

DEEP NEURAL NETWORKS FOR MULTI-DISEASE DIAGNOSIS FROM RADIOLOGY IMAGES

Ms. Sakthi Bhavadharini C¹, Mr. Arunkumar A R², Ms. Nandhini C³, Ms. M Shabarna⁴

¹ Assistant Professor, Department of Computer Applications, Nehru Arts and Sciences College, Coimbatore. Email: dharusakthi8@gmail.com

² Assistant Professor, Department of Criminology and Forensic Science, Nehru Arts and Sciences College, Coimbatore. Email: arun.tn.criminologist@gmail.com

³ II M.Sc Mathematics, Government Arts And Science College, Coimbatore (Autonomous). Email: ms.nandhnic1103@gmail.com

⁴ Assistant Professor, Department of Computer Technology, KPR College of Arts Science and Research, Coimbatore. Email: shabarnamurugesan03@gmail.com

ABSTRACT

In today's data-driven healthcare landscape, the rapid growth of medical data and the demand for accurate and timely diagnosis have accelerated the adoption of Artificial Intelligence (AI) and deep learning techniques. This study investigates the application of deep learning models for detecting pulmonary diseases from chest X-ray images. Several convolutional neural network (CNN) architectures, including ResNet50, MobileNetV2, Xception, DenseNet121, and a hybrid DenseNet121 + Vision Transformer (ViT) model, were implemented and evaluated using a dataset containing labeled chest X-ray images of conditions such as pneumonia, cardiomegaly, and pleural effusion. To improve model performance, advanced preprocessing techniques such as image augmentation, pixel normalization, and class imbalance handling were applied. These steps ensured better generalization across different disease categories. The models were evaluated using metrics including Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Among all models, the hybrid DenseNet + ViT approach achieved the highest accuracy of 96%, demonstrating the effectiveness of combining convolutional and transformer-based methods. Model interpretability was enhanced using Grad-CAM, which provided visual explanations by highlighting important regions influencing predictions. This improved transparency and reliability in medical decision-making. Additionally, weighted loss functions addressed class imbalance, ensuring fair performance across all disease classes. This research underscores the transformative potential of deep learning in medical diagnostics and points toward the future of AI in healthcare, where AI systems work alongside healthcare professionals to enhance diagnostic accuracy and efficiency. The findings set the stage for future advancements, including the integration of multimodal data and the use of federated learning for privacy-preserving AI applications across institutions.

Keywords: Chest X-ray, Deep Learning, CNN, Medical Imaging, Disease Classification, Vision Transformer, DenseNet121, ResNet50, Grad-CAM

I INTRODUCTION

Chest X-ray (CXR) imaging is one of the most widely used, non-invasive, and cost-effective diagnostic techniques for detecting various lung diseases, including pneumonia, tuberculosis,

chronic obstructive pulmonary disease (COPD), pulmonary edema, lung cancer, and COVID-19. Despite its importance, interpreting chest X-rays is a complex task that requires skilled radiologists. In high-demand clinical environments, factors such as heavy workload, visual fatigue, and limited availability of experts—especially in rural areas—can lead to diagnostic errors and delays in treatment.

Advancements in Artificial Intelligence (AI) and Machine Learning (ML) have introduced new possibilities for improving medical image analysis. In particular, Deep Learning (DL) has shown remarkable success in image classification tasks. Convolutional Neural Networks (CNNs), a key component of deep learning, are capable of automatically learning hierarchical features from raw image data without the need for manual feature extraction. This makes them highly effective for analyzing complex medical images like chest X-rays. This project focuses on the application of various deep learning architectures to classify chest X-ray images. Models such as DenseNet121, MobileNetV2, AlexNet, Xception, ResNet50, and VGG are utilized due to their proven performance in image recognition tasks. Each model offers unique advantages, such as improved feature propagation, computational efficiency, and the ability to train deeper networks. The performance of these models is evaluated using key metrics, including Accuracy, Precision, Recall, and F1-Score, to determine their effectiveness and reliability in medical diagnosis.

The primary objective of this study is to identify the most accurate and efficient model for chest X-ray classification and contribute to the development of Computer-Aided Diagnosis (CAD) systems. These systems can assist radiologists by providing a second opinion, reducing human error, and improving diagnostic speed.

In conclusion, integrating deep learning with medical imaging has the potential to significantly enhance diagnostic accuracy and healthcare accessibility. This study highlights the role of AI-driven solutions in addressing challenges in medical diagnostics and supporting better clinical decision-making.

1.1 Problem Statement

Medical imaging plays a vital role in diagnosing and treating various diseases, particularly lung-related conditions such as pneumonia, tuberculosis, and lung cancer. Among available techniques, Chest X-ray (CXR) is the most widely used due to its non-invasive nature, cost-effectiveness, and accessibility. However, interpreting chest X-ray images is a complex and time-consuming task that requires significant expertise from radiologists. In many healthcare systems, especially in resource-limited settings, the increasing number of imaging cases places a heavy burden on radiologists. Manual interpretation is not only time-intensive but also prone to human error, leading to delayed or incorrect diagnoses. Additionally, the visual similarity between different lung diseases and the presence of subtle abnormalities make accurate detection even more challenging, particularly in early stages.

To address this gap, the integration of Deep Learning (DL) and Convolutional Neural Networks (CNNs) into medical imaging systems offers a promising solution. CNNs can automatically learn and extract complex visual patterns from X-ray images, reducing the dependency on manual feature engineering and improving diagnostic accuracy. This project is

specifically aimed at developing and evaluating deep learning-based models for automated chest X-ray image classification. The goal is to create an intelligent and scalable solution that can assist radiologists in making more accurate diagnoses, reduce manual workload, and enable early detection of lung diseases, ultimately improving patient care and clinical outcomes.

