

# Task Summary and Reflection

Enno Belz

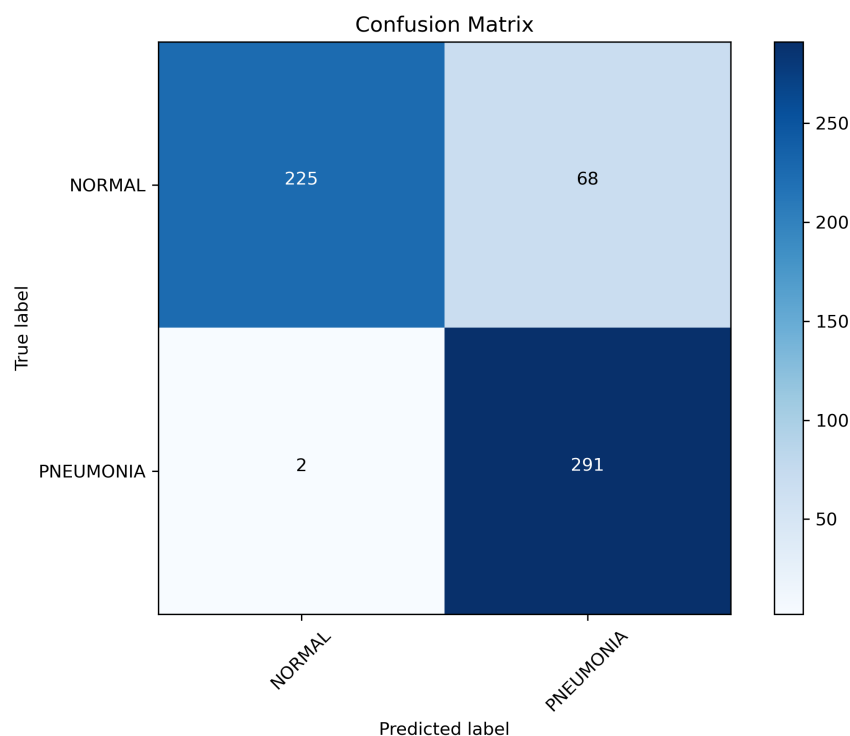
## Subtask 1

To run the delivered script the packages `nibabel` and `matplotlib` are required to be installed.

## Subtask 2

The second task involved developing a binary classifier to categorize chest X-ray images as either Normal (label = 0) or Pneumonia (label = 1). A Convolutional Neural Network (CNN) was chosen for its proven performance on image data and its potential for explainability through techniques like Grad-CAM. This could enable a deeper understanding of model decisions in downstream analyses.

The model architecture consisted of two convolutional layers with max-pooling, followed by three fully connected layers. This lightweight design serves as a "default" configuration, allowing for easy replacement with more or less advanced architectures. After only one epoch of training, the model achieved a validation accuracy of 88.05%, indicating promising initial performance.



*Figure 1: Confusion matrix illustrating the model's performance, highlighting high sensitivity for Pneumonia detection and a higher rate of misclassification for Normal samples.*

The confusion matrix (Figure 1) provides further insight into the model's predictions. It achieved a sensitivity of 99.31%, correctly identifying nearly all Pneumonia cases, with only two False Negatives. However, Normal cases were more prone to misclassification, leading to a higher proportion of False Positives and a specificity of 76.79%.

In diagnostic applications, sensitivity is often prioritized over specificity, as False Positives (misdiagnosing healthy individuals) are less critical than False Negatives (missing sick individuals), which can have severe consequences. The observed performance imbalance likely arises from the dataset's skewed class distribution, with Pneumonia samples (3,687) significantly outnumbering Normal samples (997), creating a bias toward the majority class leading to a potential overfit.

Additional challenges may include data noise, labeling errors, or subtle image quality variations. To mitigate these issues, several strategies can be employed: applying general data transformations to improve generalization. Constraining these augmentations in order to increase the number of normal samples and therefore balance the dataset is possible as well. However, the augmentation techniques have to be chosen carefully. I suggest augmentations like slight rotations, edge cropping, or vertical flipping to preserve the realism of the grayscale medical data. While undersampling Pneumonia data is an option, it is less desirable as it sacrifices valuable and potentially rare data.

Interestingly, the Pneumonia samples consist of 2,384 bacterial and 1,303 viral cases, suggesting the possibility of extending the task to a multi-class classification problem, which could better balance the dataset.

When working with imbalanced datasets, it's important to consider that these imbalances may mirror real-world scenarios. In our case, this would correspond to a substantial majority of healthy samples, as most individuals are typically healthy. Identifying the comparatively fewer sick samples is not only more challenging but also more reflective of realistic clinical settings and augmentation as well as balancing techniques have to be performed strategically.

Finally, exploring more advanced models and training routines, such as pre-trained CNNs like ResNets or different optimizers, could enhance performance. Baseline comparisons using simpler methods, such as Logistic Regression on PCA-reduced features or Lasso regularization, could also provide valuable insights into the dataset and model complexity trade-offs.

I tracked my progress on this technical task on my [GitHub](#) page, where you can find a jupyter notebook where I did the prototyping for both of the tasks.