Date: 28.04.2024

Deep learning (CNN) for detecting pneumonia in chest X-ray

Course: Introduction to Image Processing and Analysis

Course Number: 001-2-9301

Lecturer: Dr. Tarin Paz-Kagan

Submitted By:

Name: Moria Gabay **ID:** 305349433

Name: Arad Peleg **ID:** 206068801

Git-Hub-Repository - This link takes you to the project's git hub Repository and it contains: image processing code, model (CNN) code, the images after processing, a guide on how to run the code and information about the Repository both can be found in the README file.

1. Introduction

1.1. Image Processing

Image processing, a fundamental technique for manipulating and enhancing digital images, aims to remove noise and irregularities for improved interpretation. It facilitates image data storage, transmission, and machine perception. Advancements in computing technology have propelled image processing, making it indispensable across fields like remote sensing, medical imaging, and materials science. Key stages include acquisition, enhancement, segmentation, feature extraction, and classification.

The versatility of image processing techniques makes them play a crucial role in analyzing, interpreting, and manipulating visual data to extract meaningful information and enhance decision-making processes (Chitradevi & P.Srimathi, 2014).

1.2. Problem Description and Motivation

Pneumonia detection poses significant challenges in healthcare due to the shortage of radiologists, the necessity for timely diagnosis, and the critical need for accurate disease detection.

While deep learning CNN models have advanced medical image analysis, there's still room for improvement in accuracy and efficiency. Studies (Rahman et al., 2021) have demonstrated the impact of image enhancement techniques on disease detection.

To address these challenges, image processing techniques offer a promising solution. By preprocessing chest X-ray images before inputting them in CNN model, we aim to enhance performance in four stages:

Preprocessing Stage 1: Image Enhancement

- Normalization: Ensuring uniformity in image intensity levels through normalization.
- Crop-Masking: Precisely isolating the region of interest within the chest X-ray image through crop-masking.
- Resize: Adjusting the size of the image to a standardized dimension.

Preprocessing Stage 2: Feature Enhancement

- CLAHE equalization: Leveraging advanced techniques such as CLAHE equalization to effectively reduce noise and enhance image clarity. CLAHE technique offers promising avenues for enhancing the clarity and interpretability of chest X-ray images (Singh & Patel, 2017).
- Bilateral Filtering: Employing bilateral filtering to suppress noise while preserving important image edges.
- Edges using Laplacian of Gaussian: Detecting edges using the Laplacian of Gaussian technique.
- Selective Sharpening: Applying selective sharpening techniques to enhance image clarity and detail.

Preprocessing Stage 3: Segmentation

- Improved Segmentation: Leveraging advanced segmentation algorithms such as Otsu's thresholding for foreground-background separation.
- Watershed Algorithm: Utilizing the watershed algorithm for precise segmentation of distinct anatomical structures and pathological regions. Watershed segmentation has shown promise in enhancing the accuracy

of pneumonia detection by allowing CNN models to focus on specific pathology-associated areas (Liu et al., 2021).

- Closing Operation: Performing the closing morphological operation to further enhance segmentation quality and improve feature extraction.

Preprocessing Stage 4: Model Robustness

- Model Robustness: Strengthening the robustness and generalization capabilities of the CNN model through preprocessing techniques.

By integrating image processing techniques into deep learning CNN models, we anticipate improvements in accuracy, efficiency, and reliability in pneumonia detection. This approach has the potential to streamline diagnostic processes, facilitate earlier disease detection, and ultimately improve patient outcomes.

1.3.Convolution Neural Network (CNN)

A Convolutional Neural Network (CNN) processes grid-like data like images, autonomously learning hierarchical patterns inspired by the biological visual cortex. It efficiently handles vast data by assigning weights to prioritize features, requiring less pre-processing compared to traditional methods (Chen et al., 2021; Yamashita et al., 2018). The CNN architecture consists of convolutional, pooling, and fully connected layers, serving as building blocks for feature extraction and classification.

- **Convolution Layer:** extracts high-level features from input data like images using convolution operations and filters, followed by applying a non-linear activation function like ReLU, resulting in a feature map. **ReLu** is defined as $\mathbf{max}(\mathbf{0}, \boldsymbol{x}) = \begin{cases} \boldsymbol{x}, & \boldsymbol{x} > \mathbf{0} \\ \mathbf{0}, & \boldsymbol{x} < \mathbf{0} \end{cases}$ outputting x if x is positive and 0 otherwise.