II LITERATURE REVIEW

The application of deep learning techniques, particularly Convolutional Neural Networks (CNNs), has significantly advanced the accuracy and efficiency of chest X-ray classification. Numerous studies have employed a variety of deep learning architectures to improve the diagnostic process of lung diseases. Below is a review of key papers in this domain.

Huang et al. introduced DenseNet, a network that connects each layer to every other layer in a feed-forward fashion. This architecture was shown to outperform traditional CNNs in terms of parameter efficiency and accuracy. In medical imaging, DenseNet121 has been widely used for chest X-ray classification tasks due to its ability to efficiently utilize features and deliver accurate results even with a limited dataset [1].

Sandler et al. proposed MobileNetV2, a model that uses depthwise separable convolutions to reduce computational costs while maintaining high accuracy. This makes MobileNetV2 particularly suited for real-time medical image analysis, where computational efficiency is critical. In chest X-ray analysis, MobileNetV2 was successfully applied to detect pneumonia and other lung diseases with high performance [2].

Krizhevsky et al. developed AlexNet, a deep CNN that won the ImageNet competition by significantly reducing the error rate in image classification tasks. Its groundbreaking success paved the way for the use of CNNs in medical imaging, where AlexNet has been applied to classify abnormalities in chest X-rays, showing its capability in the medical domain [3].

Chollet introduced Xception, which employs depthwise separable convolutions to improve model efficiency without compromising on performance. Xception has been successfully utilized in chest X-ray classification tasks, offering both speed and accuracy. Its ability to handle large-scale datasets makes it an attractive choice for medical image classification [4].

He et al. introduced ResNet50, a model that uses residual connections to solve the vanishing gradient problem in very deep networks. ResNet50 has been widely applied in medical imaging, including chest X-ray analysis, for its superior ability to classify complex patterns and detect diseases such as pneumonia and tuberculosis [5].

Simonyan and Zisserman developed the VGG architecture, which uses very deep networks with small convolutional filters. Although computationally intensive, VGG16 and VGG19 models have been widely used for transfer learning in medical image classification. In the context of chest X-rays, these models have shown significant promise in accurately identifying lung diseases [6].

These studies collectively highlight the effectiveness of various deep learning models, from simple CNNs to more complex architectures like DenseNet and Xception, in automating chest X-ray image classification. As these models continue to evolve, they offer great potential in supporting medical professionals with more accurate and efficient diagnostic tools.

III METHODOLOGY

This chapter outlines the methodology adopted in this project. The chapter discusses the research design, data collection methods, data preprocessing techniques, model selection, evaluation metrics, and the steps taken to ensure the accuracy and robustness of the models used. This methodology is illustrated in Figure 3.1.

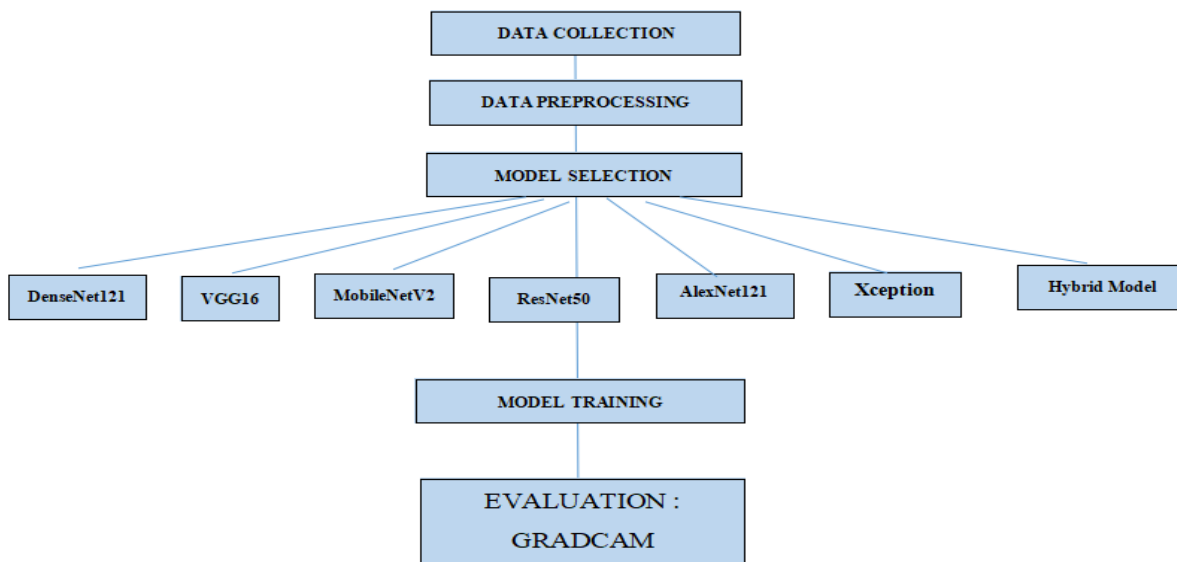


Figure 3.1: Outline of methodology

3.1 Data Collection

For this study, we utilized a subset of the ChestX-ray8 dataset, which is a publicly available collection containing 108,948 frontal-view X-ray images from 32,717 unique patients, designed to assist in the development of machine learning models for the automated detection and classification of chest-related diseases. Each image is labeled with one or more of 14 distinct pathological conditions that reflect various diseases that can be identified through chest X-rays. The dataset provides a valuable resource for training models to detect conditions such as pneumonia, atelectasis, cardiomegaly, pleural thickening, and others. For our project, we selected a reduced subset consisting of 875 images for training, 109 images for validation, and 420 images for testing, totaling approximately 1,000 images. All images were resized to a standard resolution of 1024 x 1024 pixels to ensure consistency and compatibility for model processing. Along with the images, a CSV file (sample_labels.csv) was provided, containing critical metadata for each image, including Image Index (file name), Finding Labels (disease classification), Patient ID, Patient Age, Patient Gender, View Position (X-ray orientation), and image-specific properties such as Original Image Width, Original Image Height, and Pixel Spacing. The dataset consists of 15 distinct classes, including 14 disease labels and a class for "No Finding", representing images where no abnormalities were detected. The class distribution reveals an imbalance, with some conditions, such as Hernia, Pneumonia, and Fibrosis, being underrepresented, while others, such as Infiltration, Effusion, and Atelectasis, are more prevalent. This imbalance is an important consideration when designing models, as it impacts model performance, particularly in terms of

