



CHESTLENS AI

Project Requirements Met/Not Met

Group 18

Supervisor	
Name: Email: Organization:	Dr. Mehdi Moradi moradm4@mcmaster.ca McMaster University, CAS
Team Member 1	
Name: Email: Student Number:	Abhishek Sharma shara109@mcmaster.ca 400322503
Team Member 2	
Name: Email: Student Number:	Anthony Vu vua11@mcmaster.ca 400306059
Team Member 3	
Name: Email: Student Number:	Hussein Saad saadh@mcmaster.ca 400307995
Team Member 4	
Name: Email: Student Number:	Nathan Starr starrn@mcmaster.ca 400323095
Team Member 5	
Name: Email: Student Number:	Yuvraj Jain jainy3@mcmaster.ca 400259484

P0 (Minimum Viable Product)

- The AI model should be able to identify the following 5 diseases (Atelectasis, Cardiomegaly, Consolidation, Edema, and Pleural Effusion) and No Finding.
- The application must be compatible with DICOM x-ray images.
 - This means that the application will allow uploads of DICOM images in the frontend and that the AI model in the backend must be able to process/work with DICOM x-ray images.
- The AI model for the application will use a pre-trained ML model. More details can be seen in [section 5.8](#) below.
- The AI model will be cross-trained across various datasets.
- The AI model testing requirements/methods are specified in [section 5.9](#) below.
- The expected area under the ROC curve for each disease is the following:
 - Expected area under ROC curve for Atelectasis: 0.8
 - Expected area under ROC curve for Cardiomegaly: 0.85
 - Expected area under ROC curve for Consolidation: 0.85
 - Expected area under ROC curve for Edema: 0.85
 - Expected area under ROC curve for No Finding: 0.85
 - Expected area under ROC curve for Pleural Effusion: 0.92
- The application should provide necessary information regarding the patient, including the patient's demographics, clinical/ medical history, the referring physician, and the assisting radiologist.
- The application should output a clear analysis of the x-ray by providing tags of diseases identified by the AI model.
- The front end of our application should be implemented as a website.
- The application should implement a simulated PACS server on Amazon Web Services.
- The application should provide the opportunity for radiologists to provide feedback or specific comments regarding the impressions and findings.

With the requirements specified above our project will have enough components to complete its main goal of aiding radiologists by identifying what diseases are present and the location of these diseases. This minimum viable product will allow radiologists to access the front end of our website online where they can see an analysis of x-ray images taken in the hospital and be aided in making diagnoses of Atelectasis, Cardiomegaly, Consolidation, Edema, No Finding, and Pleural Effusion with the use of disease tags identified by our AI model and provide any comments.

P1: The requirements given a priority of P1 are the next set of features that can be added to the minimal viable product to give the direct stakeholders an improved experience.

- The application should highlight/shade each region of the x-ray affected by identified disease(s).
 - This feature should effectively provide a visual mapping of areas of interest on the x-ray to better assist the radiologists.
- The application should provide additional information for identified disease(s). This additional information includes descriptions of abnormalities, disease specification, and relevant measurements.
- The AI model used in the application will be created from scratch and trained by us.

Our AUC

Finding	AUC Defined in P0	Validation AUC	Final Testing AUC
Atelectasis	0.80	0.8029	0.7969
Cardiomegaly	0.85	0.8130	0.8154
Consolidation	0.85	0.7951	0.7893
Edema	0.85	0.8861	0.8831
No Finding	0.85	0.8487	0.8483
Pleural Effusion	0.92	0.9175	0.9110

Requirements Not Met for P0-P1:

Yellow Highlighting: These requirements are mostly met.

P0: The AI model will be cross-trained across various datasets.

This requirement, while not explicitly written, was always intended to be interchangeable with P1: 'The AI model used in the application will be created from scratch and trained by us'. If this P1 requirement could not be met then we would fall back to the P0 requirement. Since we met the P1 requirement we did not cross-train the pre-trained model which was our original plan.

P0: The application should provide necessary information regarding the patient, including the patient's demographics, clinical/ medical history, the referring physician, and the assisting radiologist.

Our application provides the following patient information: patient's id, name, sex, birth date, and patient age at acquisition date. It also provides the x-ray information which includes view position and acquisition date, and previously uploaded x-rays scans. Information on further medical history and the referring physician and radiologist are not included because this information is not available in the DICOM header information which our application uses to get this information.

P0: The expected area under the ROC curve for each disease is the following:

To ensure the best result we experimented with the following training parameter combinations: optimizers (SGD vs Adam), different batch sizes (1, 8, 64) and different loss functions (BCEWithLogitsLoss vs CrossEntropyLoss). We tested these by primarily running through 10 epochs for 2 of our 6 training folders and then comparing the validation results for each. The combination we decided on was the Adam optimizer with a batch size of 64 and BCE with Logits Loss. Once we decided on our parameters we trained our model over the whole training set twice. The training set is broken into 6 folders and we trained each folder for 10 epochs each iteration through the dataset. At the end we had 20 epochs of our dataset. We used a varying learning rate as we trained.

Red Highlighting: These requirements are not-met:

P0: The application should implement a simulated PACS server on Amazon Web Services.

Since our focus was on training an AI model, creating a simulated PACS server was quickly identified as out of scope because it did not affect our ability to train our model or create a frontend website to host it. Image hosting for training was done on a department server and images for our frontend are stored in an online database thus making the simulated PACS server unnecessary.

P1: The application should provide additional information for identified disease(s). This additional information includes descriptions of abnormalities, disease specification, and relevant measurements.

In our P1 requirements we wanted to make sure our model provided additional information to help radiologists in identifying diseases beyond just a classification. We identified two ways of doing this and added both into P1 requirements. One of these methods is the red requirement above and the other is P1: 'The application should highlight/shade each region of the x-ray affected by identified disease(s). This feature should effectively provide a visual mapping of areas of interest on the x-ray to better assist the radiologists.' We decided to focus on the visual aspect of providing additional information for diseases as it worked better with our use of the MIMIC dataset allowing for the training of the highlighting feature alongside training our disease predictions without requiring additional tags not in our dataset. Since we were able to implement highlighting to provide radiologists with additional information beyond disease classification we decided not to pursue the alternative requirement for additional information.

Percent Complete: Out of our 14 requirements only two are Red and considered not met so we met 86% of our requirements.