bronchiolitis different types of pneumonia, the data set consist of children and adult digital samples. And got an average accuracy below 0.90. Xu, J., et al (2021) [20] used Chest X-Ray14 and CheXpert datasets consist of 100,000 and 200,000 front view and back view of samples. Implemented a neural network model with novel loss function and got an AUC of 0.85.

Shelke, A., et al (2021) [21] used covid-19 and normal chest X-ray samples to classify pneumonia, TB, and normal with an accuracy of 0.95, with VGG16 model, and also trained DenseNET model to classify covid-19 with a test accuracy of 0.76. Gupta, A., et al (2021) [22] and [23] implemented InstaCovNet-19 model with a pre trained model like ResNet101, Xception, InceptionV3, MobileNet, and used Covid-19 x-ray data, they did for 2 and 3 disease detection. And got an accuracy of 0.99 and 0.99 for 2 and 3 classes. Many of the researcher have used either Covid-19 samples or Chest-14 data sets for classification of multiple diseases like [31] [32] [36] [37] [39] [40] have used Covid data by using the pre trained models like ResNet, DenseNet, VGG, YOLO and ResNet models and the researchers like [33] [34] and [35] have used X-ray data set, but 90% if the model approaches completely used pre trained models, but all these models are block box model will not provide what features they extracted from which part of the image is affected with disease is not clear.

## METHODOLOGY

The proposed model integrates a Convolutional Block Attention Module (CBAM) within a deep learning framework to enhance feature extraction by incorporating both **channel-wise and spatial attention mechanisms**. CBAM includes two approaches like **Channel Attention (CA)** and **Spatial Attention (SA)**, will captures different angled features, to improve the network's discriminative power. Given an input feature map  $X \in \mathbb{R}^{C \times H \times W}$ , where **C** represents the number of channels, and **H**, **W** denote the spatial dimensions, the **Channel Attention Mechanism** computes attention weights across channels using both **average pooling** and **max pooling** operations with equation (1). This ensures the model effectively captures inter-channel dependencies, thereby emphasizing the most informative feature maps.

$$M_c(X) = \sigma \left( f_c(\text{Avg\_pool}(X)) + F_c(\text{Max\_pool}(X)) \right) \quad (1)$$

where  $f_c$  represents a two-layer convolutional network with a reduction ratio  $r$ , and  $\sigma$  is the sigmoid activation function.

First the input vector is passed through **Channel Attention module** that assigns different importance levels to each channel through a two-layer **fully connected network** with a reduction factor to reduce computational complexity. It first applies **global average pooling** to aggregate feature representations, followed by two FC layers with **ReLU activation** and a **sigmoid function will generate non linear data**. These weights are multiplied with the original feature maps, selectively enhancing relevant channels while passing less informative ones. The input feature map is refined by element-wise multiplication with the channel attention mask with equation (2).

$$X' = M_c(X) \cdot X \quad (2)$$

Subsequently, the **Spatial Attention module** complements Channel Attention  $f_c$  by focusing on significant **spatial patterns** in the feature maps. It computes spatial attention using channel-wise **average pooling**,

followed by a  $7 \times 7$  **convolutional layer** to learn spatial dependencies. The output is then passed through a **sigmoid activation function** with equation (3) to generate attention weights, which are multiplied with the input feature maps to enhance critical spatial features while reducing background noise.

$$M_c(X') = \sigma \left( f_s(\text{concat}[\text{Avg\_pool}(X'), (\text{Max\_pool}(X'))]) \right) \quad (3)$$

Where  $f_s$  is a  $7 \times 7$  convolutional layer applied to the concatenated pooled features. The final attention-enhanced feature map is obtained with equation (4).

$$X'' = M_c(X') \cdot X' \quad (4)$$

To fully leverage these attention mechanisms, the proposed model incorporates a deeper **fully connected classification network**, CBAM. This component consists of four sequential **linear hidden layers as equation (5) and (6)**, progressively reducing the feature dimensionality from **512 to 10**. Each layer utilizes **sigmoid activation** to facilitate hierarchical feature learning while mitigating vanishing gradient issues. The final layer maps extracted feature embeddings to class predictions, making the model highly effective for **multi-class classification tasks with equation (7)**.

$$H_1 = \sigma(W_1 X'' + b_1) \quad (5)$$

$$H_2 = \sigma(W_2 H_1 + b_2) \quad (6)$$


$$y_p = \text{softmax}(W_3 H_2 + b_3) \quad (7)$$

Where  $H_1, H_2$  are hidden layers with sizes 128 and 64, respectively, and  $W_i, b_i$  represent weight matrices and biases. By integrating **CBAM with a deeper classification network**, the proposed model significantly improves **feature selection and classification performance**.