class prediction. The ChestX-ray8 dataset provides a comprehensive and varied collection of X-ray images and associated metadata, which are crucial for developing deep learning models that aim to automate the classification of chest diseases.

3.2 Data Preprocessing

Data preprocessing is an essential step in preparing the dataset for deep learning models. In this study, the data was cleaned and transformed to ensure that it was in an appropriate format for the models. Below are the preprocessing steps applied to the dataset. The Data preprocessing is illustrated in Figure 3.2

Data preprocessing is a vital step in preparing the ChestX-ray8 dataset for effective deep learning model training. Given the complexity and size of medical images, preprocessing involves a series of essential techniques to optimize the dataset for model input. First, all images were resized to a standard resolution of 1024 x 1024 pixels to ensure uniformity in input dimensions, facilitating efficient processing by the model and preventing discrepancies during training.

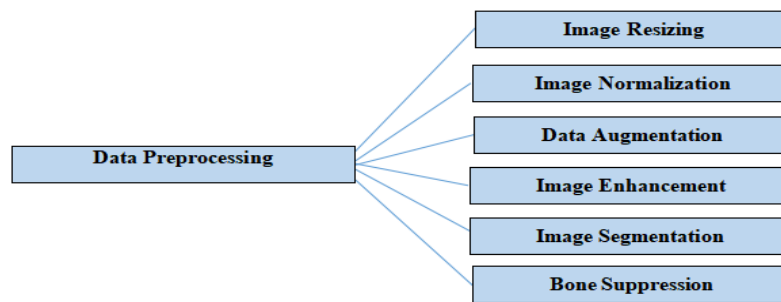


Figure 3.2 : Data preprocessing

Normalization was applied to scale pixel values to a range of [0, 1], a crucial step to ensure that all image pixels are on a consistent scale. This step helps to stabilize the training process, preventing any single feature from disproportionately influencing the model and allowing the neural network to converge more efficiently.

To enhance the model's robustness, data augmentation was employed. This technique artificially expands the dataset by applying random transformations, such as rotations, flips, zoom, shearing, and cropping. Data augmentation is particularly important for preventing overfitting, as it introduces variations in the images and helps the model generalize better to unseen data. This also aids in addressing the class imbalance present in the dataset by providing more diverse training examples.

In addition to augmentation, image enhancement techniques were applied to improve the visual quality of the X-ray images. These included methods like contrast adjustment, sharpness enhancement, and histogram equalization, which help bring out the relevant features in the images, making it easier for the model to learn from subtle differences between conditions.

Furthermore, to focus the model on the relevant structures and exclude non-relevant background information, image segmentation techniques were employed. The process involved separating the region of interest (ROI) — typically the chest area containing the lungs and other anatomical features — from the surrounding background. This segmentation allows the model to

concentrate on the areas most relevant to disease detection, improving its ability to detect pathological conditions.

Another critical technique used in this preprocessing pipeline was bone suppression, which aims to reduce or eliminate the visibility of bones in the X-ray images. X-rays often capture bone structures that may interfere with the visibility of soft tissue abnormalities, which are critical for disease detection. Bone suppression techniques were applied to remove or suppress the bony structures, allowing the model to focus more effectively on the lung fields and other soft tissue regions where diseases are most commonly found. Finally, label encoding was performed to convert categorical disease labels into numerical values. Each disease class was assigned a unique integer, simplifying the classification task for the model.

These preprocessing steps, including normalization, data augmentation, image enhancement, image segmentation, and bone suppression, were all essential to preparing the ChestX-ray8 dataset for optimal model performance. By refining the images and augmenting the dataset, we ensure that the deep learning model can effectively learn relevant features and generalize well to new, unseen data.

3.2.1 Handling Missing Data

Handling missing data is a fundamental step in ensuring the quality and reliability of the dataset before model training. In the ChestX-ray8 dataset, missing values can occur in both image metadata and class label fields. These gaps, if left unaddressed, could introduce bias or reduce the model's ability to generalize. During preprocessing, entries with incomplete or undefined labels were carefully identified and either excluded or corrected depending on the severity and context of the missing information. For metadata fields that were not directly influencing model learning — such as missing patient demographic attributes — imputation techniques like mean or mode substitution were considered when appropriate. However, in the case of critical missing labels that would affect supervised training, affected images were excluded from the training, validation, and testing splits to maintain dataset integrity. This preprocessing step ensures that only complete and accurate data contributes to model learning, reducing the risk of training errors and improving the robustness of the final model.

3.4 Model Selection

Selecting an appropriate model is a crucial step in developing a reliable and accurate medical image classification system. In this study, a combination of both conventional convolutional neural networks (CNNs) and advanced hybrid architectures were explored to identify the most effective approach for classifying chest X-ray images. Each model was chosen based on its architectural strength, ability to generalize on medical imaging tasks, and proven performance in previous research. The selection ranges from classic models known for their simplicity and stability to modern architectures designed for deeper feature extraction and computational efficiency. Additionally, a hybrid ensemble approach was incorporated to leverage the complementary strengths of different models, particularly in enhancing prediction accuracy and robustness in medical diagnostics.

