Predicting Pulmonary Edema Using Convolutional Neural Networks with Lung and Heart Image Segmentation

Davila-Garcia, David [1]; Potdar, Yash [1]; Morocho, Marco [1]

ddavilag@ucsd.edu, ypotdar@ucsd.edu, mmorocho@ucsd.edu

[1] University of California, San Diego

UC San Diego Health

UC San Diego HALICIOĞLU DATA SCIENCE INSTITUTE

Introduction

Convolutional neural networks (CNNs) have shown promising results in identifying diseases from medical images. Justin Huynh et al. demonstrated the use of CNNs in diagnosing cardiogenic pulmonary edema by training them on chest radiographs and the clinical serum biomarker NT-proBNP. However, recent literature suggests that NT-proBNP concentrations can be influenced by confounding factors such as renal failure, age, sex, and body mass index (BMI). Additionally, segmentation of the original radiographs to isolate the heart and lungs has been shown to improve the performance of CNN models. We hypothesize that the addition of segmentation or clinical data will improve the performance of the CPE classifier.

Methods

In this project, we investigated the impact of confounding factors, including renal function and BMI, and clinical findings of pneumonia and acute heart failure, on the accuracy of CNN models for diagnosing CPE. We utilized Creatinine as a clinical serum biomarker to measure renal function. Additionally, we incorporated the use of lung and heart segmentation to isolate the relevant structures. Our dataset consisted of chest radiographs and NT-proBNP data from 16,619 patients from UCSD Health.

We trained four Modified-ResNet152 CNN architectures using an 80%/10%/10% train-validation-test split with differing inputs: (A) Original Radiographs only, (B) Original Radiographs + Clinical Data, (C) Original Radiographs + Heart/Lung Segmentation and (D) Original Radiographs + Heart/Lung Segmentation + Clinical Data. Each model was trained for 20 epochs (n = 13,295) using the mean absolute error loss function and Adam Optimizer with a learning rate of 0.001. After training, each model was selected based on the lowest mean absolute error on an unseen validation set (n = 1,662), and model performance was evaluated by calculating the accuracy and AUC of each model on an unseen test set (n = 1,662).

