# Deep Learning- Final Course Project

Lital Binyamin (ID. 315637231) Yuval Saadaty (ID. 205956634)

Submitted as final project report for the DL course, BIU, 2024

## 1 Introduction

This project tackles the urgent global health issue of diagnosing pneumonia by applying deep learning to analyze chest X-ray (CXR) images. Pneumonia is a major cause of death globally, particularly affecting children under five, making its accurate and prompt diagnosis crucial for enhancing patient outcomes. Traditional diagnostic methods depend heavily on radiologists' expertise and are challenged by the intricate nature of CXR interpretations and the overwhelming number of cases. Utilizing a comprehensive dataset of CXR images, our study develops deep learning models designed to accurately distinguish between healthy individuals and those with pneumonia, further categorizing cases into bacterial and viral pneumonia.

Additionally, the project explores the use of embedding vectors for classifying new images, applies anomaly detection to identify cases of pneumonia in images previously deemed healthy, and incorporates model explainability technique to demystify the decision-making process of the neural networks. This advancement holds the potential to improve healthcare delivery and accessibility, ultimately contributing to better patient care and health outcomes.

### 1.1 Related Works

**Pneumonia Diagnose**

Distinguishing between normal lung conditions and pneumonia in chest X-ray imaging hinges on detailed visual analysis, with particular emphasis on the lung fields and their peripheries. In cases of normal lung health, the focus is predominantly on the lung fields themselves, which should appear clear and unobstructed, allowing the detailed structures such as the ribs and the outline of the heart to be visible. This clarity indicates that the lungs are filled with air and free of fluid or infection. In contrast, diagnosing pneumonia involves a more extensive examination that includes the areas around the lungs. Pneumonia can cause the lung fields to show areas of opacity, signaling infection or fluid buildup. However, the disease can also impact the spaces around the lungs, potentially leading to complications such as pleural effusion, where fluid accumulates in the pleural cavity surrounding the lungs. This condition can further obscure the

lung fields and complicate the diagnosis, making it essential to assess both the lungs and the adjacent areas carefully.

A study that delves into these diagnostic distinctions is "Differentiation of Bacterial and Viral Pneumonia in Children" by Korppi et al.[2]. This research highlights the importance of comprehensive imaging analysis to accurately identify pneumonia and its underlying causes, emphasizing the need for careful examination beyond the lung fields to include the surrounding regions for signs of complications or associated conditions.

**Deep Learning for X-ray**

Deep learning has significantly advanced the field of medical imaging, particularly in analyzing X-ray images for various lung conditions. By automating the detection and classification of lung abnormalities, deep learning models have transformed diagnostic practices, offering more accurate, faster, and consistent analyses compared to traditional methods. These technologies have been applied to a broad range of pulmonary diseases, including pneumonia, tuberculosis, lung cancer, and fibrosis, enhancing the ability to diagnose, monitor, and plan treatment for these conditions.

A seminal work in this area is "Deep learning-based algorithm for detecting lung cancer on chest radiographs" by Ardila et al.[1], which demonstrates the potential of deep learning models in identifying lung cancer from chest X-rays with a level of accuracy comparable to that of radiologists. This study is part of a growing body of research that leverages deep learning to improve the early detection of lung cancer, potentially increasing survival rates through timely intervention.

In the realm of infectious diseases, deep learning models have also made significant contributions. The outbreak of COVID-19 presented an unprecedented challenge to global healthcare systems, prompting the rapid development of AI tools to aid in the fight against the pandemic. "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images" by Wang et al.[5] is a notable example. This study introduced a specialized deep learning model designed to detect COVID-19 from chest X-rays, providing a crucial tool for screening and early detection amidst widespread testing constraints.

Additionally, the use of deep learning for diagnosing tuberculosis from chest X-rays, as explored in "Automated detection of tuberculosis in chest radiographs using a convolutional neural network" by Lakhani et al.[3], showcases the broad applicability of these models across various lung diseases. This work underscores the potential of deep learning to support global health initiatives, particularly in regions where access to radiologists and diagnostic tools is limited.

