

ANALYSIS OF COVID-19 CHEST X-RAYS:

Report 2: Modeling Report

Saniya Arfin, Yvonne Breitenbach, Alexandru Buzgan

May 2025

Contents

1	Introduction	2
2	Methods	2
2.1	Prepare Data for Modelling	2
2.1.1	Prepare Data for Machine Learning Models	3
2.1.2	Prepare Data for Deep Learning Models	3
2.2	Metrics to Evaluate the Model Results	3
2.3	Used Models: Machine Learning	3
2.3.1	? Decision Trees ?	3
2.3.2	? Random Forst ?	3
2.3.3	Used Models: Optimized Machine Learning	3
2.4	Used Models: Deep Learning	3
2.4.1	? Model 1 ?	3
2.4.2	? Model 2 ?	3
2.5	Transfer Learning	3
3	Results	4
3.1	Results ? Decision Trees ?	4
3.2	Results ? Random Forst ?	4
4	Conclusion	4
5	Example include image and table	4

1 Introduction

- What kind of machine learning problem is your project like? (classification, regression, clustering, etc)
- What task does your project relate to? (fraud detection, facial recognition, sentiment analysis, etc)?
- ...

2 Methods

2.1 Prepare Data for Modelling

Just a summary of the points we found out in report 1, that we have to do with the images. Maybe we need a different order.

- convert to grayscale
- Contrast Enhancement with CLAHE
- resize the masks and combine them with images
- Pixel Value Normalization
- Label Encoding: Convert categorical class labels (e.g., COVID, Normal, Pneumonia, Lung Opacity) into numerical format for compatibility with machine learning algorithms.
- Dataset Splitting: Partition the dataset into training, validation, and test subsets to enable robust model evaluation and prevent data leakage.
- Address Class Imbalance: Since the distribution between the classes is not balanced (for example Viral Pneumonia images are far less numerous than the Normal images) we have to address this problem before the training stage. Appropriate strategies will be applied, such as:
 - Class weighting during model training
 - Oversampling/undersampling
 - Data augmentation for minority classes
- This ensures the model does not become biased toward the majority classes and maintains fair performance across all categories.

Maybe we need two different data preparation pipelines: one for the machine learning models and another one for the deep learning models

2.1.1 Prepare Data for Machine Learning Models

2.1.2 Prepare Data for Deep Learning Models

2.2 Metrics to Evaluate the Model Results

Idea is to determine metric which we want to use to evaluate the performance of all our tested models.

- What is the main performance metric used to compare your models? Why this one?
- Did you use other qualitative or quantitative performance metrics? If yes, detail it.

2.3 Used Models: Machine Learning

Some ideas, questions for this chapter:

- What algorithms have you tried?
- Describe which one(s) you selected and why?
- Did you use parameter optimization techniques such as Grid Search and Cross Validation?
- Have you tested advanced models? Bagging, Boosting, Deep Learning... Why?

2.3.1 ? Decision Trees ?

just a random example

2.3.2 ? Random Forst ?

just a random example

2.3.3 Used Models: Optimized Machine Learning

Romain mentioned something like that, that there is a library for that.

2.4 Used Models: Deep Learning

2.4.1 ? Model 1 ?

2.4.2 ? Model 2 ?

2.5 Transfer Learning

we have to find out more about that, but Romain mentioned it.

3 Results

Some ideas, questions for this section:

- Have you analyzed the errors in your model?
- Did this contribute to his improvement? If yes, describe.
- Have you used interpretability techniques such as SHAP, LIME, Skater... (Grad-CAM for Deep Learning...)
- What has (or not) generated a significant improvement in your performance?

3.1 Results ? Decision Trees ?

just a random example

3.2 Results ? Random Forst ?

just a random example

4 Conclusion

Overall Conclusion Which model performed best What worked, what did not work? Why?

5 Example include image and table

Put images you want to include to this report in the subfolder: figures_report_2 and use relative paths inside this report. Upload them on git aswell.

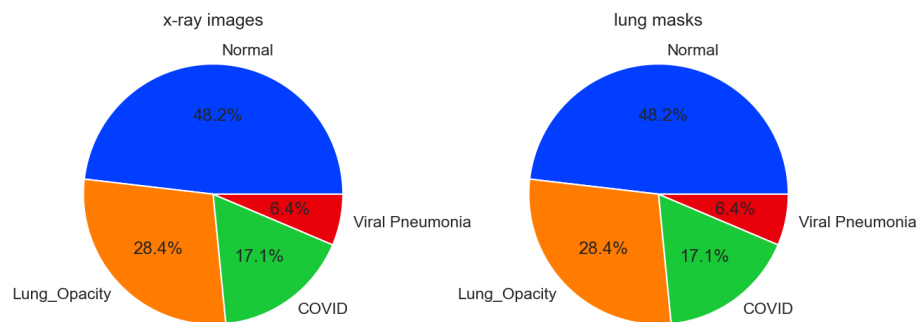


Figure 1: Examples of image

X	Y	Z
1	2	3
4	5	
6		7

Table 1: Example of table

Figure 1 shows ...

Table 1 shows ...