Lung Segmentation Network

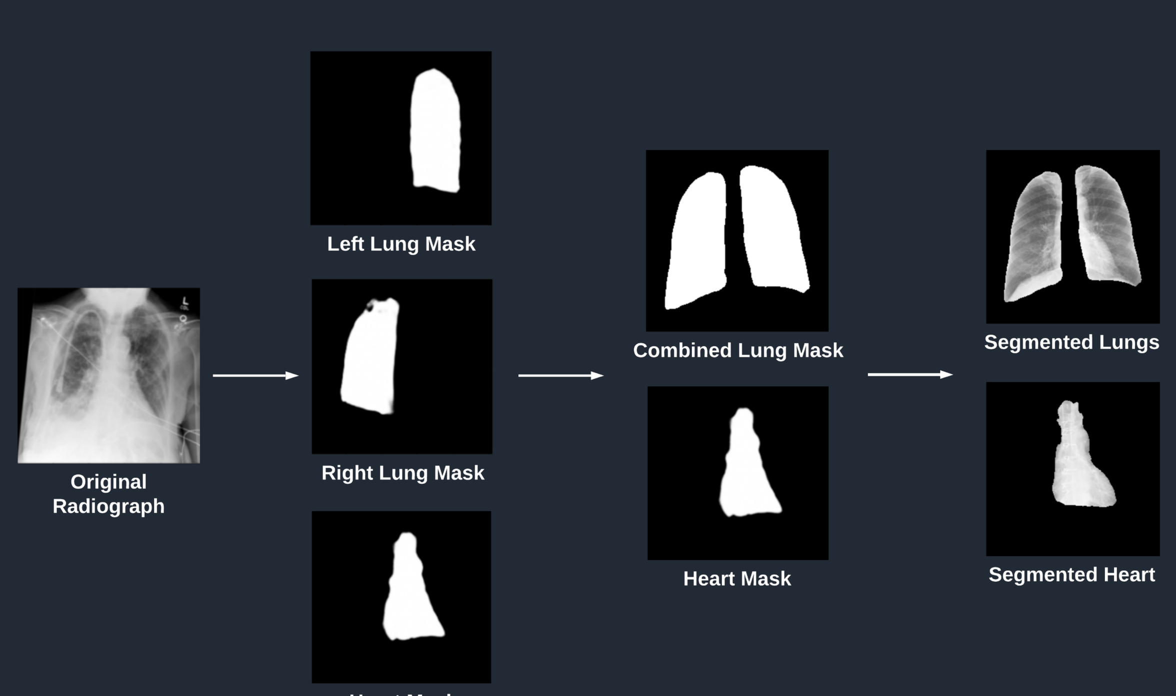
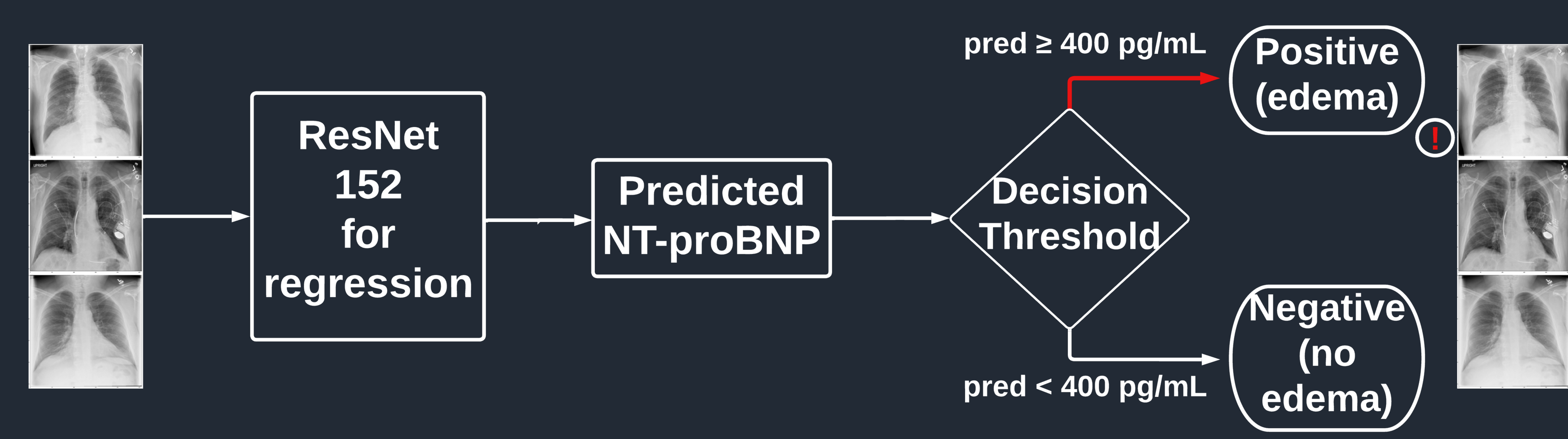


Figure 1. Heart and Lung Segmentation Diagram

Model (A): Original Images Architecture



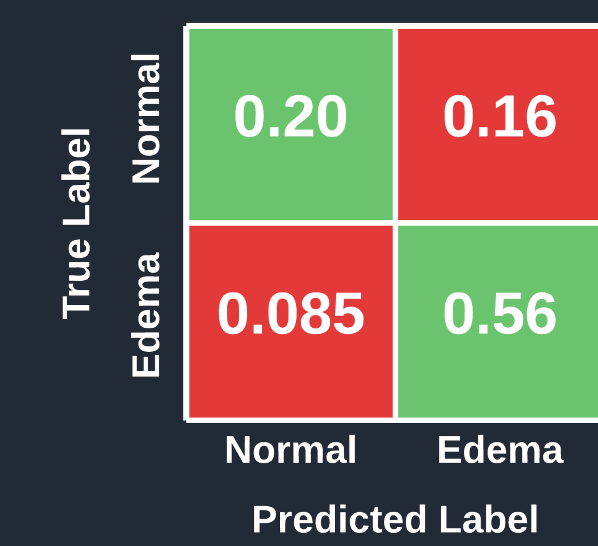
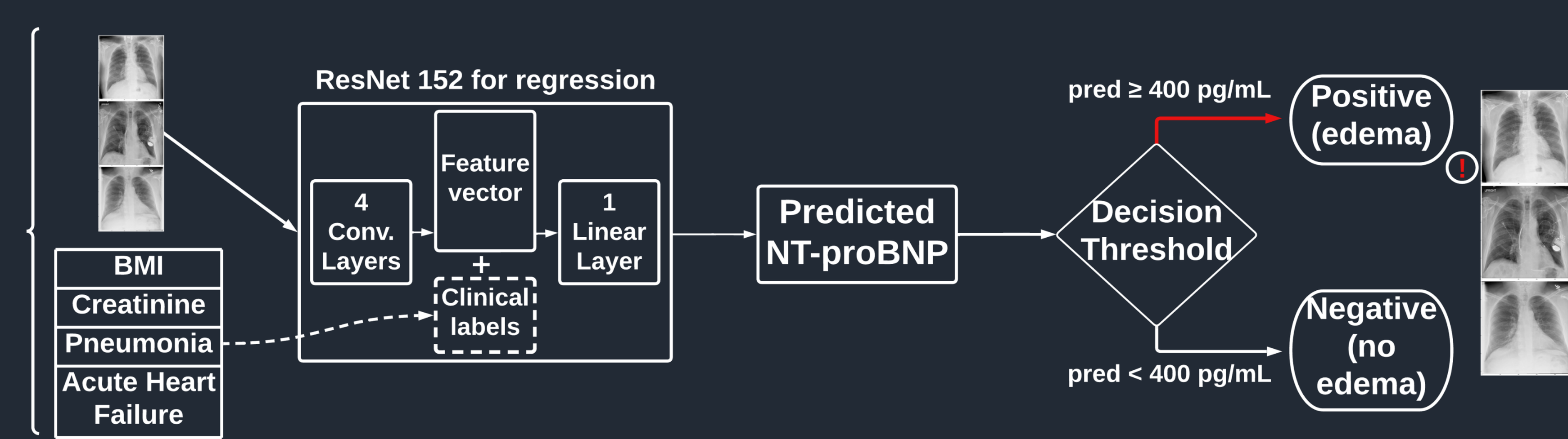