**Deep Learning for X-ray Pneumonia Classifications**

Specifically, in the realm of diagnosing pneumonia from chest X-rays, deep learning models have demonstrated remarkable capabilities. These models are trained to scrutinize X-ray images for the telltale signs of pneumonia, automating the detection process with high levels of accuracy. This not only aids radiologists by reducing their workload but also ensures consistency in diagnosis, potentially catching cases that might be challenging to identify through man-

ual review alone. A pivotal study in this field is "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," by Rajpurkar et al.[4]. This research showcased a deep learning model that achieved radiologist-level accuracy in identifying pneumonia from chest X-rays, underscoring the technology's potential to augment traditional diagnostic methods.

# 2 Solution

## 2.1 General approach

Our primary approach to solving the problem of pneumonia classification from chest X-ray (CXR) images leverages the power of transfer learning from a pretrained ImageNet model, specifically ResNet-18. We chose ResNet-18 for its balance between depth and complexity, aiming to reduce the risk of overfitting while maintaining sufficient representational capacity for our tasks. Transfer learning, by either fine-tuning the entire network, offers a robust foundation due to ResNet-18's proven effectiveness on diverse image recognition tasks. This strategy accelerates training and improves performance even with limited medical imaging data.

Initially, we will apply this method to binary classification (healthy vs. pneumonia). Once the binary model is trained, we plan to utilize it as a pretrained model for a more challenging three-way classification task (healthy, bacterial pneumonia, and viral pneumonia). This stepwise approach allows us to tackle increasingly complex classification problems, leveraging the foundational knowledge learned from the binary task to enhance our model's performance on the three-class problem.

Further exploration into the latent space of our models will be conducted using t-SNE and k-NN (k-Nearest Neighbors). By visualizing the high-dimensional feature representations in a lower-dimensional space with t-SNE and applying k-NN for classification, we aim to adopt a more classical approach to understand the downstream task's nuances. This method provides insight into how well the model's learned representations cluster and differentiate among the classes in a visually interpretable manner.

Anomaly detection for identifying pneumonia cases in ostensibly healthy X-ray images will be approached using Autoencoder. The choice of AE is motivated by the ability to learn rich, hierarchical representations of data in an unsupervised manner. In an unsupervised manner, AE is trained on a dataset consisting primarily of normal cases (healthy X-rays in this context). This training enables the AE to learn a rich, hierarchical representation of what constitutes a "normal" chest X-ray. For anomaly detection, the hypothesis is that when an anomalous X-ray (e.g., one showing signs of pneumonia) is input into the AE, the reconstruction error will be significantly higher compared to normal X-rays.

Lastly, we plan to apply the occlusion sensitivity method for model explainability. This technique involves systematically obscuring different parts of the

input image and observing the impact on the model's output. By identifying regions most critical for the model's pneumonia classification decisions, we can gain insights into its reasoning process. This analysis is particularly valuable for aligning the model's focus areas with those used by radiologists in clinical settings, thereby validating that our model's diagnostic criteria are consistent with professional medical evaluations and the findings highlighted in the literature review.

## 2.2 Design

Our deep learning project focused on pneumonia classification from chest X-ray images involved several critical stages: data preprocessing, model architecture adaptation, training, and evaluation.

**Data Handling** Upon analyzing the dataset, a significant train imbalance was observed: 3,875 pneumonia cases versus 1,341 normal cases. Such disparity necessitated an approach that could effectively manage class imbalance to ensure the model did not bias towards the majority class.

**Architecture and Model Adaptation** The core of our classification model was built on the ResNet-18 architecture, chosen for its efficiency and ability to capture deep features without the complications of vanishing gradients. In our study, we explored multiple popular architectures including VGG16, AlexNet, and ResNet18 for pneumonia classification. Through rigorous experimentation and evaluation, we found that ResNet18 consistently outperformed VGG16 and AlexNet, achieving the highest accuracy and demonstrating superior generalization capabilities on our chest X-ray dataset. For the binary classification task, we modified ResNet-18's final layer to output two classes (normal and pneumonia), aligning with our objective. For the multi-class classification, we expanded this to three outputs, leveraging the binary model's weights as a foundation to enhance learning efficiency.