## DATASET AND PREPROCESSING

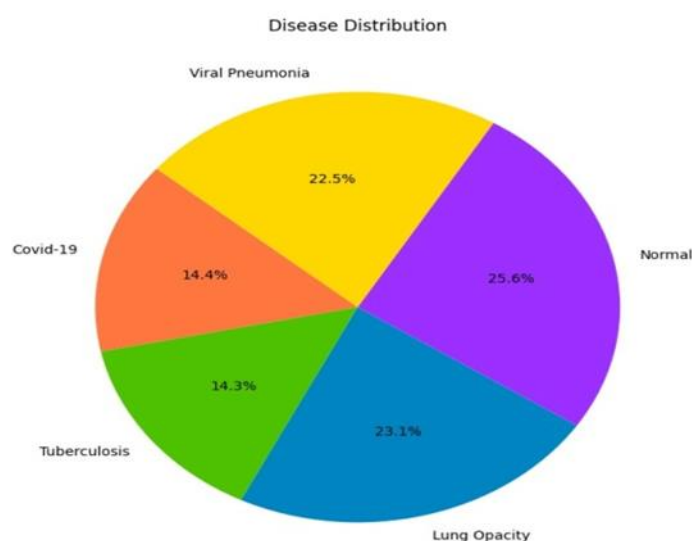
The dataset used in this study comprises chest X-ray (CXR) images categorized into five distinct classes: **COVID-19, Tuberculosis, Lung Opacity, Normal, and Viral Pneumonia**. These classes represent a diverse range of pulmonary conditions, allowing the model to generalize well across different respiratory diseases. The images were sourced from multiple datasets available on Kaggle, each containing varying numbers of samples per class. Specifically, the dataset consists of **704 COVID-19, 700 Tuberculosis, 1125 Lung Opacity, 1250 Normal, and 1100 Viral Pneumonia images as shown in Figure 1**. Given the inherent challenges in medical image classification, such as imbalanced class distributions and variations in image quality, appropriate preprocessing techniques were applied to ensure optimal model performance.

**Table 1** sample image from **Tuberculosis** and corresponding Image vector after normalization.

|   |  |
|---|--|
|  | <p>[0.19548959, 0.032374833, 0.8091909, 2.011572, 0.7416156, 0.9703078, 0.021886688, 0.29910952, 0.3221549, 0.47343925, 0.59963065, 0.2573957, 1.913524, 0.5280585, 2.5191782, 0.933251, 0.030268257, 2.0389175, 0.8341835, 0.48514938, 0.4156689, 0.49167904, 1.6397014, 0.5832021, 1.8869599, 1.9788432, 1.2028855, 0.17473242, 0.6688879, 0.16237481, 0.73982537, 1.5041256, 0.8798962, 0.40174264, 0.7051053, 0.06426582, 0.6714343, 0.8731386, 0.81580436, 0.22636929, 2.9304593, 0.4138742, 1.2477546, 0.5506517, 2.8935385, 1.3012506, 0.124193214, 3.0198662, 2.1468701, 0.051048316, 0.86909646, 0.050680228, 0.5549408, 0.017481085, 0.3513011, 1.3187522, 1.16335, 4.310297, 1.0183234, 3.028105, 0.10314953, 0.1580841, 0.2086794, 0.39904353, 0.058631927, 0.44340855, 0.5203699, 0.10102598, 0.8773394, 0.1447529, 0.10900897, 0.58900833, 0.4284491, 0.1427274, 1.3970008, 1.0086709, 0.1421501, 2.3806024, 0.17476553, 1.0110075, 0.39966834, 0.5158894, 0.9014908, 0.09578547, 0.021607142, 0.1526772, 0.93262655, 0.8762444, 1.7545271, 0.48778483, 1.5623596, 0.34287184, 0.17646903, 0.14250919, 1.0557555, 0.517391, 2.116281, 0.5693241, 0.80620086, 2.4996798, ...]</p> |
|---|--|

Rather than relying solely on raw pixel data, a feature extraction approach was implemented using **Img2Vec**, a deep learning-based vectorization method. Each image was converted into a **512-dimensional feature vector**, capturing essential spatial and textural information necessary for classification. This transformation was performed using the **PIL (Pillow) library**, ensuring that all images were standardized into a three-channel RGB format before vectorization. By leveraging deep feature embeddings instead of raw pixel values, the model could focus on high-level patterns in the images, enhancing its robustness against variations in image resolution and contrast.

Once the feature vectors were generated, the dataset is divided into **training and testing sets** at an **80-20 split ratio for training and testing at random state of 40**. The labels were also preserved during the split to maintain a balanced distribution across different disease classes.