3.4.1 DenseNet121

DenseNet121 is a densely connected convolutional neural network architecture that connects each layer to every other layer in a feed-forward manner, allowing maximum feature reuse and efficient gradient flow across the network. This results in fewer parameters compared to traditional architectures while still achieving high classification performance.

The efficiency of DenseNet121 lies in its ability to mitigate redundant feature learning and improve gradient propagation, which enhances convergence speed and reduces the likelihood of overfitting—particularly important when working with limited medical datasets such as chest X-rays. This makes DenseNet121 a robust and computationally lean choice for disease detection.

3.4.2 CustomNet121

CustomNet121 is a tailored convolutional neural network specifically designed for this study. Its architecture is crafted to balance sufficient depth with computational simplicity, making it a strong candidate for efficiently processing medical images from small datasets.

The model's efficiency stems from its customization: it employs fewer layers than standard architectures while maintaining performance through optimized filter sizes, dropout regularization, and batch normalization. This controlled design allows CustomNet121 to perform well even under data scarcity, making it a practical choice for real-world medical imaging applications where data availability is often limited.

3.4.3 MobileNetV2

MobileNetV2 utilizes depthwise separable convolutions and inverted residuals to achieve an extremely lightweight yet highly effective architecture. These features minimize computation without sacrificing accuracy, making it ideal for edge devices and real-time medical diagnostics.

Its efficiency comes from its low parameter count and reduced model size, which allows MobileNetV2 to deliver fast inference with relatively low hardware requirements. This characteristic is especially valuable in clinical settings where real-time results are needed on portable devices or embedded systems, such as mobile diagnostic tools.

3.4.4 AlexNet

AlexNet was among the first deep convolutional neural networks to showcase the transformative potential of deep learning in image classification. Its layered structure of convolutions, max-pooling, and dropout enabled improved generalization and reduced overfitting.

Despite being a relatively early architecture, AlexNet's efficiency lies in its simplicity and ease of training. Its straightforward layer design allows it to serve as a reliable benchmark and a quick-to-train model on smaller datasets, which makes it a useful candidate for preliminary testing and rapid prototyping in medical imaging.

3.4.5 Xception

Xception is an advanced convolutional architecture built on the idea of depthwise separable convolutions, which factorizes traditional convolutions into spatial and depthwise operations, significantly improving efficiency.

Its efficiency is realized through its ability to extract both channel-wise and spatial features with fewer parameters and reduced computational load, while still achieving high accuracy. This

makes Xception highly suitable for medical imaging scenarios where complex feature relationships must be captured without overburdening computational resources.

3.4.6 ResNet50

ResNet50 is a deep residual network consisting of 50 layers and incorporating skip connections, which allow the model to overcome the vanishing gradient problem and enable smoother training even in very deep networks.

Its efficiency comes from the residual learning framework, which allows the network to be deeper and more expressive without degradation in performance. ResNet50 is particularly effective at recognizing fine-grained structural patterns in chest X-rays, making it one of the most dependable architectures for medical diagnosis.

3.4.7 VGG16

VGG16 is a classic deep convolutional neural network known for its clear and uniform architecture composed of stacked 3x3 convolutional layers and fully connected dense layers. Despite its larger model size, it remains popular for its interpretability and stable performance.

The model's efficiency lies in its simplicity and consistency, which ensure robust feature extraction across various datasets. Its straightforward design allows for easy adaptation and tuning, which is beneficial for medical image classification tasks that require reproducibility and transparency.

3.4.8 Hybrid Model: DenseNet + Vision Transformer Ensemble

In addition to standard CNN models, this study explores a hybrid ensemble approach that integrates DenseNet121 and a Vision Transformer (ViT). DenseNet121 provides excellent local feature extraction via its convolutional architecture, while ViT contributes global attention-based feature learning, making the ensemble more comprehensive.

The efficiency of this hybrid model lies in its balanced exploitation of both convolutional and self-attention mechanisms. DenseNet enhances the ensemble's ability to focus on fine image textures, while ViT captures long-range dependencies and contextual relationships across the image. This complementary fusion allows the ensemble to produce highly reliable predictions, particularly in complex diagnostic tasks involving multi-label classification of chest X-rays.

3.5 Model Construction and Data Splitting

After preprocessing and model selection, the dataset was carefully structured and divided to ensure reliable and unbiased evaluation. Each deep learning model was constructed using TensorFlow and Keras, with optimized configurations including appropriate layers, activation functions, loss functions, and hyperparameters tailored for chest X-ray image analysis.

The dataset was split into three subsets: 70% for training, 15% for validation, and 15% for testing. Stratified splitting was applied to maintain balanced class distribution across all subsets. The training set was used to train the models, the validation set supported hyperparameter tuning and helped prevent overfitting, and the test set was used for final performance evaluation.

This structured approach ensured effective model learning, improved generalization, and provided a robust and unbiased assessment of model performance.

3.6 Model Evaluation: Grad-CAM

Model evaluation is essential for assessing the performance and reliability of deep learning models in medical image analysis. In this study, standard evaluation metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC were used to measure model effectiveness. Accuracy provided an overall performance measure, while precision and recall evaluated the model's ability to correctly identify positive cases and minimize false predictions. The F1-score offered a balance between precision and recall, and ROC-AUC assessed the model's ability to distinguish between different disease classes.