For anomaly detection, we explored the data's latent space using a Autoencoder (AE). The AE architecture was designed with convolutional layers for encoding and transpose convolutional layers for decoding. To determine anomalies, we utilized a thresholding approach based on the Receiver Operating Characteristic (ROC) curve. Specifically, we identified a threshold by comparing False Positive Rate (FPR) and False Negative Rate (FNR) values from the ROC curve, aiming for a small discrepancy (less than 5e-3) between these rates to pinpoint anomalous instances effectively.

**Hyperparameters** For classification tasks: Optimizer: SGD with a learning rate of 0.001, momentum of 0.9, and weight decay of 0.01. Loss Function: Focal Loss (with alpha=0.25 and gamma=2), selected to mitigate the impact of the class imbalance by reducing the weight of easy-to-classify examples and focusing on harder cases. Learning Rate Scheduler: StepLR, decreasing the learning rate by a factor of 0.1 every 10 epochs to fine-tune model performance.

**Loss Function** Focal Loss was pivotal in addressing the dataset's imbalance, enabling the model to prioritize learning from more challenging examples

and thus improving its ability to detect pneumonia cases amidst a majority of pneumonia images. For AE we use the Mean Squared Error (MSE) Loss.

**Data Augmentation** To enhance model robustness and generalization, we applied data augmentation techniques, including resizing (to 256), center cropping (to 224), and normalization (using preset ImageNet means and standard deviations). These transformations introduced necessary variability and helped mitigate overfitting. For VAE we also use CLAHE augmentation to emphasize the area around lung.

**Training and Convergence** Both the classification and anomaly detection tasks exhibited convergence after approximately 10 epochs, signaling efficient learning dynamics and the effectiveness of the chosen model configurations and training strategies.

**Challenges and Solutions** The main technical challenge was the dataset's imbalance, which we addressed using Focal Loss. Another challenge was ensuring model generalization, tackled through strategic data augmentation and careful tuning of hyperparameters.

# 3 Experimental results

**Experimental Settings** We divided our dataset into training, validation, and test sets (Kaggle partitions). The models were trained using the predefined hyperparameters and architecture modifications described previously.

**Metrics Evaluated** The primary metrics for evaluating our model's performance were:

Validation Loss and Accuracy: Monitored after each epoch to determine the convergence of the model and its performance on unseen data during the training phase. Test Accuracy: Calculated as the percentage of correctly classified instances out of the total test set, providing a direct measure of the model's generalization ability. Precision and Recall: Precision (the ratio of true positives to the sum of true and false positives) indicates the model's accuracy in identifying pneumonia cases. Recall (the ratio of true positives to the sum of true positives and false negatives) measures the model's ability to detect all relevant cases of pneumonia.

**Binary Classification Results** We experimented with three models—VGG, ResNet, and AlexNet—and observed that ResNet achieved the highest accuracy among them. We will show the result of ResNet18 model.

- Validation Loss and Accuracy: The model train set attained an accuracy rate of 87.50%. Notably, the peak validation accuracy observed was 93.75%.
- Test Accuracy: On the independent test set, the model secured an accuracy of 86.38%.
- Precision for Pneumonia: the model achieved approximately 82.9% of X-rays classified as pneumonia were accurately identified.

- Recall for Pneumonia: the model's effectiveness in identifying 98.5% of all actual pneumonia cases within the test dataset.

The graph in Figure 1 illustrates the model's loss function across training epochs. A consistent decrease in loss indicates effective learning, with stabilization towards the later epochs suggesting convergence.