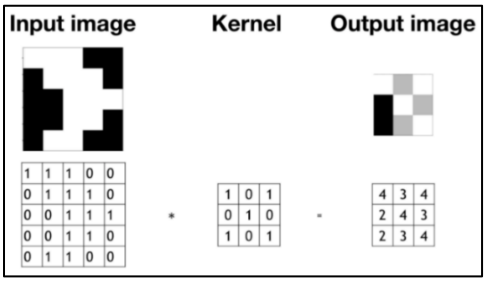

Figure 1.1 - illustrates the output of a convolution between the input image and a filter (kernel).

- **Pooling Layer:** reduces feature map dimensions, improving efficiency by methods like max and average pooling, extracting respective maximum and average values within patches.

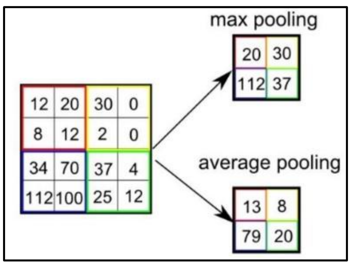

Figure 1.2 - feature map obtained after aggregation operations.

- **Fully Connected Layers:** finalize the CNN, connecting every input node to each output node with learnable weights and employing tailored non-linear activation functions, like a sigmoid function for binary classification.

  $\mathbf{sigmoid}(x) = \frac{1}{1+e^{-x}},$

To train a CNN, data is split into three sets: training, validation, and testing. Input data passes through network layers with weights. Nodes combine inputs, undergo non-linear transformations, and are evaluated using a loss function. Back-propagation iteratively adjusts weights to optimize model performance (Yamashita et al., 2018).

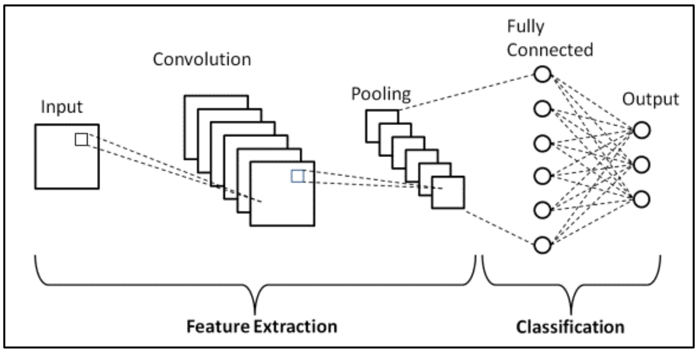

Figure 1.3 - Transition process of input between CNN layers until it becomes an output.

1.3.1. CNNs for detection of pneumonia in chest X-rays

Have gained popularity for their ability to assist radiologists in identifying pneumonia types accurately and cost-effectively. While deep learning cannot replace expert interpretation, it offers valuable medical decision support, saving time. Many researchers analyze chest radiographs using CNNs, with promising results reported, primarily focusing on pneumonia diagnosis, and occasionally classifying bacterial or viral pneumonia types (Rahman et al., 2020).

In our case we use the following CNN methods:

- **Binary classification:** Involves categorizing data into two distinct classes. We designate images as NORMAL (0) or PNEUMONIA (1), training our model to effectively discern between the two categories, which is essential for accurate pneumonia detection.
- **Callbacks:** Utilizing automated tools for enhanced training efficiency, we employ Model Checkpoint Model Checkpoint to save the best model and Early Stopping to prevent overfitting by halting training if no progress is detected.
- **Binary cross entropy:** A function to assess and refine our model's capability to distinguish lung x-ray images with or without pneumonia. This function, ideal for binary classifications for minimizing prediction error and enhancing the accuracy of detecting lung conditions.
- **Exponential Schedule:** We employed an exponential learning rate schedule in our deep learning model, gradually reducing the learning rate during training to prevent convergence to suboptimal solutions and ensure smoother training.
- **Adam:** The optimization process plays a crucial role in improving deep learning models, aiming to minimize prediction errors and enhance accuracy by dynamically adjusting neural network weights. This adaptability enables more efficient learning from the dataset.