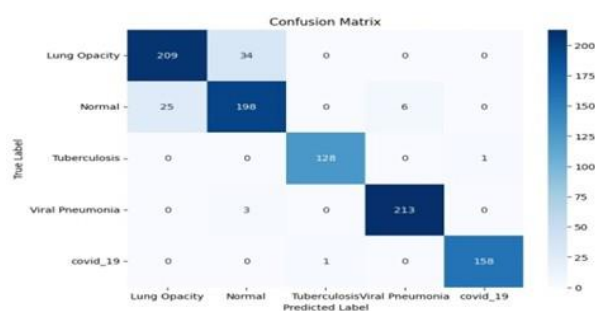
**Figure 1** number of samples for each disease

## RESULT ANALYSIS

The training process of the proposed model was conducted over **ten epochs**, with the objective of minimizing the classification loss and enhancing the model's ability to distinguish between multiple lung diseases. The loss values recorded at each epoch demonstrate a **consistent downward trend**, indicating that the model effectively learns discriminative features from the input images. In the initial epoch, the loss was relatively high **0.8149**, reflecting the model's initial state with randomly initialized parameters. However, with subsequent iterations, a **sharp decline** in loss was observed, dropping to **0.2628** by the second epoch. This suggests that the model rapidly adapted to key patterns in the data within the early training phase.

From the **third epoch onward**, the rate of loss reduction became more gradual, signifying that the model was approaching a more refined representation of the features necessary for classification. By the **fifth epoch**, the loss had decreased to **0.1751**, demonstrating significant learning progress. Interestingly, even though the initial plan was to train for five epochs, the training was extended to ten epochs. The loss values from the **sixth to tenth epoch** continued to fluctuate slightly but maintained an overall downward trend, with the final loss reaching **0.1452**.





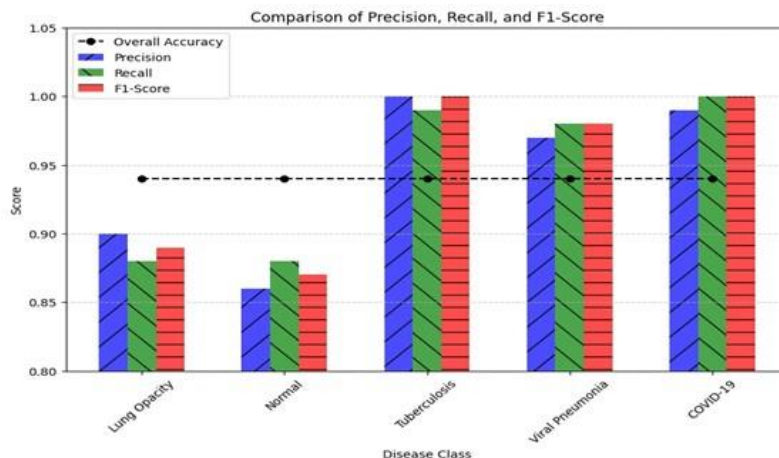
**Figure 2** confusion matrix of proposed CBAM with attention model.

The proposed **CNN-based model with attention mechanisms** demonstrated **high classification accuracy** across five disease categories: **Lung Opacity**, **Normal**, **Tuberculosis**, **Viral Pneumonia**, and **COVID-19**. The results indicate that the model effectively differentiates between various lung diseases, showcasing its reliability for automated diagnosis.

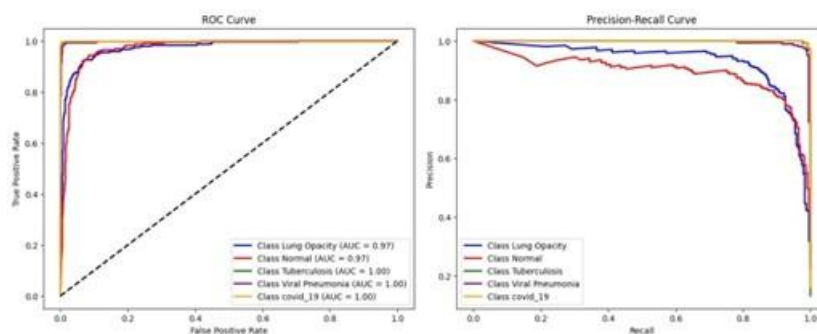
For **Lung Opacity**, the model correctly identified **88%** of cases, with a slight margin of misclassification. Similarly, for **Normal** cases, the model achieved an accuracy of **88%**, successfully distinguishing them from diseased samples, though some instances were incorrectly labeled.

The model performed exceptionally well in detecting **Tuberculosis**, accurately identifying **99%** of cases. A similar trend was observed for **Viral Pneumonia**, where the model achieved an accuracy of **98%**, with minimal errors.

For **COVID-19**, the model reached an accuracy of **100%**, correctly identifying all cases with no false negatives. This high accuracy is crucial in clinical applications, ensuring reliable identification of COVID-19 cases and minimizing the risk of undetected infections as illustrated in Figure 3.