The confusion matrix presented in Figure 2 was obtained from evaluating the model on the test set. It provides a detailed breakdown of the model's predictions across the two classes, showcasing its precision and recall capabilities.

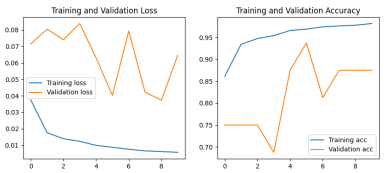



Figure 1: The steady decrease in the loss function across epochs while increase accuracy across epochs, further suggests that the model is effectively learning from the training data.

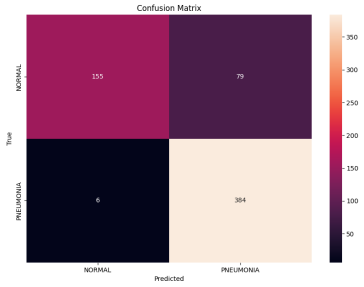



Figure 2: Confusion matrix for the binary classification task.

From these results, the model demonstrates robust performance in distinguishing between healthy and pneumonia-affected chest X-rays, with high overall accuracy and an especially strong ability to identify actual cases of pneumonia, as indicated by the high recall rate.

**Multi-Class Classification Results**

Transitioning from binary to multi-class classification, where the model aimed to distinguish among normal, viral pneumonia, and bacterial pneumonia cases, demonstrated commendable efficacy, albeit with expected complexities inherent to a more nuanced classification task. We experimented with three models—VGG, ResNet, and AlexNet—and observed that VGG achieved the highest accuracy among them. We will show the result of VGG model.

- **Validation Loss and Accuracy:** The model recorded a validation loss of 1.2715 and an accuracy rate of 75.00%, which was also noted as the peak validation accuracy.
- **Test Accuracy:** On the comprehensive test set, a test accuracy of 78.36% was achieved.

The confusion matrix for the multi-class classification provides a granular view of the model's predictive performance across the three categories:

- Normal Cases: True Positives (Normal correctly identified): 143 False Positives (Normal incorrectly identified as Virus or Bacteria): 91 (70 as

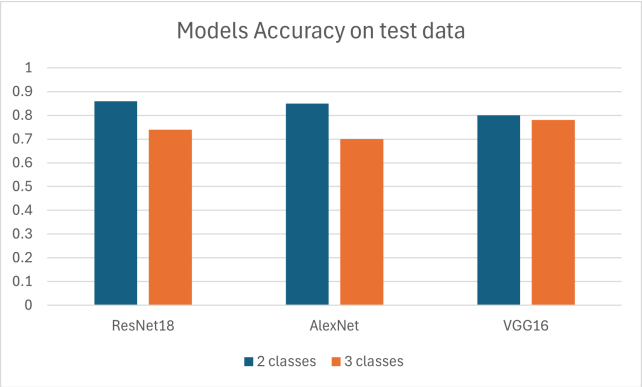



Figure 3: The histogram depicts the distribution of accuracy scores achieved by different models.

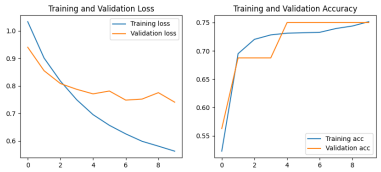



Figure 4: Loss function and Accuracy function across epochs for multi-class.

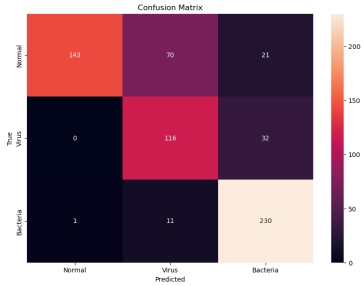



Figure 5: Confusion matrix for multi-class.

Virus + 21 as Bacteria), overall 0.681% precision for normal cases and 0.61% recall.

- Viral Pneumonia Cases: True Positives (Virus correctly identified): 116 False Positives (Virus incorrectly identified as Normal or Bacteria): 32 (0 as Normal + 32 as Bacteria), overall 0.554% precision for normal cases and 0.784% recall.