2. Database Description

Data collection is essential for robust deep learning (DL) systems, particularly in medical applications where accuracy is paramount. Effective DL relies on extensive, high-quality datasets. Incorporating image preprocessing enhances dataset quality, aiding model absorption of relevant information during training. However, gathering medical data, such as patient X-rays, presents challenges due to privacy issues, limiting access to large, freely available databases.

- **Examples of Data**:
  Our data is divided into two categories. x-ray images with pneumonia and normal images (without pneumonia) the images are pre labeled with 0 and 1 according to their categories (**NORMAL, PENUMONIA**)

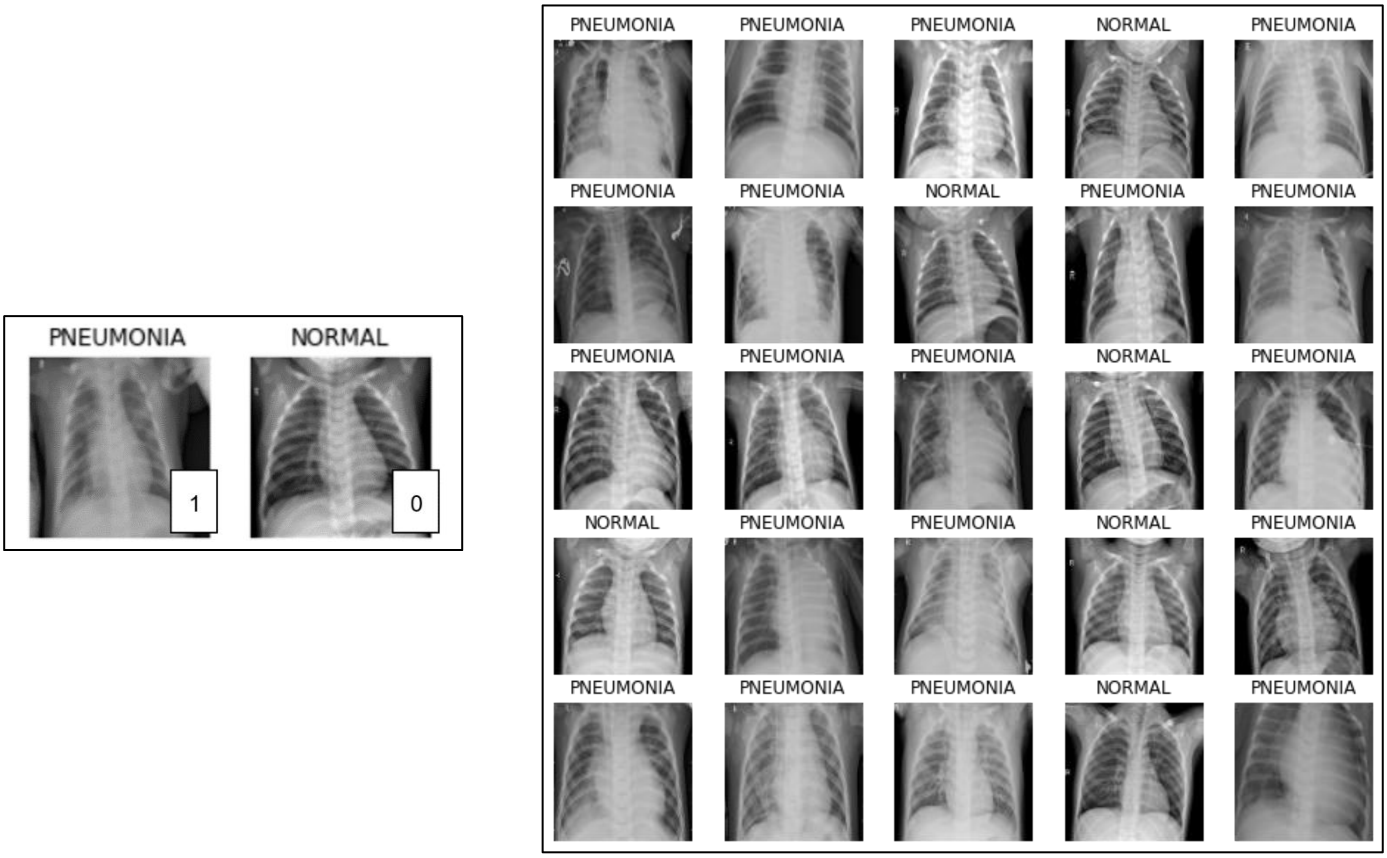

Figure 2.2 - displays image categories (NORMAL, PNEUMONIA) labeled as (0, 1) on the left, with examples of raw images from the dataset (Before Preprocessing) on the right

