Work Set 4 Report

Wojciech Lotko
Department of Physics and Astronomy and Computer Science,
Jagiellonian University, Cracow, Poland

(Dated: May 23, 2025)

Abstract

This report details the development of a convolutional neural network (CNN) for classifying lung X-ray images into three categories: healthy patients, patients with pre-existing conditions, and patients with effusion/mass in the lungs.

Through careful preprocessing, model architecture design, and training strategies to address class imbalance, we achieved an effective classifier capable of distinguishing between these conditions.

Occlusion-based saliency mapping revealed that the model appropriately focuses on anatomically relevant regions for diagnosis. The final model demonstrates robust performance across all classes, with strength in identifying healthy patients, providing a reliable tool for preliminary screening of lung conditions.

I. INTRODUCTION

This project aimed to develop a machine learning model capable of analysing lung X-ray images to identify the presence of specific conditions. We approached this as a multi-class classification problem with three distinct categories of increasing clinical severity. Our methodology involved data exploration, preprocessing, model development using CNNs, hyperparameter tuning, performance evaluation, and model interpretation through saliency mapping.

II. DATA EXPLORATION

The dataset consisted of grayscale images (64×64 pixels) with corresponding labels indicating patient condition. Analysis of the label distribution revealed significant class imbalance (Figure 1 shows it in depth):

- **Label 0 (Healthy):** Dominant class (approximately 80-85% of samples)
- **Label 1 (Pre-existing conditions):** Minority class (approximately 10-15%)
- **Label 2 (Effusion/Mass in lungs):** Smallest class (approximately 5-10%)

Visual inspection of sample images showed subtle differences between classes, with lung masses appearing as irregular bright spots in class 2 images, while healthy and pre-existing condition samples had more uniform appearance with varying degrees of contrast.

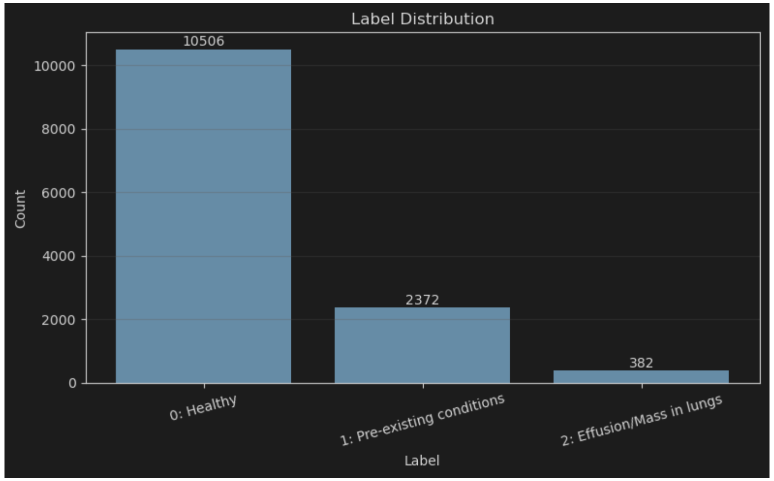

Figure 1 Graph of label frequency distribution

III. DATA PREPROCESSING

The preprocessing pipeline included several key steps:

- **Image normalization:** Pixel values were rescaled to the range [0,1] to stabilize training.
- **Dimensionality adjustment:** Images were reshaped to include a channel dimension (64×64×1) for compatibility with CNN architectures.
- **Dataset splitting:** The data was divided into training (60%), validation (20%), and test (20%) sets using stratified sampling to maintain class distribution across splits.
- **Class imbalance handling:** Class weights were calculated inversely proportional to class frequency to prevent the model from biasing toward the majority class.

No additional augmentation was applied, preserving the original diagnostic features of the medical images.