- Bacterial Pneumonia Cases: True Positives (Bacteria correctly identified): 230 False Positives (Bacteria incorrectly identified as Normal or Virus): 17 (1 as Normal + 11 as Virus), overall 0.975% precision for normal cases and 0.931% recall.

From the multi-class classification results, the model shows a commendable ability to distinguish among normal, viral pneumonia, and bacterial pneumonia cases, though the task introduces added complexity as reflected in the slightly lower accuracy rates compared to the binary classification. The high precision

and recall for bacterial pneumonia cases illustrate the model's effectiveness in identifying this specific condition accurately. However, the precision and recall for normal and viral pneumonia cases, while still substantial, indicate a need for further refinement to reduce misclassifications and improve the model's discriminative power across all classes.

**t-SNE Visualization and k-NN Classification Results** Utilizing t-SNE for dimensionality reduction and k-NN classification, we analyzed the model's ability to differentiate between classes and its consistency in class separability and classification accuracy. The binary classification t-SNE plot reveals distinct separation between the classes, indicating clear class discriminability. For the multi-class task, while clusters for each class are observable, the separation is less pronounced. In k-NN classification, precision and recall for pneumonia in the binary task and for each category in the multi-class scenario were calculated based on the confusion matrices, highlighting the model's nuanced capability to differentiate among normal, viral, and bacterial pneumonia cases. The precision and recall for each class are discussed, reflecting the model's performance.

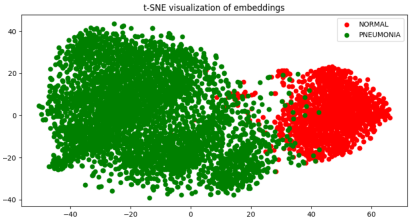



Figure 6: Binary task confusion matrix with KNN predictions.

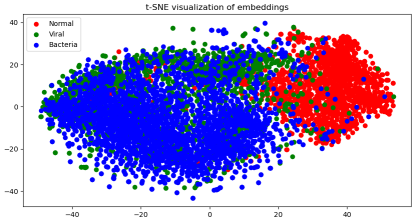



Figure 7: Multi-class task confusion matrix with KNN predictions.

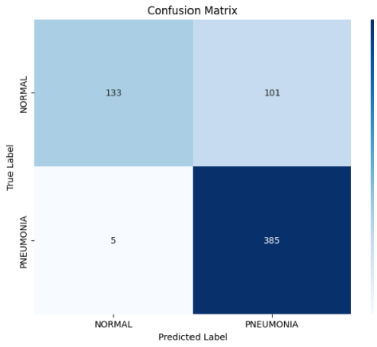



Figure 8: Loss function and Accuracy function across epochs for multi-class.

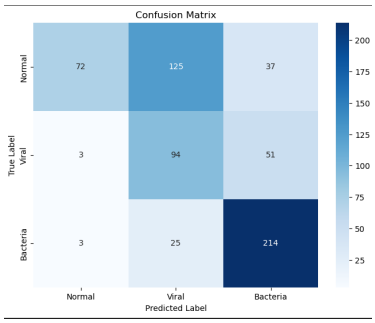



Figure 9: Confusion matrix for the binary classification task for multi-class.

**Anomaly Detection** In our attempt to identify pneumonia cases as anomalies within a dataset presumed healthy, we employed Autoencoders (AE) to learn a generative model of the lung X-ray images. Anomaly detection in this context relied on the model's ability to reconstruct input images with differing levels of fidelity, where higher reconstruction errors could indicate anomalous or pneumonia-affected cases. To evaluate the effectiveness of our anomaly detection method using the Autoencoder (AE) model on chest X-ray images, we conducted experiments on a dataset comprising ostensibly healthy individuals without specific instances of pneumonia. The AE model was trained to learn the latent representation of normal chest X-ray images in an unsupervised manner. We then utilized the learned representations to identify anomalous cases, such as those indicative of pneumonia, based on reconstruction errors. Our approach achieved promising results with an error rate of 0.299, a precision of 0.796, sensitivity of 0.700, specificity of 0.701, and an F1-score of 0.745.