- **Data Characteristics:**
  - Our images are in JPEG format, which was not chosen deliberately; however, it offers the advantage of smaller file sizes, facilitating storage.
  - Given the varying sizes of the images, it is crucial to resize them uniformly to a standardized dimension, such as 180 x 180, before inputting them into the neural network.
  - **Consideration of the data volume and its distribution is imperative.**

To construct our model, we employed a dataset comprising **5,856 chest X-ray images** of children aged one to five years. The dataset was divided as follows:

The training dataset, comprising 80% of the original dataset, was further partitioned into a training set (80%) and a validation set (20%).During training, the model iteratively learns to recognize patterns and extract features from the data, refining its decision-making accuracy.

In numbers → **5,232 chest images for the training set** - 3,883 with pneumonia and 1,349 normal images We divided the training set into two data sets, a training set (4200 chest images) and a validation set (1032 chest images).

The test dataset, usually comprising 20% of the total data, is utilized to evaluate the model after training. This step assesses the model's performance on new, unseen data to verify its accuracy beyond the training scope.

In numbers → **624 chest images for the test set** - 390 with pneumonia and 234 normal images, without pneumonia.

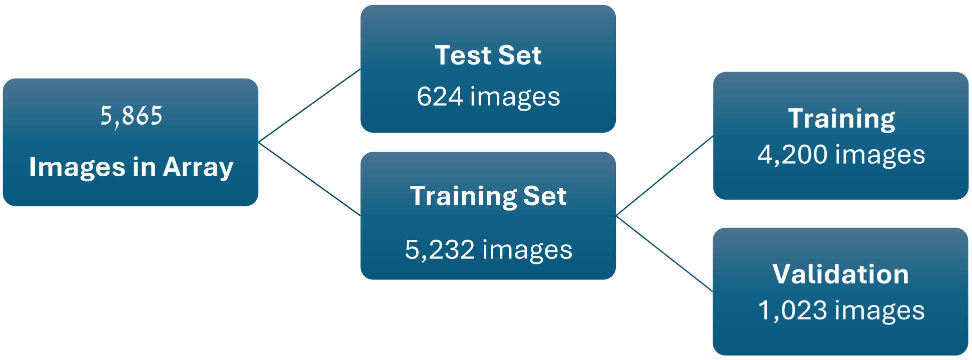

Figure 2.3 - Tree that shows the distribution of data visually.

Balancing the dataset: crucial for CNN models, which perform optimally on balanced data. If the dataset has an unequal number of pneumonia and normal images, balancing is achieved by assigning higher weights to images in the underrepresented category. This approach helps mitigate biases (Over and under fitting) and improves the model's accuracy during training.

- We have 3,883 "PENUMONIA" images compared to 1349 "NORMAL" images.
- **class_weight –** we will balance the data by giving higher weight to "NORMAL" images.

$$\textbf{weight\_for\_ NORMAL} = \frac{1}{COUNT\_NORMAL} \times \frac{TRAIN\_IMG\_COUNT}{2}$$

$$\textbf{weight\_for\_ PENUMONIA} = \frac{1}{COUNT\_PNEUMONIA} \times \frac{TRAIN\_IMG\_COUNT}{2}$$

- **Data Acquisition:**

Cell Press - publishes over 50 high-quality scientific journals covering various fields. It supports open access, with 50% of its articles available as such. For our model, we utilized 5,856 chest X-rays sourced from Cell Press. All images underwent quality control review by three expert doctors in imaging interpretation.