Figure 2A. Model (A) Confusion Matrix

Model (B): Original Images with 4 Clinical Labels Architecture



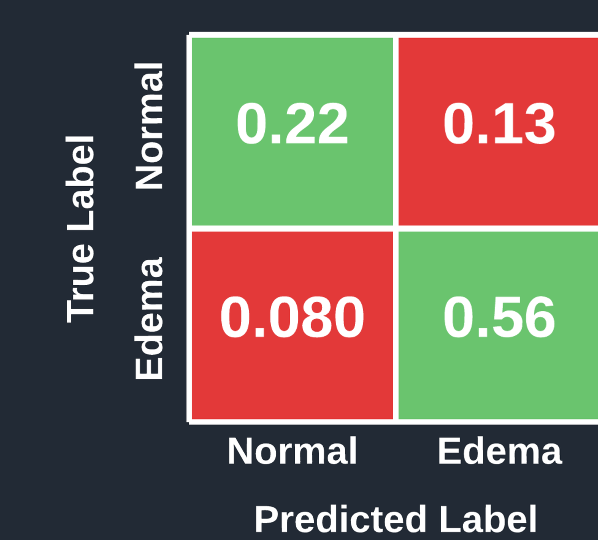
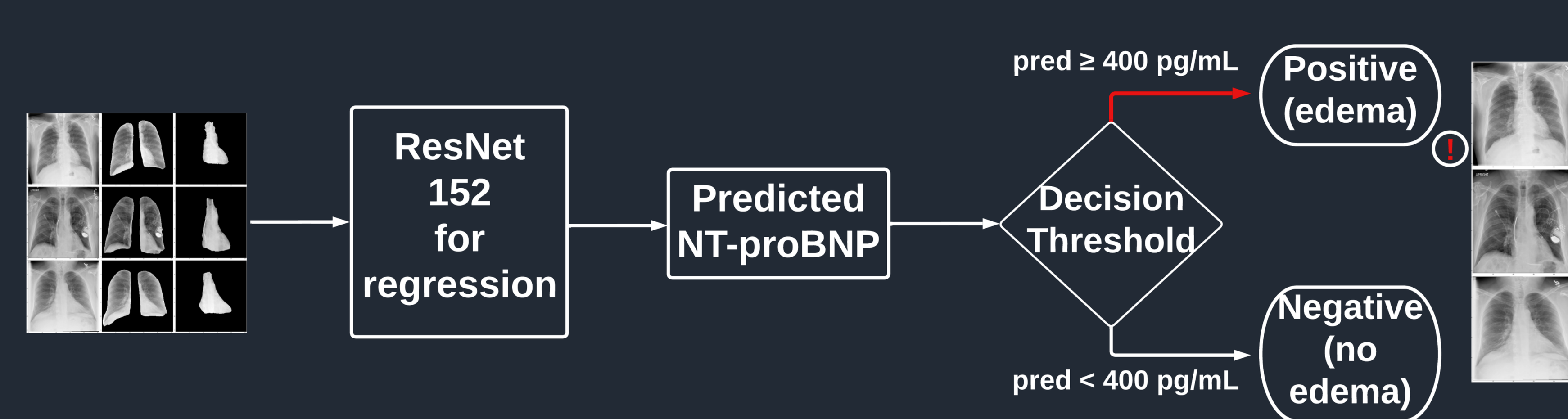