**Figure 3** class wise performance of CBAM with attention model



**Figure 4** ROC and precision, recall curves of proposed model

From the figure 4 ROC curve and the Precision-Recall (PR) curve for different disease classes, including Lung Opacity, Normal, Tuberculosis, Viral Pneumonia, and COVID-19. The ROC curve on the left demonstrates the difference between positive and negative instances across different threshold values. The AUC values indicate strong classification performance, with Tuberculosis and COVID-19 achieving near-perfect scores  $AUC = 1.00$ , while other classes also maintain high discrimination capabilities. The PR curve on the right illustrates the trade-off between precision and recall, highlighting the model’s robustness in correctly identifying positive cases. The curves suggest that the classifier performs exceptionally well across all classes, particularly for Tuberculosis and COVID-19, which maintain high precision across various recall levels.

The figure 5 presents Grad-CAM visualization, which provides interpretability for the deep learning model’s decision-making process. The left side displays the original chest X-ray image, while the right side overlays the Grad-CAM heatmap, revealing the most influential regions contributing to the model’s classification. The highlighted areas illustrate the regions the model focuses on while predicting a specific disease. The visualization confirms that the model effectively localizes disease-relevant regions, such as lung opacities and abnormalities, further validating its clinical applicability.

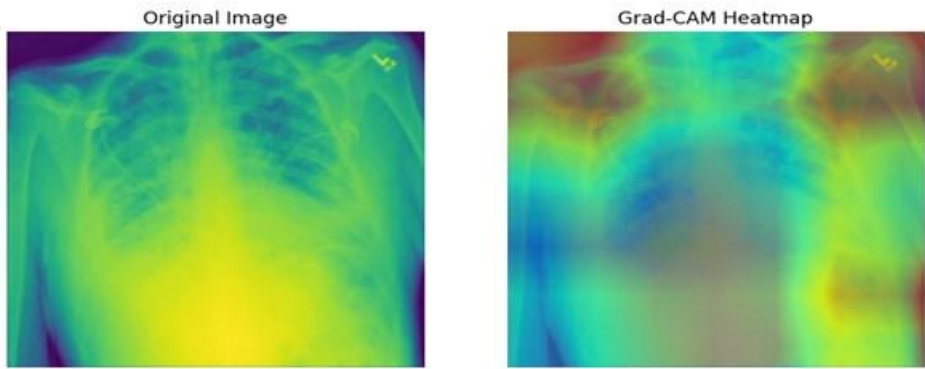


Figure 5 features extracted from X-ray samples for training

Ablation Study

To evaluate the impact of different architectural modifications on classification performance, an **ablation study** was conducted by systematically training and evaluating multiple model variations, including the **Spatial Attention, channel attention CBAM, Deeper Layers, and Reduced Hidden Layers as illustrated in table 2**. The results provide insights into how attention mechanisms and architectural depth influence learning efficiency and generalization.

Table 2 ablation model with the parameters

| Model Name        | Number of Layers | Number of Parameters | Attention Mechanism | Notable Modifications                              |
|-------------------|------------------|----------------------|---------------------|--|
| Channel Attention | 3                | ~74K                 | Channel Attention   | Feature recalibration using global average pooling |
| Spatial Attention | 3                | ~72K                 | Spatial Attention   | Convolution-based spatial features                 |
| CBAM              | 3                | ~76K                 | Channel + Spatial   | Combination of channel and spatial attention       |

|                       |   |       |      |  |
|-----------------------|---|-------|------|--|
| Deeper Layers         | 4 | ~100K | None | Increased hidden layers for deeper to increase complexity                |
| Reduced Hidden Layers | 3 | ~50K  | None | Fewer neurons per layer to reduce model complexity compare to base model |

The **Base Model** training showed a steady decrease in loss, starting from **0.78** in the first epoch to **0.15** by the tenth epoch, demonstrating effective convergence. Introducing **Spatial Attention** initially resulted in a higher loss but improved learning dynamics, with the final loss reaching **0.15**, indicating enhanced feature representation through spatial attention mechanisms.

Further refinement using **CBAM** exhibited an initial loss of **0.90**, which was slightly higher than other models, but it progressively reduced to **0.15**, suggesting that incorporating both spatial and channel attention optimizes feature selection.

Increasing the **model depth** further improved learning stability as shown in Figure 6, with the **Deeper Layers** model reaching a lower final loss of **0.14**, signifying better hierarchical feature extraction. Conversely, reducing the number of hidden layers led to **slower convergence**, with the **Reduced Hidden Layers** model having the highest initial loss (**0.98**) and a relatively higher final loss of **0.16**, indicating suboptimal learning due to insufficient feature abstraction.

