
Multimodal Chest X-ray Classification with CNN

Misumi Matsudo
1010062254, misum.matsudo@mail.utoronto.ca
[Link to Google Colab](#)

Introduction

Chest radiography is one of the most widely used imaging techniques in clinical practice, but interpretation requires significant expertise and time. Misdiagnosis or delays can have serious consequences, especially in resource-limited settings. Automated classification systems have the potential to support physicians by providing rapid and consistent assessments of common thoracic conditions.

The goal of my project is to build a classifier that can distinguish between four categories in chest X-ray images: No Finding, Pneumonia, Effusion, and Cardiomegaly. I will first establish a baseline using a simple convolutional neural network (CNN) trained on images alone. After this, I will investigate whether incorporating patient metadata such as age, gender, and view position can improve classification performance, particularly for minority classes. This problem is interesting because it highlights both the promise of deep learning in medicine and the challenges posed by class imbalance. Deep learning is an appropriate approach since CNNs are capable of extracting complex spatial features directly from raw images, making them well-suited for medical image analysis..

Illustration

Figure 1 illustrates the overall idea of the proposed system. The baseline model takes only chest X-ray images as input, processes them through a CNN, and outputs one of four disease categories. In the improved version, metadata (age, gender, view position) is concatenated with the image features before classification, creating a multimodal architecture.

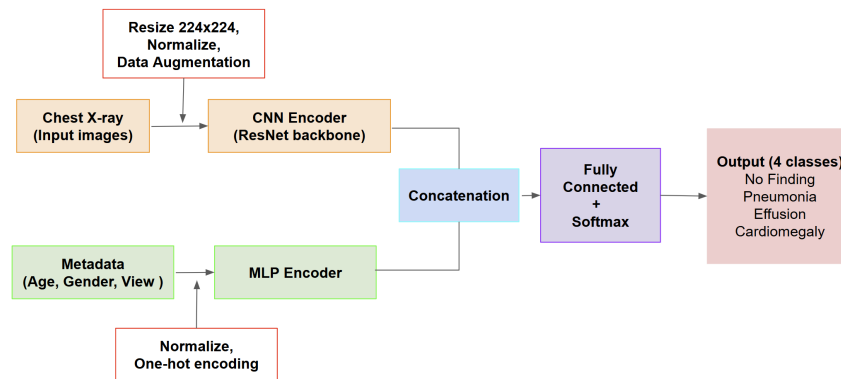


Figure 1: Proposed architecture: CNN for image encoding, Meta Data branch, fusion via Classifier, and final Output.

Background & Related Work

I will use the NIH ChestX-ray14 dataset introduced by Wang et al. (2017), which contains 112,120 frontal-view chest radiographs labeled with 14 disease categories. This dataset is one of the largest publicly available resources for chest radiography and has been widely adopted in deep learning

research. Recent research has shown that deep learning can provide strong results for chest X-ray analysis but also highlighted several open challenges. Khader et al. (2023) demonstrated that multimodal models that combine image features with non-image clinical parameters can achieve higher diagnostic performance than image-only approaches, motivating the integration of metadata in my project. Chen et al. (2025) provided a comprehensive review of CNN-based methods for medical image classification and concluded that deep learning models are highly effective across multiple diagnostic tasks, which supports my decision to begin with a CNN baseline. Baltruschat et al. (2019) compared different CNN architectures on the ChestX-ray14 dataset and showed that combining metadata such as age and body position with image features improved classification accuracy. However, they also reported that small sample sizes for certain classes introduced high variability, which aligns with the class imbalance problem I expect to face in my own work. In addition, Irvin et al. (2019) emphasized that evaluation metrics beyond accuracy, such as AUROC and recall, are essential for medical AI because overall accuracy can mask failures on clinically important but rare conditions.

Data Processing

From the ChestX-ray14 dataset, I selected four categories: No Finding, Pneumonia, Effusion, and Cardiomegaly. The filtered subset contains 6,433 images with a highly imbalanced distribution: 5,999 No Finding (93.25%), 34 Pneumonia (0.53%), 309 Effusion (4.80%), and 91 Cardiomegaly (1.41%).

All images will be resized to 224×224 pixels and normalized for compatibility with pretrained CNNs. To mitigate imbalance, I will apply class-weighted loss and data augmentation (random rotations, flips, brightness adjustments) to the training set only. The dataset will be split into 70% train, 15% validation, and 15% test.

Architecture

The final model will be a CNN-based classifier, extended to incorporate patient metadata for multi-modal learning. The image branch will use a pretrained CNN backbone such as ResNet to extract high-level visual features from resized chest X-rays. In parallel, tabular metadata features (age, gender, view position) will be encoded using a small multilayer perceptron. The outputs of both branches will be concatenated and passed through fully connected layers and a softmax layer to predict one of four disease categories. This design allows the model to leverage both imaging and non-imaging information and should provide a stronger representation than an image-only baseline.

Baseline Model

As my baseline, I will train this CNN using only image data. I expect that the model will perform well on the majority class (No Finding) but will struggle with minority classes, particularly Pneumonia, which makes up only about 1% of the dataset. I will evaluate the baseline not only with overall accuracy but also with confusion matrices, precision, recall, and F1-scores per class. These metrics will likely reveal that the high accuracy is misleading, as the classifier may fail to detect rare conditions. This sets up a clear motivation for improvement.

Ethical Considerations

The use of artificial intelligence in healthcare raises important ethical concerns, particularly related to fairness and equity. Ueda et al. (2023) reviewed the impact of AI on healthcare and argued that ensuring fairness directly affects patient safety and equitable access to medical services. Their findings underscore the need to address disparities that may arise from biased datasets or imbalanced model performance. In my project, this means I must go beyond overall accuracy and evaluate per-class performance to ensure that minority conditions, such as Pneumonia or Cardiomegaly, are not systematically overlooked.

References

- [1] Wang et al., NIH ChestX-ray14 dataset 2017par [2] Khader et al., Multimodal Deep Learning for Integrating Chest Radiographs and Clinical Parameters: A Case for Transformers, Radiology 2023.
- [3] Chen et al., A Review of Convolutional Neural Network Based Methods for Medical Image Classification, Computers in Biology and Medicine 2025.
- [4] Baltruschat et al., Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification, PLoS ONE 2019.
- [5] Irvin et al., CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison, AAAI 2019.
- [6] Ueda et al., Fairness of Artificial Intelligence in Healthcare: Review and Recommendations, PMC 2023.