Figure 2B. Model (B) Confusion Matrix

Model (C): Segmented Images Architecture



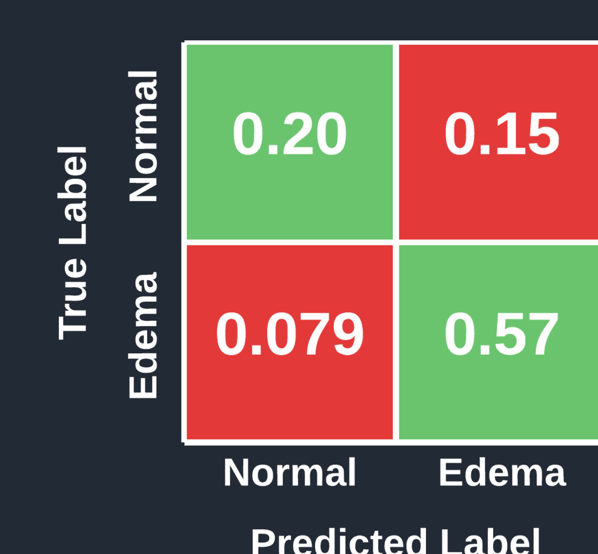
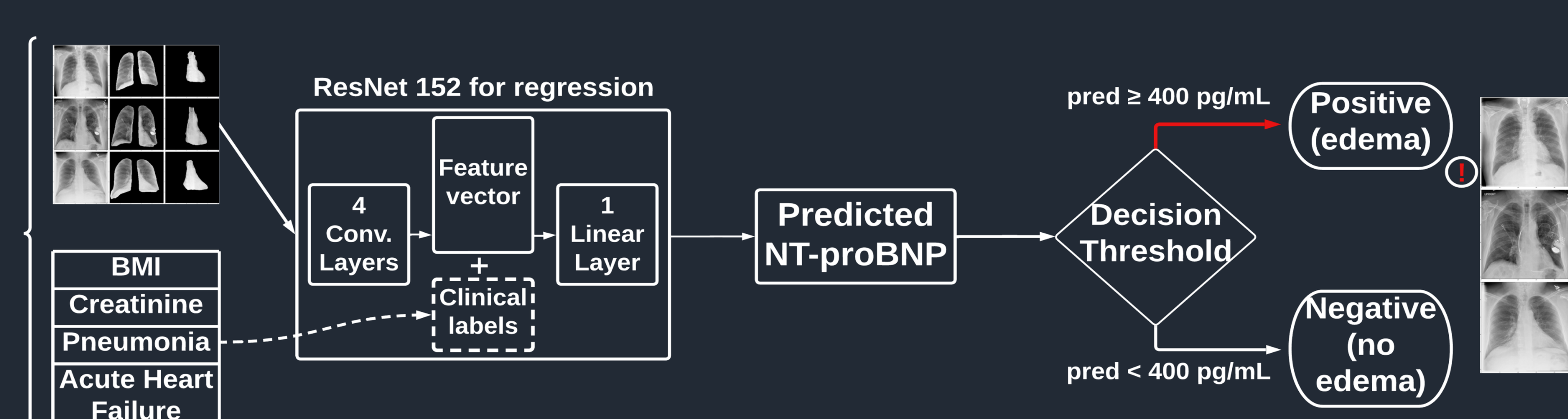