Link For Website: [Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning: Cell]

Link for Data Download: [Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification - Mendeley Data]

Note: If you want to download the data you only need to download ChestXray2017.zip

3. Results

3.1. Work-Flow

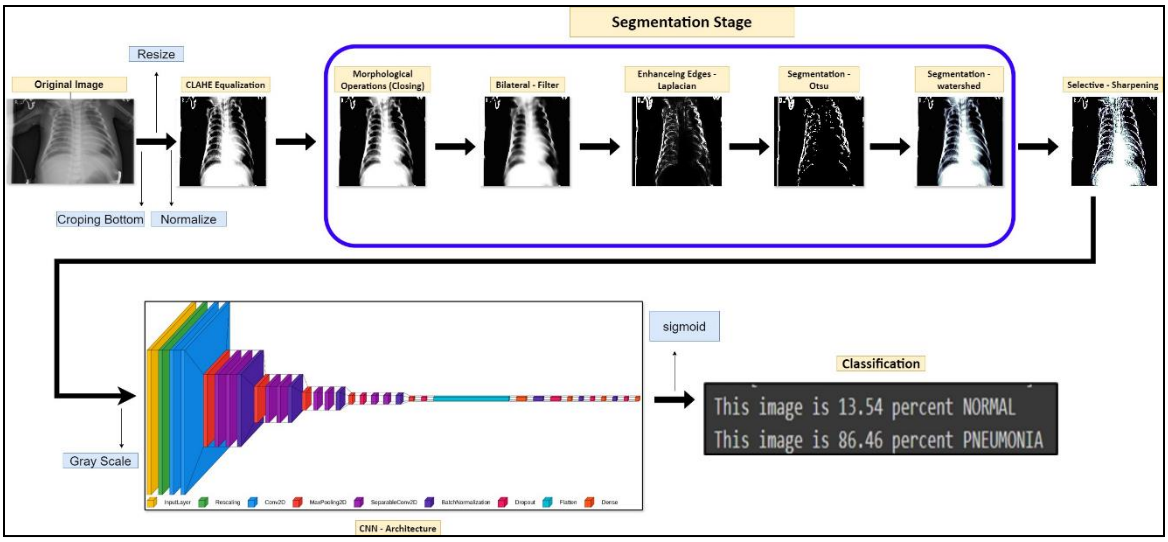

Figure 3.1 – workflow of our project showing each stage of image preprocessing and CNN for classification.

3.2. Model Evaluation

3.2.1. Confusion Matrix

The confusion matrix compares ground truth labels with model predictions, with columns representing predictions and rows representing actual cases. While not a direct performance measure, it serves as a reference point for evaluating model results.

In the matrix we can see that there are 4 basic metrics:

- True positive – (TP): when a chest X-ray with pneumonia is correctly classified.
- True negative (TN): when a chest X-ray without pneumonia is correctly classified.
- False positive (FP): when a chest X-ray without pneumonia is mistakenly classified as having pneumonia.
- False negative (FN): when a chest X-ray with pneumonia is mistakenly classified as not having pneumonia.

Only using deep learning (CNN)

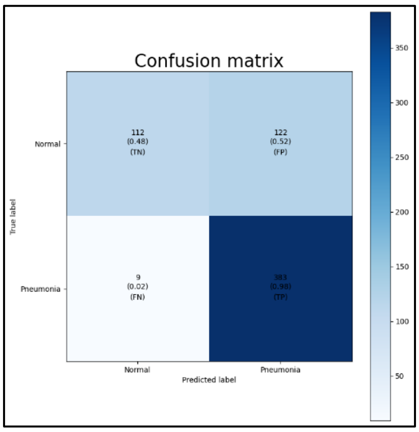

Image Pre-Processing + CNN (Presentation)

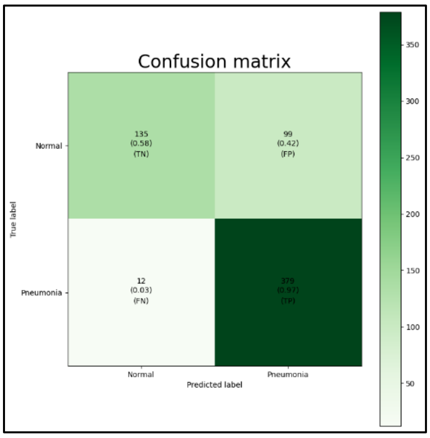

Image Pre-Processing + CNN (Final)