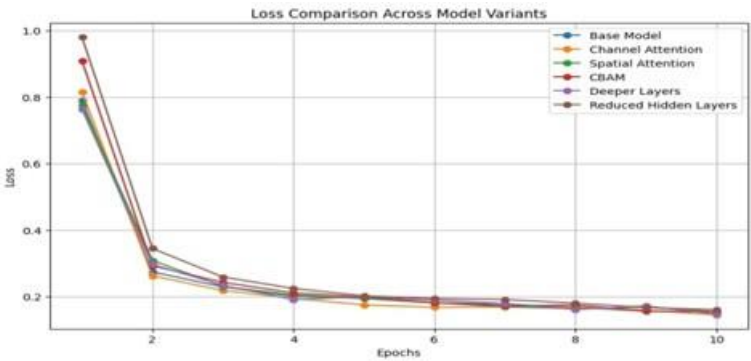


Figure 6 Ablation models training loss comparison

**Base Model** demonstrated strong classification ability, achieving an overall performance of **0.94** across all disease classes as shown in figure 9. Among individual categories, **Tuberculosis** and **COVID-19** exhibited the highest recognition, both exceeding **0.99**, indicating that the model effectively differentiates these conditions. **Lung Opacity** and **Normal** cases showed slightly lower recognition, around **0.89** and **0.87**, respectively.

Integrating **Deeper Layers** resulted in marginally reduced overall performance, reaching **0.93**. **Tuberculosis** and **COVID-19** maintained near-perfect identification, both above **0.99**. However, **Lung Opacity** showed a drop, reaching **0.88**, while **Normal** cases increased slightly to **0.86**.

From Figure 7 the **Channel Attention** model performed similarly, with an overall value of **0.93**. The identification of **Lung Opacity** and **Normal** was slightly lower than the **Base Model**, registering **0.88** and **0.85**, respectively. Meanwhile, **Viral Pneumonia**, **Tuberculosis**, and **COVID-19** maintained high recognition, exceeding **0.98**.



The **Spatial Attention** model showed a slight decrease, achieving **0.92** overall. **Lung Opacity** and **Normal** cases had the most noticeable impact, dropping to **0.87** and **0.84**, respectively. The performance in identifying **Viral Pneumonia**, however, remained stable at **0.97**, with **Tuberculosis** and **COVID-19** continuing to be well-recognized.

The **CBAM** model achieved a comparable **0.92**, with **Lung Opacity** improving to **0.88**, while **Normal** cases showed a slight decline to **0.84**. The recognition of **Tuberculosis**, **Viral Pneumonia**, and **COVID-19** remained consistently high.

The **Reduced Hidden Layers** model exhibited similar overall performance to **Deeper Layers** as illustrated in **Figure 8**, with a value of **0.93**. **Lung Opacity** recognition dropped slightly to **0.87**, while **Normal** cases improved to **0.86**. **Tuberculosis**, **Viral Pneumonia**, and **COVID-19** remained at strong levels, surpassing **0.98**.