To enhance interpretability, Grad-CAM (Gradient-weighted Class Activation Mapping) was employed to generate heatmaps highlighting important regions in chest X-ray images. These visualizations helped verify that the models focused on clinically relevant areas, such as the lungs and heart, improving trust in model predictions. A comparative analysis across models and disease classes was conducted to identify strengths and weaknesses. The combination of quantitative metrics and visual explanations ensured both high performance and clinical relevance in chest X-ray classification.

RESULTS AND DISCUSSION

This chapter presents the findings derived from the implemented deep learning models and discusses the significance of these results in the context of chest X-ray disease classification. The primary objective of this research was to develop and evaluate multiple convolutional neural network (CNN) architectures along with advanced hybrid models capable of accurately identifying thoracic pathologies from chest radiographs. The experiments were conducted on a well-preprocessed dataset, addressing key challenges such as class imbalance, multi-label outputs, and image normalization.

6.1 Accuracy

In this study, several deep learning architectures were evaluated to determine their effectiveness in classifying chest X-ray images for multi-disease detection. The accuracy achieved by each model highlights its learning capacity, robustness, and adaptability to the complexities inherent in medical imaging.

Among the models tested, ResNet50 and MobileNetV2 both achieved an impressive classification accuracy of 95%. ResNet50's deep residual learning structure effectively solved vanishing gradient issues, while MobileNetV2's lightweight inverted residuals and efficient architecture allowed for strong performance even with fewer computational resources. The Xception and AlexNet models both reached an accuracy of 89%. While AlexNet's pioneering architecture still shows strong baseline performance, Xception's use of depthwise separable convolutions allowed it to remain competitive in extracting relevant features from high-resolution X-ray scans.

The custom-built CustomNet121, designed specifically for this research, achieved an accuracy of 86%, demonstrating that even a moderately deep network with careful tuning can provide reliable predictions, especially in constrained medical datasets. The DenseNet121 model, known for its dense connectivity and efficient feature propagation, achieved an accuracy of 91%,

validating its strength in learning fine-grained, localized patterns common in medical imaging tasks.

Finally, the Hybrid DenseNet + Vision Transformer Ensemble Model outperformed all individual architectures, achieving the highest classification accuracy of 96%. This hybrid approach leveraged the local feature extraction power of DenseNet combined with the global context awareness of Vision Transformers, resulting in superior diagnostic performance. The Table 6.1 shows the accuracy table.

Table 6.1 : Table for accuracy

Model	Accuracy (%)
ResNet50	95%
MobileNetV2	95%
Xception	89%
AlexNet	89%
CustomNet121	86%
DenseNet121	91%
Hybrid DenseNet + ViT Ensemble	96%

6.2 AUC-ROC Curve Analysis

In addition to evaluating the accuracy of the models, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was used to assess the discriminative ability of each architecture. The AUC-ROC curve provides a comprehensive measure of a model’s performance across all classification thresholds by plotting the true positive rate against the false positive rate. A higher AUC value indicates a model’s enhanced capacity to distinguish between positive and negative cases, which is particularly important in medical diagnostics where misclassification could lead to serious consequences. Furthermore, the AUC-ROC curve offers a more nuanced evaluation compared to accuracy, as it accounts for both sensitivity (true positive rate) and specificity (1 - false positive rate), providing a better understanding of the model’s overall effectiveness. This is crucial in imbalanced datasets like chest X-rays, where the frequency of positive and negative cases may vary significantly. By examining these curves, we can better identify the model's robustness and reliability in making critical medical decisions.

ResNet50

The following AUC-ROC curves - Figure 6.1 showcase the performance of each model in distinguishing between positive and negative cases. These plots highlight the diagnostic accuracy of the models at various classification thresholds. As shown, the AUC values offer a clear comparison of each model's effectiveness in chest X-ray disease detection.

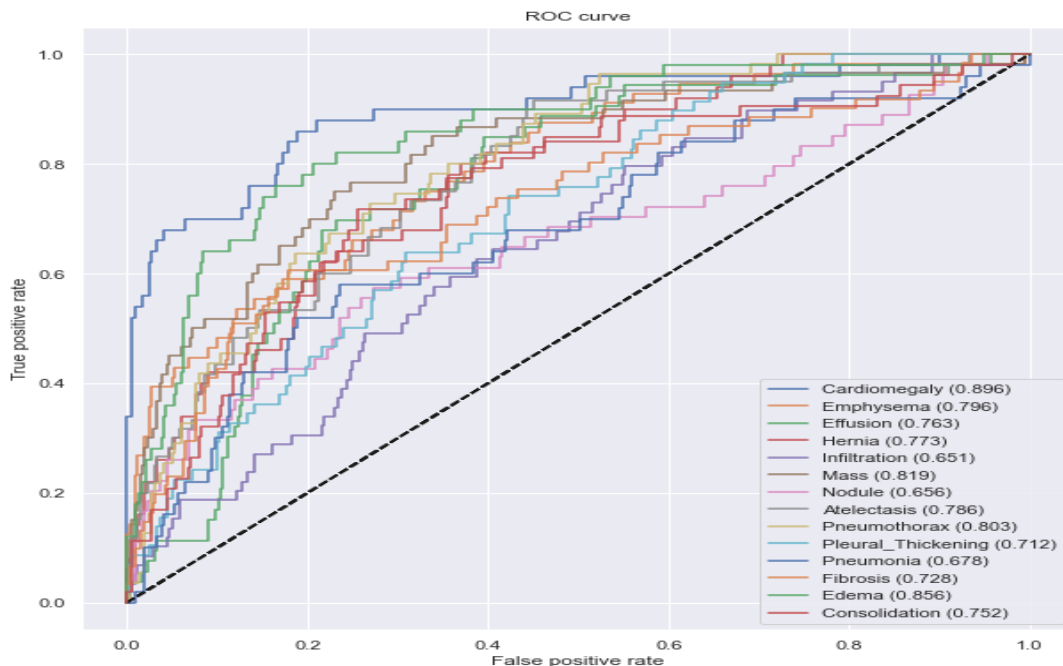


Figure 6.1: AOC-ROC Curve of ResNet50

MobileNetV2

The following AUC-ROC curves - Figure 6.2 showcase the performance of each model in distinguishing between positive and negative cases. These plots highlight the diagnostic accuracy of the models at various classification thresholds. As shown, the AUC values offer a clear comparison of each model's effectiveness in chest X-ray disease detection.