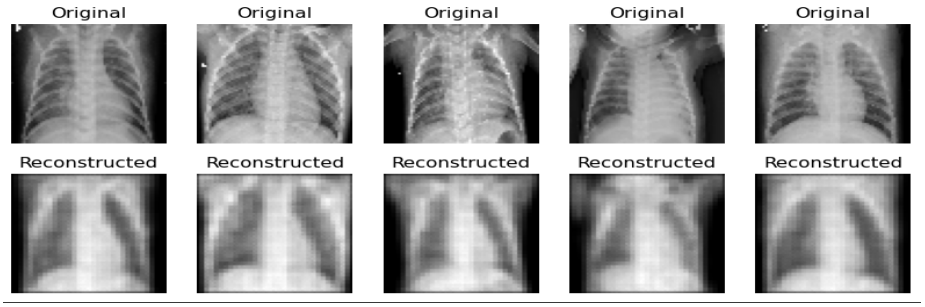
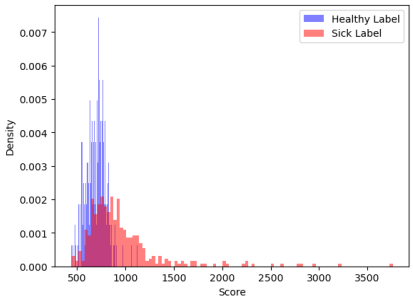
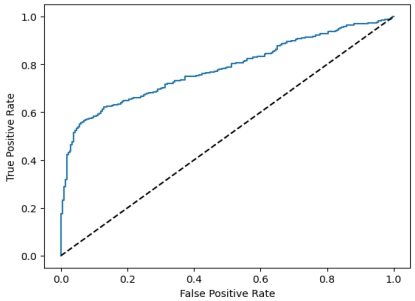



Figure 10: AE reconstructed images.




(a) Overlaid histograms based on predicted scores for two classes

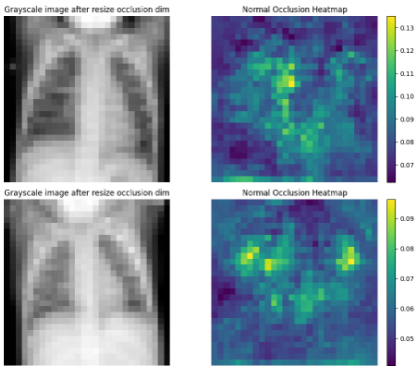



(b) ROC curve illustrating the trade-off between TP rate and FP rate

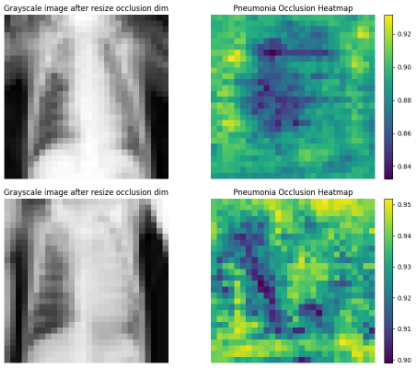
Figure 11: Analyzing Autoencoder (AE) Predictions

# 4 Explainability

In our pursuit to enhance the interpretability of the deep learning models used for classifying chest X-ray images, we implemented the occlusion method. This technique involves systematically obscuring different parts of an input image and observing the impact on the model's output. By analyzing changes in the model's confidence levels as various regions of an image are occluded, we can infer which parts of the image are most critical for the model's decision-making process.