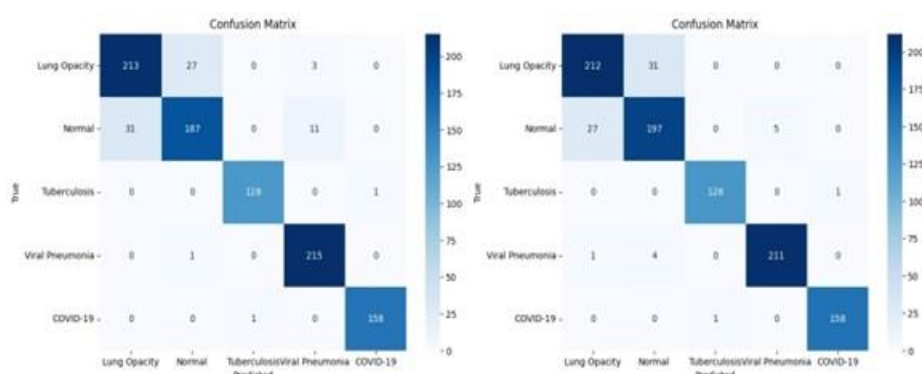


Figure 7 confusion matrix of spatial and channel attention models

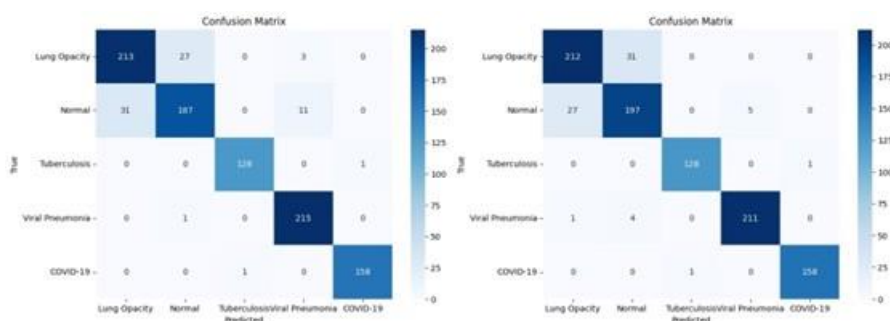


Figure 8 confusion matrixes of deeper and reduced layers models

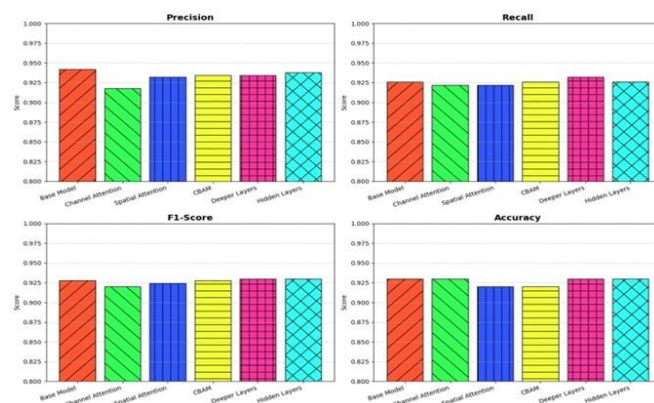
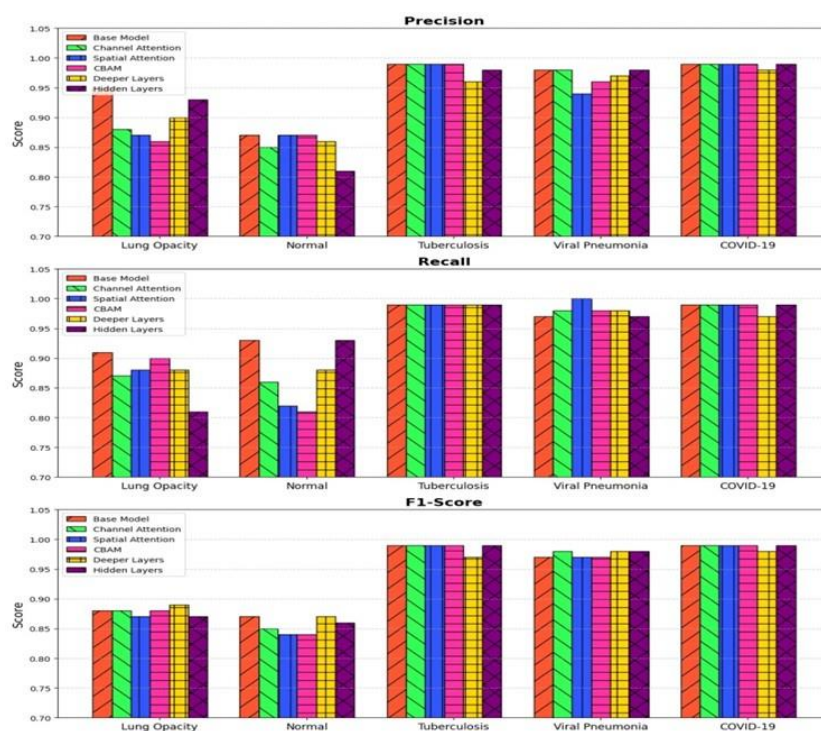


Figure 9 comparison of ablation models performance



**Figure 10** comparison of ablation models performance on precision recall and F1 score

The figure 10 presents a comparative analysis of different model variations based on precision, recall, and F1-score across five disease classes: Lung Opacity, Normal, Tuberculosis, Viral Pneumonia, and COVID-19. The models compared include the base model, channel attention, spatial attention, CBAM, deeper layers, and additional hidden layers; each represented using distinct hatch patterns. The precision plot indicates that while most models perform consistently well for Tuberculosis, Viral Pneumonia, and COVID-19, there are notable variations for Lung Opacity and Normal classes, with deeper layers and hidden layers showing improvements. The recall plot demonstrates that all models achieve near-perfect recall for Tuberculosis, Viral Pneumonia, and COVID-19, while some variations show differences in the Lung Opacity and Normal classes. The F1-score plot consolidates these findings, reinforcing those models incorporating attention mechanisms and deeper architectures generally enhance classification performance. The results suggest that architectural modifications, particularly attention-based enhancements, contribute to improved disease classification, particularly in challenging cases like Lung Opacity and Normal conditions.

## CONCLUSION

This study introduces a deep learning-based multiple disease classification model that utilizes spatial and channel feature extraction along with an attention approach to enhance performance. The proposed model demonstrates **94.0%** accuracy, **94.4%** precision, **94.6%** recall and F1-score of **93.4%**, outperforming traditional CNN-based models. The ablation study confirms the contribution of each feature extraction component in refining the classification capability of the model. While the results are promising, challenges such as dataset imbalance and model generalization require further investigation. Future work will focus on expanding the dataset, incorporating transformer-based architectures, and optimizing the attention mechanism to improve robustness and clinical applicability. And also includes multiple disease detection with explainable AI, to interpret the disease affected regions in the image.