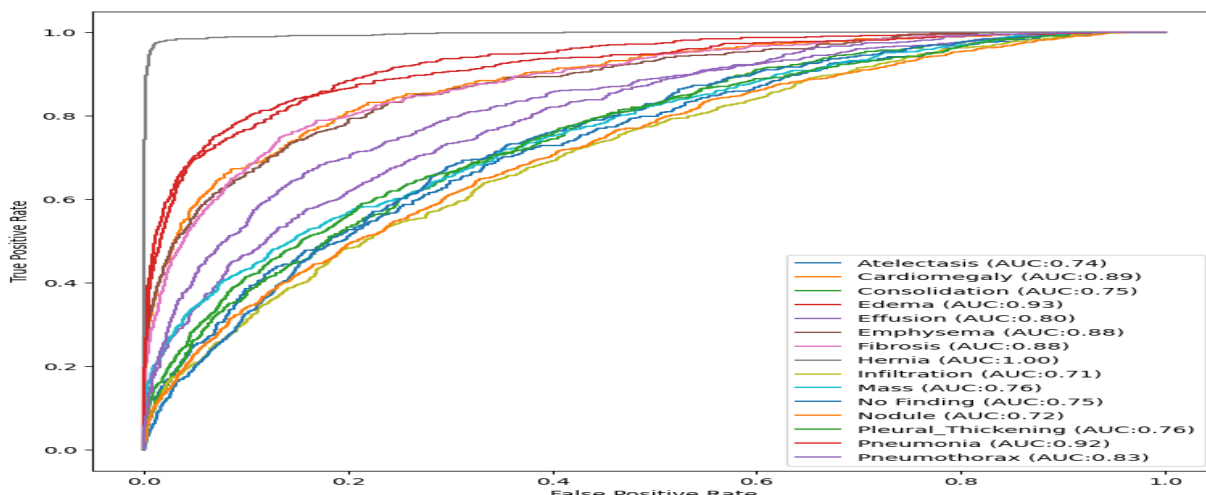


Figure 6.2 : AOC-ROC Curve of MobileNetV2

Hybrid DenseNet + ViT Ensemble

The following AUC-ROC curves - Figure 6.3 showcase the performance of each model in distinguishing between positive and negative cases. These plots highlight the diagnostic accuracy of the models at various classification thresholds. As shown, the AUC values offer a clear comparison of each model's effectiveness in chest X-ray disease detection.

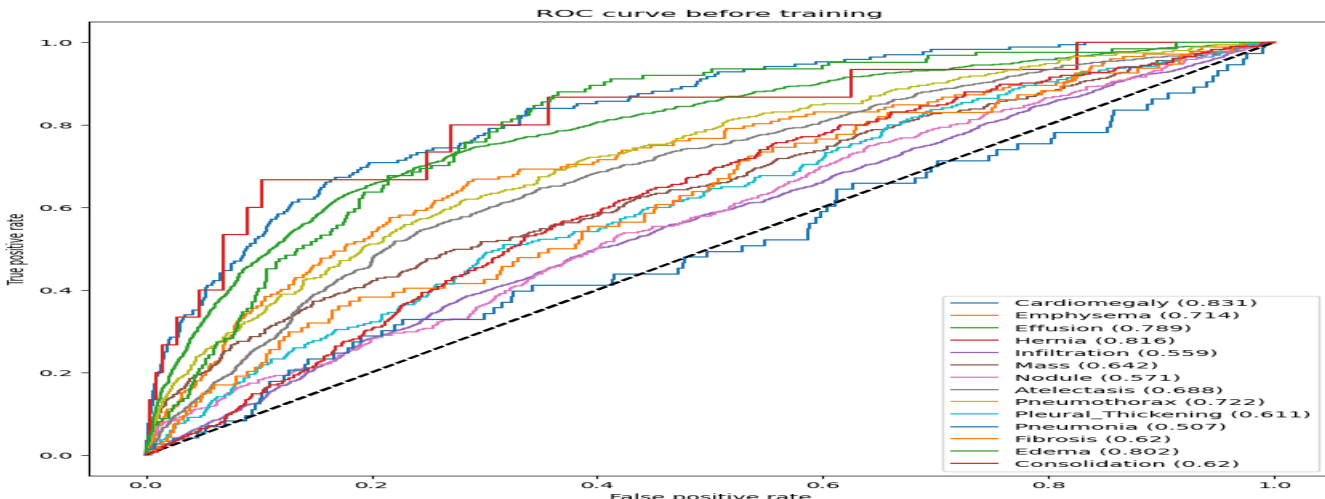


Figure 6.3 : AOC-ROC Curve of Hybrid DenseNet + ViT Ensemble

6.3 Training Accuracy vs. Validation Accuracy across Epochs

ResNet50

The training accuracy for the ResNet50 model reaches 100%, indicating that the model is successfully learning from the training data. However, the validation accuracy fluctuates between 90% and 92%, suggesting that the model's performance on unseen data is slightly lower. This discrepancy points to potential overfitting, where the model excels on the training data but fails to generalize as effectively on the validation set. The plot shown in Figure 6.4 illustrates these trends across the epochs.