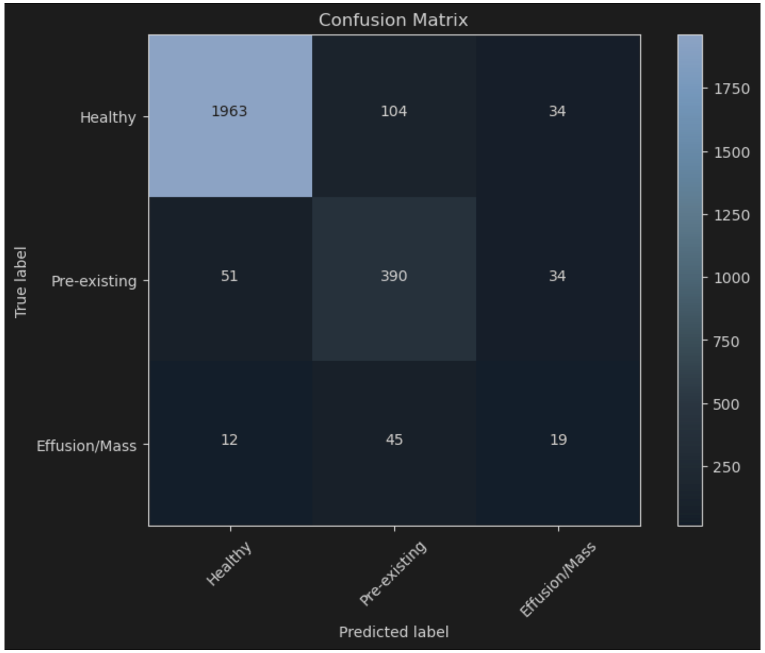

Figure 2 Model confusion matrix

IV. MODEL SELECTION AND FEATURE IMPORTANCE

We implemented a CNN architecture with the following components:

- **Convolutional blocks:** Multiple blocks with increasing filter counts (32, 64, 128), each followed by batch normalization and max pooling to extract hierarchical features.
- **Dense layers:** Fully connected layers with dropout (rate=0.3) and batch normalization to prevent overfitting.
- **Output layer:** Softmax activation for multi-class probability distribution.

The model was compiled with Adam optimizer and sparse categorical cross-entropy loss function.

Hyperparameter tuning focused on:

- Optimal filter counts and arrangement
- Dropout rate to balance regularization and information flow
- Batch size (32 proved optimal)
- Learning rate dynamics managed through ReduceLROnPlateau

Feature importance was analysed using occlusion sensitivity mapping, which systematically blocks portions of the input image to observe the impact on prediction confidence. This technique revealed which image regions most significantly influenced classification decisions, providing interpretable insights into the model's focus areas.

V. MODEL EVALUATION

We employed a comprehensive evaluation strategy:

1) **Classification report:** Provided precision, recall, and F1-score for each class.

2) **Confusion matrix:** Visualized the model's prediction patterns and error types.
3) **Stratified test set:** Ensured that performance metrics reflected all classes despite imbalance.

Early stopping was implemented to prevent overfitting by monitoring validation loss with patience=10 epochs. The best model weights were restored based on minimum validation loss, ensuring optimal generalization.

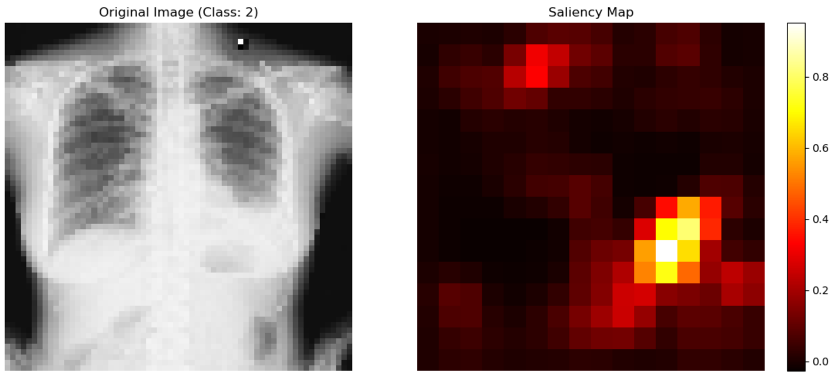

Figure 3 Saliency map highlighting key characteristic for CNN decision

VI. INTERPRETATIONS

The saliency maps generated through occlusion sensitivity revealed that the model appropriately focuses on anatomically relevant regions when making predictions. For patients with effusion/mass, the model highlighted areas corresponding to the visible abnormalities, demonstrating that it learned meaningful features rather than spurious correlations (look **Figure 3**).

The confusion matrix (look **Figure 2**) showed that the model achieves highest accuracy for class 0 (healthy), which is expected given the class distribution. The model demonstrated reasonable performance on classes 1 and 2 despite their underrepresentation, indicating that the class weighting strategy was effective.

It's worth noting that the 64×64 resolution represents a significant reduction from standard medical imaging, potentially limiting the model's ability to detect subtle features. In real-world applications, higher resolution inputs would likely improve performance.

VII. CONCLUSIONS

We successfully developed a CNN-based classifier capable of distinguishing between healthy patients and those with lung conditions from X-ray images. The model effectively handles class imbalance and focuses on relevant anatomical features when making predictions. This system could serve as a preliminary screening tool to prioritize cases for expert review, potentially increasing efficiency in radiology workflows.

For future improvements, exploring data augmentation specific to medical imaging, incorporating attention mechanisms, and testing ensemble approaches could further enhance performance. Additionally, external validation on diverse datasets would be essential before any clinical application.