## REFERENCES

- [1] Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal*, 64, 923-935.

- [2] Yimer, F., Tessema, A. W., & Simegn, G. L. (2021). Multiple lung diseases classification from chest X-ray images using deep learning approach. *Int. J*, 10, 2936-2946.
- [3] Kabiraj, A., Meena, T., Reddy, P. B., & Roy, S. (2022, October). Detection and classification of lung disease using deep learning architecture from x-ray images. In *International Symposium on visual computing* (pp. 444-455). Cham: Springer International Publishing.
- [4] Banerjee, S., Misra, R., Prasad, M., Elmroth, E., & Bhuyan, M. H. (2020). Multi-diseases classification from chest-x-ray: A federated deep learning approach. In *AI 2020: Advances in Artificial Intelligence: 33rd Australasian Joint Conference, AI 2020, Canberra, ACT, Australia, November 29–30, 2020, Proceedings 33* (pp. 3-15). Springer International Publishing.
- [5] Kim, S., Rim, B., Choi, S., Lee, A., Min, S., & Hong, M. (2022). Deep learning in multi-class lung diseases' classification on chest X-ray images. *Diagnostics*, 12(4), 915.
- [6] Allaouzi, I., & Ahmed, M. B. (2019). A novel approach for multi-label chest X-ray classification of common thorax diseases. *IEEE Access*, 7, 64279-64288.
- [7] Khan, E., Rehman, M. Z. U., Ahmed, F., Alfouzan, F. A., Alzahrani, N. M., & Ahmad, J. (2022). Chest X-ray classification for the detection of COVID-19 using deep learning techniques. *Sensors*, 22(3), 1211.
- [8] Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., & Saalbach, A. (2019). Comparison of deep learning approaches for multi-label chest X-ray classification. *Scientific reports*, 9(1), 6381.
- [9] Nawaz, M., Nazir, T., Baili, J., Khan, M. A., Kim, Y. J., & Cha, J. H. (2023). CXray-EffDet: chest disease detection and classification from X-ray images using the EfficientDet model. *Diagnostics*, 13(2), 248.
- [10] Ibrahim, D. M., Elshennawy, N. M., & Sarhan, A. M. (2021). Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Computers in biology and medicine*, 132, 104348.
- [11] Shamrat, F. J. M., Azam, S., Karim, A., Ahmed, K., Bui, F. M., & De Boer, F. (2023). High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. *Computers in Biology and Medicine*, 155, 106646.
- [12] Chen, J. I. Z. (2021). Design of accurate classification of COVID-19 disease in X-ray images using deep learning approach. *Journal of ISMAC*, 3(02), 132-148.
- [13] Sanida, M. V., Sanida, T., Sideris, A., & Dasygenis, M. (2024). An advanced deep learning framework for multi-class diagnosis from chest X-ray images. *J*, 7(1), 48-71.
- [14] Moses, D. A. (2021). Deep learning applied to automatic disease detection using chest X-rays. *Journal of Medical Imaging and Radiation Oncology*, 65(5), 498-517.
- [15] Hussain, E., Hasan, M., Rahman, M. A., Lee, I., Tamanna, T., & Parvez, M. Z. (2021). CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos, Solitons & Fractals*, 142, 110495.
- [16] Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine*, 121, 103792.
- [17] Pillai, A. S. (2022). Multi-label chest X-ray classification via deep learning. *arXiv preprint arXiv:2211.14929*.
- [18] Yan, C., Yao, J., Li, R., Xu, Z., & Huang, J. (2018, August). Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics* (pp. 103-110).
- [19] Chen, K. C., Yu, H. R., Chen, W. S., Lin, W. C., Lee, Y. C., Chen, H. H., ... & Lu, H. H. S. (2020). Diagnosis of common pulmonary diseases in children by X-ray images and deep learning. *Scientific reports*, 10(1), 17374.
- [20] Xu, J., Li, H., & Li, X. (2021). MS-ANet: deep learning for automated multi-label thoracic disease detection and classification. *PeerJ Computer Science*, 7, e541.
- [21] Shelke, A., Inamdar, M., Shah, V., Tiwari, A., Hussain, A., Chafekar, T., & Mehendale, N. (2021). Chest X-ray classification using deep learning for automated COVID-19 screening. *SN computer science*, 2(4), 300.
- [22] Gupta, A., Gupta, S., & Katarya, R. (2021). InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray. *Applied Soft Computing*, 99, 106859.