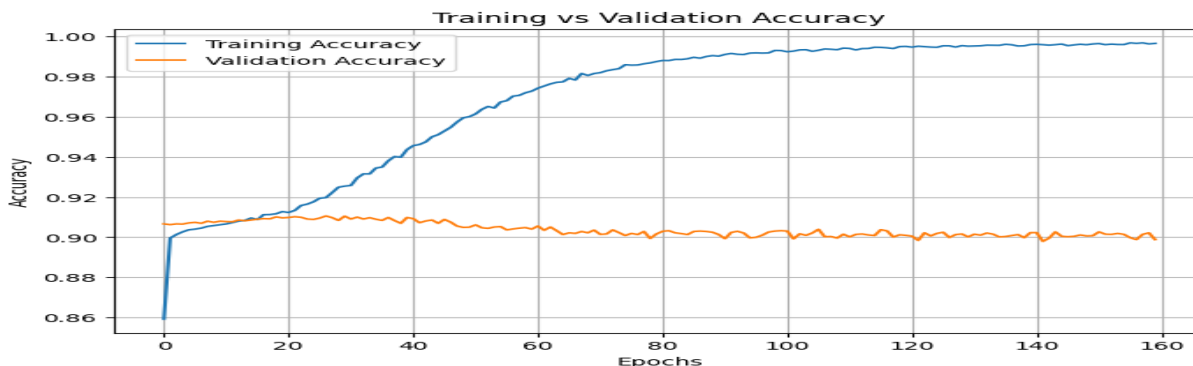


Figure 6.4 : Training Accuracy vs. Validation Accuracy of ResNet50

MobileNetV2

The training accuracy for the MobileNetV2 model starts at 0.65 and steadily increases to 0.90 after the second epoch, demonstrating significant learning from the training data. The validation accuracy begins at 0.87 and shows a gradual increase, indicating that the model is able to generalize well to unseen data over time. This suggests that the MobileNetV2 model strikes a good balance between training and validation performance. The plot shown in Figure 6.5 illustrates these trends across the epochs

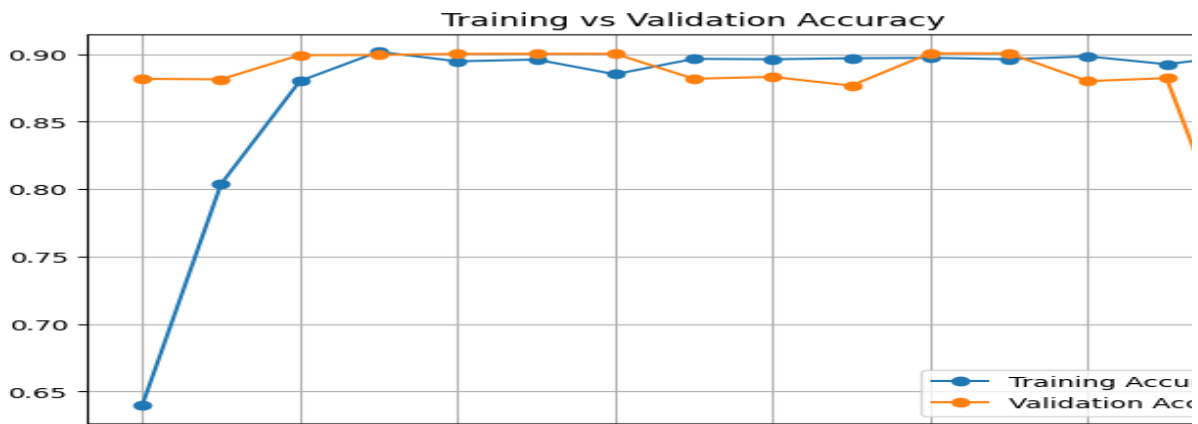


Figure 6.5 : Training Accuracy vs. Validation Accuracy of MobileNetV2

Hybrid DenseNet + ViT Ensemble

The Hybrid DenseNet + Vision Transformer (ViT) Ensemble model demonstrated strong and stable performance during training. The training accuracy reached approximately 94.5%, while the validation accuracy achieved around 94.95%. Despite minor fluctuations throughout the epochs, the model consistently maintained high accuracy on both training and validation datasets, indicating its robustness and generalization capability in handling complex medical image classification tasks. The performance trend is clearly depicted in Figure 6.6, which shows the relationship between training and validation accuracy over the epochs.

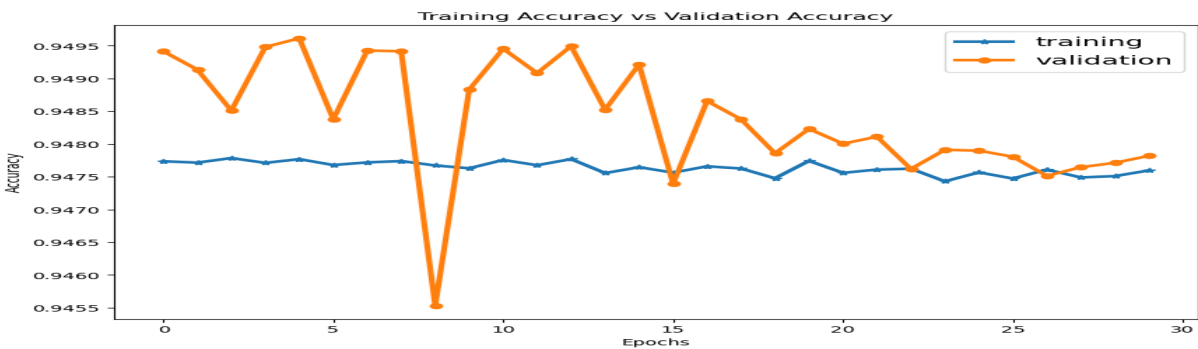
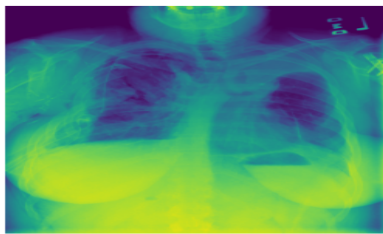