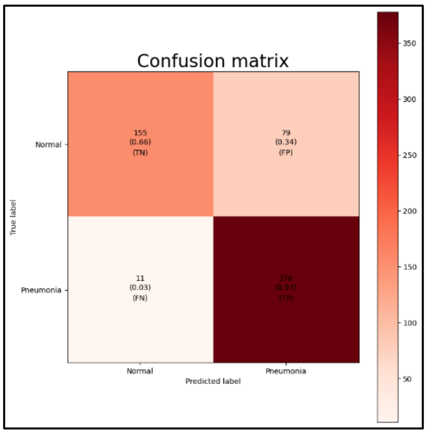

Figure 3.2 - confusion matrix outputs in each stage of the project

Note: We obtained these results by averaging multiple model runs.

3.2.2. Model Performance

Using these basic metrics as a foundation, we can develop additional evaluation metrics to assess the model's performance.

- **Binary Accuracy:** Measures the proportion of correct predictions (both true positives and true negatives) among the total number of cases evaluated. $\frac{TP+TN}{TP+TN+FP+FN}$
- **Precision:** Measures the proportion of true positive predictions in the total positive predictions made. $\frac{TP}{TP+FP}$
- **Recall:** Measures the proportion of true positive predictions out of all actual positive cases. $\frac{TP}{TP+FN}$
- **Loss function**: evaluates how well a model's predictions match the actual target values, guiding the optimization process by quantifying error or deviation.

We'll demonstrate the metrics extracted from the confusion matrix for the test dataset at each project phase, enabling comparative analysis. These metrics assess the model's effectiveness with new images post-training and validation.

Only using deep learning (CNN)	Image Pre-Processing + CNN (Presentation)	Image Pre-Processing + CNN (Final)
Loss: 2.3936	Loss: 0.7154	Loss: 0.5661
Binary Accuracy: 0.79	Binary Accuracy: 0.8227	Binary Accuracy: 0.855
Precision: 0.7524	Precision: 0.7911	Precision: 0.8274
Recall: 0.9897	Recall: 0.9696	Recall: 0.9705

Figure 3.3 - results of metrics to evaluate the model of each stage of the project.

The pneumonia detection model achieves an **85.5%** accuracy and outstanding recall at **97.05%,** indicating it identifies nearly all actual cases. However, its precision at **82.74%** suggests some false positives, and a loss of **0.5661** reflects moderate prediction errors. Overall, it effectively diagnoses pneumonia but has room for precision enhancement.

3.2.3. Receiver Operating Characteristic (ROC)

The ROC curve evaluates binary classifiers by plotting true positive against false positive rates, summarizing performance with an AUC value ranging from 0.5 to 1. Threshold selection balances false positives and negative.

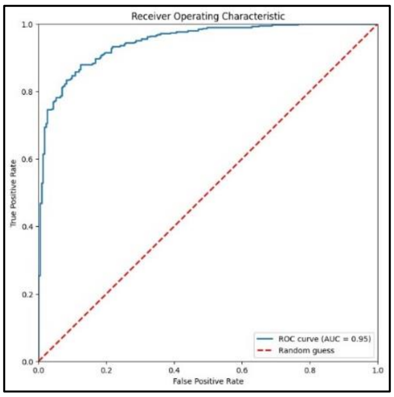

Figure 3.4 - shows the ROC curve, with the AUC (blue) representing the binary classifier's performance compared to the Random Guess line (red).

With an AUC of **0.95**, our model effectively discriminates between positive and negative classes, indicating its potential to perform well on new X-ray images.
Note: this graph presents one of the runs of our model that was the best.

4. Discussion

In the workflow of our project, we used a more extensive set of preprocessing techniques, and it has clearly manifested its efficacy in improving our model's performance in detecting pneumonia from chest X-rays. We meticulously engineered a sequence of preprocessing steps, including Clahe equalization, Morphological Operations (Closing), bilateral filtering, edge enhancement with Laplacian, segmentation via Otsu and the watershed algorithm, and selective sharpening. In our opinion with this approach each step contributed to the refinement of the image features critical for accurate classification. These steps, visualized in our workflow, systematically enhance the image quality, and highlight diagnostic features, preparing the input for the CNN in an optimized manner.
The measured improvements for evaluating the model performance metrics through these stages are significant. From the initial deep learning model to the final model with comprehensive preprocessing, we observed a loss decrease by a lot from **2.3936** to **0.5661** which means most predictions in the test set are right and that error between the predicted label and truth label has decreased, an accuracy increases from **79%** to **85.5%** which is an improvement of **6%** overall in detecting normal and pneumonia images in general , a precision rise from **75.24%** to **82.74%** which means a better prediction for normal lung images , and recall which went a little down from **98.97%** to **97.05%** but yet indicating the model's effectiveness in predicting images with the presence of pneumonia. These metrics not only demonstrate an optimized model but also underscore the crucial role of image preprocessing in deep learning for medical imaging.