(a) We can see in normal occlusion heatmap how the model predominantly focused on specific areas within the lung fields




(b) We can see in normal occlusion heatmap how the model predominantly focused on specific areas within the lung fields

Figure 12: Comparison between normal and pneumonia occlusion heatmaps

The application of the occlusion method revealed distinct patterns in how the neural network processes images for pneumonia and normal cases. For pneumonia-affected X-rays, the model's attention was not limited to specific areas showing opacities or fluid accumulation; rather, it also significantly factored in the overall layout and structure of the lung fields and their peripheries. This behavior aligns with the clinical approach to diagnosing pneumonia, which involves a detailed analysis beyond the lung fields to include the surrounding regions for signs of infection or fluid buildup.

Conversely, for X-rays classified as normal, the model predominantly focused on specific areas within the lung fields. These areas are expected to appear clear and unobstructed, showcasing the lungs filled with air and free of any infection or fluid, as highlighted in the introductory discussion on diagnosing normal lung conditions. The emphasis on these clear, unobstructed areas by the neural network mirrors the critical features a radiologist would look for when confirming a normal chest X-ray.

This distinction in the model's focus areas—predominantly on certain regions for normal cases and including the layout for pneumonia cases—underscores the complexity of diagnosing pneumonia from X-ray images. It reflects the

necessity of comprehensive imaging analysis, as detailed in "Differentiation of Bacterial and Viral Pneumonia in Children" by Korppi et al.[2]. The study emphasizes the importance of careful examination beyond the immediate lung fields to include peripheral areas, which can show signs of complications or associated conditions crucial for accurate diagnosis.

# 5 Discussion

This project has underscored the potential of deep learning in enhancing the diagnostic process for pneumonia through chest X-ray imaging. By leveraging transfer learning, t-SNE visualization, anomaly detection, and explainability methods, we have developed models that not only improve classification accuracy but also offer insights into the neural network's decision-making processes.

The implementation of the occlusion method for explainability has particularly highlighted the neural network's alignment with clinical diagnostic practices, focusing on lung fields and their peripheries for pneumonia cases and clear, unobstructed areas for normal cases. This alignment with clinical expectations enhances trust in the model's diagnostic capabilities.

## 5.1 Future Research

Future research should focus on addressing the challenge of mode collapse, which remains a significant hurdle in applying generative models for medical anomaly detection. One potential direction could involve exploring alternative architectures or training strategies that encourage the model to preserve the diversity of the input data in its latent space representation. Techniques such as adversarial training, which have shown promise in mitigating mode collapse in generative adversarial networks (GANs), could be adapted for use in VAEs or other generative models suited for medical imaging tasks.

Additionally, further investigation into model explainability and interpretability will be crucial for integrating deep learning tools into clinical workflows. Developing methods that provide more granular insights into model decisions can help bridge the gap between AI outputs and clinical reasoning, ensuring that these tools complement the expertise of medical professionals.

Overall, while challenges remain, the insights gained from this project highlight the transformative potential of deep learning in medical imaging. With continued research and development, these technologies can significantly contribute to early detection, accurate diagnosis, and personalized treatment planning, ultimately enhancing patient care and outcomes.

# 6 Conclusion

In conclusion, this project has demonstrated the effectiveness and potential challenges of applying deep learning models to the classification of chest X-ray

images for pneumonia diagnosis. Through a comprehensive approach that included transfer learning, anomaly detection, and explainability techniques, we gained valuable insights into the capabilities and limitations of current models. The experience underscored the importance of alignment with clinical diagnostic criteria and highlighted areas for future research, particularly in improving model robustness and interpretability. As we move forward, addressing these challenges will be key to unlocking the full potential of deep learning in healthcare.

# 7 Code

To view the **Train Notebook**, click here.
To view the **Test Notebook**, click here.

# References


[1] Diego Ardila, Atilla P Kiraly, Sreenivasa Bharadwaj, Bokyung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961, 2019.

[2] Matti Korppi et al. Differentiation of bacterial and viral pneumonia in children. *Thorax*, 48(7):679–683, 1993.

[3] Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):574–582, 2017.

[4] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[5] Linda Wang, Zhong Qiao Lin, and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):1–12, 2020.