Figure 2C. Model (C) Confusion Matrix

Model (D): Segmented Images with 4 Clinical Labels Architecture



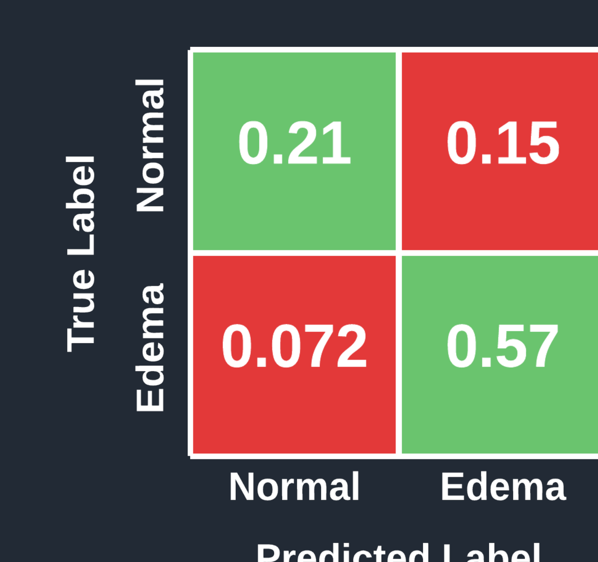


Figure 2D. Model (D) Confusion Matrix

Results

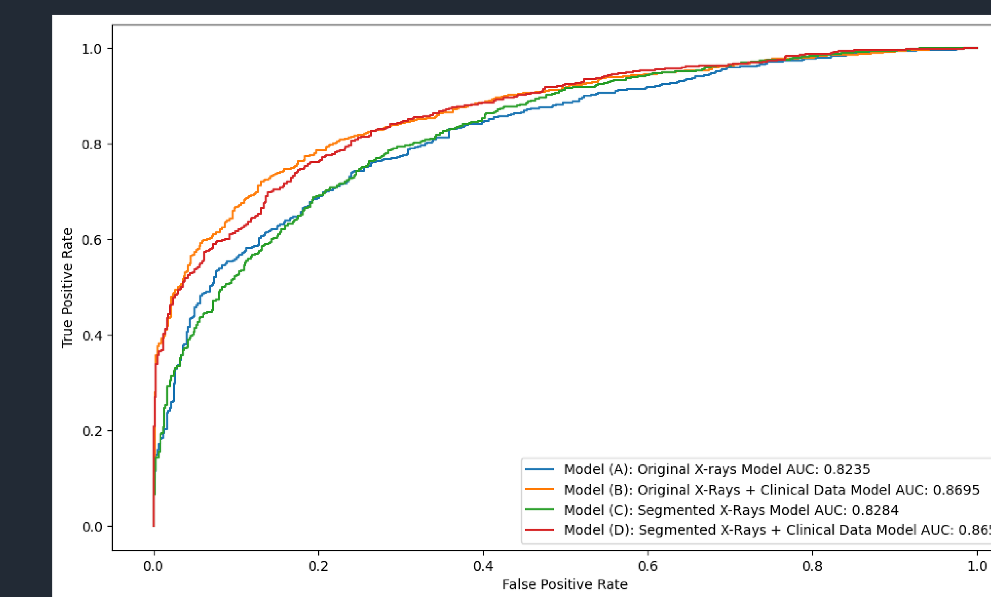


Figure 3. ROC Curves for each ResNet152 model by Input Data

Table 1: ResNet152 Model Performance by Input Data

Input Data	Train L1-loss	Test L1-loss	Accuracy	AUC	Pearson R
Model A: Original X-rays	0.442	0.531	0.756	0.824	0.646
Model B: Original X-rays with Clinical Data	**0.392**	**0.455**	**0.787**	**0.869**	**0.738**
Model C: Original X-rays, Lung & Heart Segmentations	0.518	0.514	0.768	0.828	0.656
Model D: Original X-rays, Lung & Heart Segmentations, with Clinical Data	0.428	0.468	0.783	0.866	**0.738**

Model (B), which used the original radiographs and clinical data as inputs, achieved the highest accuracy of 0.787 and AUC of 0.869 on the test set. Model (D), which used all inputs, performed marginally worse than Model (B), with an accuracy of 0.783 and AUC of 0.866 on the test set.

Conclusion

Our project demonstrates the importance of considering confounding factors, other clinical data, and image segmentation when training CNN models to diagnose CPE from chest radiographs. We found that incorporating clinical data improved CNN model performance. However, our results did not indicate a significant improvement in model performance with the addition of heart and lung segmentation. Further investigation is needed to determine the optimal use of image segmentation in CNN models.

Acknowledgments

Credits to our mentor Dr. Albert Hsiao (alhsiao@alumni.ucsd.edu) for his guidance, and Amin Mahmoodi for providing the segmentation network.

References

[1] Huynh, J. et al., "Deep Learning Radiographic Assessment of Pulmonary Edema: Optimizing Clinical Performance, Training With Serum Biomarkers,"

[2] Chen, H. H. et al., "Natriuretic peptide measurement in heart failure."

[3] Welsh, P., Campbell et al., "Reference Ranges for NT-proBNP (N-Terminal Pro-B-Type Natriuretic Peptide) and Risk Factors for Higher NT-proBNP Concentrations in a Large General Population Cohort."

[4] Sofat, S. et al., "Lung Field Segmentation in chest radiographs: A historical review ..."