

# CS3 Rubric – Detecting Pneumonia

DS 4002 – Spring 2025 - Kaitlin Blakeslee

Due: TBD

Submission format: Upload link to github repo to canvas

## Individual Assignment

**General Description:** Submit to canvas a link to your case study repository as well as a single page pdf reflecting on this assignment.

Preparatory Assignments – Everything in the course

**Why am I doing this?** Pneumonia is a leading cause of hospitalization and death in the elderly and young children. Early and accurate detection is critical. You will explore how deep learning can assist clinicians by building and evaluating a convolutional neural network (CNN) for automated diagnosis.

**What am I going to do?** The GitHub repository for this case study can be found at <https://github.com/kaitlinblakeslee/CS2DS4002/tree/main>. Begin by reading the CS3-Hook Document and the articles included in the MATERIALS folder to familiarize yourself with the project's context and real-world relevance. Then, navigate to the DATA folder and access the RSNA Pneumonia Detection Challenge dataset. Spend time exploring the dataset to understand the clinical problem it represents. Once you have a strong grasp of the task, your challenge is to train and evaluate a binary classification model using a Convolutional Neural Network (CNN) architecture—specifically, ResNet50, a powerful deep learning model commonly used in medical imaging research. You will apply image preprocessing and augmentation techniques to optimize model performance and reduce overfitting. After training your model, you must report its performance using at least three evaluation metrics, such as accuracy, recall, and AUC. Finally, you will reflect on your model's clinical relevance and limitations, considering how this tool might function in a real hospital setting and where caution or improvement is needed.

Your final deliverables should include:

- Github repository
- Single page pdf reflection

All of this will be submitted electronically via a link to a github repository.

### Tips for success:

- Clean, documented code (comment where needed!)
- Clear README with setup instructions and project summary
- Thoughtful model evaluation and explanation of results
- Realistic reflection on model limitations and next steps
- Don't forget: keep everything in the GitHub repo!

**How will I know I have Succeeded?** You will meet expectations on this case study when you follow the criteria in the rubric below.

Formatting	<ul style="list-style-type: none"> <li>• One Github repository (submitted via a link on Canvas)</li> <li>• The repository should contain: <ul style="list-style-type: none"> <li>o Data and Code <ul style="list-style-type: none"> <li>▪ Submit the code you write for any data cleaning and model building for this assignment</li> <li>▪ Include comments throughout your code</li> <li>▪ Include any additional data used besides what was included in the DATA folder (RSNA dataset)</li> </ul> </li> <li>o Reflection <ul style="list-style-type: none"> <li>▪ Submit the document as a single page pdf file</li> </ul> </li> </ul> </li> </ul>
Data and Code	<ul style="list-style-type: none"> <li>• <u>Goal</u>: All written code will be included in this folder of a repository</li> <li>• The original repository already has a DATA folder with RSNA data but if other data was used include it in this folder as well</li> <li>• Your code should include an ‘exploratory data analysis’ jupyter notebook with any data cleaning and visualizations representing the RSNA dataset (number of pneumonia vs non-pneumonia cases, x-ray image example, other demographic information, etc.)</li> <li>• Your code should include a ‘model jupyter’ notebook with your CNN model and evaluation metrics</li> </ul>
Reflection	<ul style="list-style-type: none"> <li>• <u>Goal</u>: Reflect on your results, challenges, and what you learn throughout this project.</li> <li>• Analyze your results and what your main takeaways were from this assignment <ul style="list-style-type: none"> <li>o 2 paragraphs long</li> </ul> </li> <li>• Discuss challenges and what you did to overcome them <ul style="list-style-type: none"> <li>o 1 paragraph</li> </ul> </li> <li>• Reflect on what you learned from this project and what you will incorporate from this project into future work <ul style="list-style-type: none"> <li>o 1 paragraph</li> </ul> </li> <li>• PDF format</li> <li>• One page maximum</li> </ul>

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for the rubric template. This structure is pulled from [Streifer & Palmer \(2020\)](#).