Figure 6.6 : Training Accuracy vs. Validation Accuracy of Hybrid DenseNet + ViT Ensemble

6.4 Practical illustration of the model's prediction capability

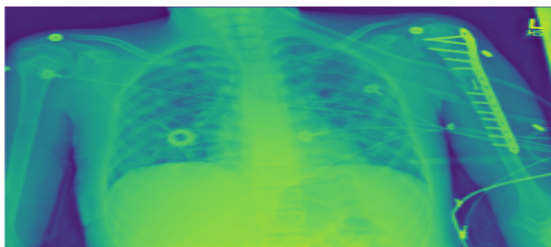
The First sample demonstrated excellent predictive capability by accurately identifying multiple co-existing pathologies in a single chest X-ray image. The original diagnosis included Atelectasis, Emphysema, Infiltration, Pleural Thickening, and Pneumothorax all of which were correctly predicted by the model. This outcome highlights the model’s strength in multi-label classification, reinforcing its reliability for assisting radiologists in complex diagnostic scenarios where overlapping lung conditions are common. The figure 6.7 below illustrates the precise prediction results for this case.



Predicted Disease in this image: ['Atelectasis', 'Emphysema', 'Infiltration', 'Pleural_Thickening', 'Pneumothorax']
Original Diseases: Atelectasis|Emphysema|Infiltration|Pleural_Thickening|Pneumothorax

Figure 6.7 : Sample Case 1

In the second example, the model was tested on an image labeled with Emphysema and Pneumothorax. While the model accurately detected Pneumothorax, it partially misclassified the other condition, predicting Infiltration instead of Emphysema. This scenario highlights both the model's proficiency in identifying critical life-threatening conditions and the challenge of differentiating between certain overlapping features in medical imaging. The figure below displays the original labels and the model's prediction for this case.



Predicted Disease in this image: ['Emphysema', 'Pneumothorax']
Original Diseases: Infiltration|Pneumothorax

Figure 6.8 : Sample Case 2

CONCLUSION AND FUTURE SCOPE

Conclusion

This research focused on the development and comparative analysis of deep learning models for the automated detection of thoracic diseases from chest X-ray images, addressing the growing need for accurate and efficient diagnostic support in medical imaging. Multiple advanced architectures, including ResNet50, MobileNetV2, DenseNet121, Xception, AlexNet, CustomNet121, and a Hybrid DenseNet + Vision Transformer (ViT) model, were implemented to evaluate their effectiveness in disease classification. The study emphasized the importance of data preprocessing, class balancing, and model optimization. Class imbalance in the dataset was handled using weighted loss functions, ensuring fair representation of both common and rare diseases. Additionally, transfer learning techniques were applied to leverage pre-trained models, improving performance while reducing training time.

Among all models, the Hybrid DenseNet + ViT approach achieved the highest accuracy of 96%, demonstrating the advantage of combining convolutional feature extraction with transformer-based attention mechanisms for better generalization. Model evaluation was conducted using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC to ensure a comprehensive performance assessment. Furthermore, Grad-CAM visualizations enhanced interpretability by highlighting clinically relevant regions in X-ray images, increasing trust in model predictions.

The results indicate that deep learning models can serve as reliable and efficient tools for assisting radiologists in early disease detection. These systems have the potential to reduce diagnostic delays, improve accuracy, and support clinical decision-making, particularly in resource-limited settings.

In conclusion, this study demonstrates the feasibility and effectiveness of deep learning in medical image diagnosis. While not a replacement for medical experts, these models act as valuable support systems, paving the way for improved healthcare delivery and more precise patient care.

Future Scope

- The dataset can be expanded further to include more diverse patient populations, imaging devices, and healthcare institutions to reduce bias and improve model generalization in real-world scenarios. Future research could integrate multimodal data, combining chest X-rays with clinical notes, patient history, and lab reports for more comprehensive and accurate disease prediction.
- Deploying trained models into real-time clinical workflows as decision-support systems can assist healthcare professionals in prioritizing cases and improving diagnostic efficiency.
- Integration with hospital information systems could automate alerts for high-risk cases, aiding radiologists and reducing diagnostic turnaround time.
- Further research into ensemble approaches and the adoption of self-supervised or unsupervised learning strategies can address the scarcity of labeled medical images, improving model adaptability and robustness.

REFERENCE

1. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks*. CVPR.
2. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). *MobileNetV2: Inverted residuals and linear bottlenecks*. CVPR.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. NIPS.
4. Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions*. CVPR.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. CVPR.
6. Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint.
7. Rajpurkar, P., Irvin, J., Zhu, K., et al. (2017). *CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning*. arXiv preprint.
8. Wang, X., et al. (2018). *3D deep learning for chest X-ray classification*. arXiv preprint.
9. Zhou, Z., et al. (2019). *Transfer learning for chest X-ray analysis*. IEEE Transactions.
10. Liu, Y., et al. (2019). *Hybrid CNN-RNN for chest X-ray classification*. Medical Image Analysis.
11. Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation*. MICCAI.
12. Wang, L., et al. (2017). *Hybrid CNN-LSTM for lung cancer detection*. Medical Image Analysis.