Moreover, the Receiver Operating Characteristic (ROC) curve of our final model highlights its predictive capability and reliability. With an Area Under the Curve (AUC) of 95%, the model demonstrates superior ability to distinguish between pneumonia and non-pneumonia cases, surpassing random chance by a significant margin. This high AUC value indicates strong separability and suggests that the model will generalize effectively to new images. While the displayed result may represent a single run, the consistency of the **95**% score across multiple runs underscores its reliability.

5.Conclusion

The findings from our project clearly emphasize the substantial role that a comprehensive preprocessing pipeline plays in optimizing the performance of Convolutional Neural Networks for the detection of pneumonia in chest X-ray imagery. Our approach has led to significant improvements in our model accuracy. This is reflected in the pronounced decrease in loss and an appreciable increase in binary accuracy. Furthermore, the precision of our model rose indicating a greater probability of true negative identifications (normal lungs), which is crucial for patient diagnosis and treatment planning. While we noted a minor reduction in recall, the value remains high ensuring that our model reliably identifies most true pneumonia cases. This slight trade-off resulted in a model that balances the need for both sensitivity and precision in medical diagnostics. Additionally, the good AUC attests to the model's discriminative power and confirms its utility as a reliable diagnostic tool.
The accuracy of the model is relatively good, but in the case of human life, 100% accuracy is necessary, so further research is needed to improve the model as much as possible. This work shows beneficial impact of preprocessing on the efficacy of CNNs in a medical imaging context and lays the groundwork for more research on this topic because of the potential it has to save human lives. Further research could involve expanding the dataset to include a wider array of imaging conditions, experimenting with additional preprocessing methods, and employing more sophisticated CNN architectures that include the optimized parameters to enhance model performance and using transfer learning to adapt models trained on one task to another related task could also gain beneficial insights. Additionally, we should aim to translate these research findings into practice by evaluating the model in real-world clinical settings and assessing its impact on the diagnostic workflow. Investigating model interpretability and ensuring the explainability of AI decisions will enhance trust and adoption among clinicians.

In the end the target of this work is to harness the power of deep learning and image processing to provide robust, automated tools that support radiologists, reduce diagnostic errors, and ultimately improve patient care outcomes. With these advancements, we may see in the future a transformative shift in radiological practices, leveraging the capabilities of AI to amplify human expertise.

6. Bibliography

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. In *Remote Sensing* (Vol. 13, Issue 22). https://doi.org/10.3390/rs13224712

Chitradevi, B., & P.Srimathi. (2014). An Overview on Image Processing Techniques. *International Journal of Innovative Research in Computer and Communication Engineering*.

Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability 2021, Vol. 13, Page 1224*, *13*(3), 1224. https://doi.org/10.3390/SU13031224

Rahman, T., Chowdhury, M. E. H., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., & Kashem, S. (2020). Transfer learning with deep Convolutional Neural Network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences (Switzerland)*, *10*(9). https://doi.org/10.3390/app10093233

Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. Bin, Islam, M. T., Al Maadeed, S., Zughaier, S. M., Khan, M. S., & Chowdhury, M. E. H. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*, *132*. https://doi.org/10.1016/J.COMPBIOMED.2021.104319

Singh, B., & Patel, S. B. (2017). Efficient Medical Image Enhancement using CLAHE Enhancement and Wavelet Fusion. *International Journal of Computer Applications*, *167*(5), 1–5. https://doi.org/10.5120/IJCA2017913277

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. In *Insights into Imaging* (Vol. 9, Issue 4). https://doi.org/10.1007/s13244-018-0639-9