- [23] Ravi, V., Acharya, V., & Alazab, M. (2023). A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. *Cluster Computing*, 26(2), 1181-1203.
- [24] Ho, T. K. K., & Gwak, J. (2020). Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities. *IEEE access*, 8, 160749-160761.
- [25] Ge, Z., Mahapatra, D., Chang, X., Chen, Z., Chi, L., & Lu, H. (2020). Improving multi-label chest X-ray disease diagnosis by exploiting disease and health labels dependencies. *Multimedia Tools and Applications*, 79, 14889-14902.
- [26] Chandra, T. B., Singh, B. K., & Jain, D. (2022). Disease localization and severity assessment in chest x-ray images using multi-stage superpixels classification. *Computer Methods and Programs in Biomedicine*, 222, 106947.
- [27] Visuña, L., Yang, D., Garcia-Blas, J., & Carretero, J. (2022). Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning. *BMC Medical Imaging*, 22(1), 178.
- [28] Joshi, R. C., Yadav, S., Pathak, V. K., Malhotra, H. S., Khokhar, H. V. S., Parihar, A., ... & Dutta, M. K. (2021). A deep learning-based COVID-19 automatic diagnostic framework using chest X-ray images. *Biocybernetics and Biomedical Engineering*, 41(1), 239-254.
- [29] Ravi, V., Narasimhan, H., Chakraborty, C., & Pham, T. D. (2022). Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images. *Multimedia systems*, 28(4), 1401-1415.
- [30] Elhanashi, A., Saponara, S., & Zheng, Q. (2023). Classification and localization of multi-type abnormalities on chest X-rays images. *IEEE Access*, 11, 83264-83277.
- [31] Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., & Yakoi, P. S. (2024). Pneumonia classification using deep learning from chest X-ray images during COVID-19. *Cognitive computation*, 16(4), 1589-1601.
- [32] Karar, M. E., Hemdan, E. E. D., & Shouman, M. A. (2021). Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. *Complex & Intelligent Systems*, 7(1), 235-247.
- [33] Karaddi, S. H., & Sharma, L. D. (2023). Automated multi-class classification of lung diseases from CXR-images using pre-trained convolutional neural networks. *Expert Systems with Applications*, 211, 118650.
- [34] Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T., & Nguyen, H. Q. (2021). Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing*, 437, 186-194.
- [35] Al-qaness, M. A., Zhu, J., AL-Alimi, D., Dahou, A., Alsamhi, S. H., Abd Elaziz, M., & Ewees, A. A. (2024). Chest x-ray images for lung disease detection using deep learning techniques: a comprehensive survey. *Archives of Computational Methods in Engineering*, 31(6), 3267-3301.
- [36] Nahiduzzaman, M., Goni, M. O. F., Hassan, R., Islam, M. R., Syfullah, M. K., Shahriar, S. M., ... & Kowalski, M. (2023). Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19. *Expert Systems with Applications*, 229, 120528.
- [37] Akter, S., Shamrat, F. J. M., Chakraborty, S., Karim, A., & Azam, S. (2021). COVID-19 detection using deep learning algorithm on chest X-ray images. *Biology*, 10(11), 1174.
- [38] El Asnaoui, K., & Chawki, Y. (2021). Using X-ray images and deep learning for automated detection of coronavirus disease. *Journal of Biomolecular Structure and Dynamics*, 39(10), 3615-3626.
- [39] Nigam, B., Nigam, A., Jain, R., Dodia, S., Arora, N., & Annappa, B. (2021). COVID-19: Automatic detection from X-ray images by utilizing deep learning methods. *Expert Systems with Applications*, 176, 114883.
- [40] Bhosale, Y. H., & Patnaik, K. S. (2023). PulDi-COVID: Chronic obstructive pulmonary (lung) diseases with COVID-19 classification using ensemble deep convolutional neural network from chest X-ray images to minimize severity and mortality rates. *Biomedical Signal Processing and Control*, 81, 104445.
- [41] Shamrat, F. J. M., Azam, S., Karim, A., Islam, R., Tasnim, Z., Ghosh, P., & De Boer, F. (2022). LungNet22: a fine-tuned model for multiclass classification and prediction of lung disease using X-ray images. *Journal of Personalized Medicine*, 12(5), 680.
- [42] Öztürk, Ş., Turah, M. Y., & Çukur, T. (2025). Hydravit: Adaptive multi-branch transformer for multi-label disease classification from chest X-ray images. *Biomedical Signal Processing and Control*, 100, 106959.



- [43] Venkatesh, T., Ramesh, D. (2025). An Enhanced Deep Learning Model to Detect Lung Diseases from Chest-Xrays. In: Kumar, A., Gunjan, V.K., Senatore, S., Hu, YC. (eds) Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1. ICDSMLA 2023. Lecture Notes in Electrical Engineering, vol 1273. Springer, Singapore. [https://doi.org/10.1007/978-981-97-8031-0\\_32](https://doi.org/10.1007/978-981-97-8031-0_32)