

Project Proposal

Team Name: Group 33

Abhishek Tummalapalli - 1066956

Fangzhou Wang -1339824

Sayan Bachhar - 1413063

Yusuf Berdan Güzel - 1051639

1 Research Objective

Medical research has shown that there is a strong association between pneumonia and worse outcomes, stating that patients with pneumonia are associated with an increased risk of two-to-four-fold for heart failure hospitalisations and cardiovascular mortality (Shen, Jhund, Anand, et al., 2021).

This project aims to answer the following research question:

How likely are adult ICU patients diagnosed with pneumonia to develop heart failure within six months of their initial ICU admission?

This is to help with the development of better preventive care and overall improvement of patient outcomes. This requires us to investigate the factors connecting the two conditions, thus catching it at an early stage.

2 Data Source & Phenotyping

We will use the MIMIC-IV database, a repository of de-identified health records from ICU patients between 2008 and 2019. This dataset includes clinical data such as demographics, diagnoses, vital signs and laboratory results.

We will begin by retrieving data using ICD-10 codes and then pre-process it to address missing values and standardise formats. Relevant variables include patient demographics (e.g., age, sex), co-morbid conditions (e.g., chronic obstructive pulmonary disease, diabetes, prior myocardial infarction), and specific laboratory results like arterial blood gases and complete blood counts (Zhang et al., 2023).

We will also include clinical variables linked to cardiovascular complications in pneumonia patients, such as elevated blood urea nitrogen (BUN) levels and low oxygen saturation (Desai et

al., 2022). Patients with BUN levels ≥ 30 mg/dL and oxygen saturation below 90% are at higher risk of heart failure.

Additional biomarkers include elevated B-type natriuretic peptide (BNP) levels, a predictor of heart failure, and troponin levels, which indicate myocardial injury (Desai et al., 2022). We will assess acute inflammatory responses using markers like C-reactive protein (CRP) and interleukins (e.g., IL-6, IL-10), which are associated with both acute myocardial infarction and heart failure in pneumonia patients.

3 Methodology

3.0 Steps

Section	Details
Data Pre-processing	Standardise and normalise the data as required, address missing values and outliers and manage class imbalance.
Feature Selection	Recursive Feature Elimination (RFE) to select the most predictive features.
Model Selection	Start with a baseline model, Logistic Regression with regularisation. Implement Decision Trees. Explore more complex models: Random Forests and AdaBoost. Conduct hyperparameter tuning to optimise model performance.
Model Evaluation	In addition to traditional metrics like F1-Score, use the Net Reclassification Index (NRI) and Integrated Discrimination Improvement (IDI) to assess improvements in risk prediction. Perform a detailed analysis by patient subgroups, such as age and comorbidities, to ensure the model performs well across all relevant patient demographics.
Results Visualisation	Use interactive dashboards (Jupyter Dashboards) to visualise the model's predictions and feature importance. Provide insights based on the analysis, highlighting the most significant risk phenotypes.

3.1 Performance Metrics

- Net Reclassification Index (NRI): This metric will measure the improvement in risk prediction provided by the model.
- Integrated Discrimination Improvement (IDI): This metric will evaluate the improvement in sensitivity and specificity when comparing the new model to a baseline model.
- Calibration curves (AUC): These curves will assess whether the model underestimates or overestimates the risk of heart failure by comparing predicted and actual probabilities.

3.2 Expected Outcomes

- We aim to develop a model that measures the risk of heart failure due to pneumonia by identifying risk phenotypes in advance.
- We will provide a tool that healthcare professionals can use to identify pneumonia patients at risk of heart failure, enabling early intervention and potentially improving patient outcomes.
- We anticipate optimised management of these patients, potentially leading to reduced incidences of heart failure.

References

Shen, L., Jhund, P., Anand, I., et al. (2021). Incidence and outcomes of pneumonia in patients with heart failure. *Journal of the American College of Cardiology*, 77(16), 1961–1973. <https://doi.org/10.1016/j.jacc.2021.03.001>

Desai, A., Aliberti, S., Amati, F., Stainer, A., & Voza, A. (2022). Cardiovascular complications in community-acquired pneumonia. *Microorganisms*, 10(11), 2177.

Zhang, Y., Peng, Y., Zhang, W., & Deng, W. (2023). Development and validation of a predictive model for 30-day mortality in patients with severe community-acquired pneumonia in intensive care units. *Frontiers in Medicine (Lausanne)*, 10, 1295423.

Steyerberg, E. W., Eijkemans, M. J. C., Harrell, F. E., & Habbema, J. D. F. (2000). Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in Medicine*, 19 (8), 1059-1079.

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115 (7), 928-935.

Pencina, M. J., D'Agostino, R. B., D'Agostino, R. B., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, 27 (2), 157-172.

Corrales-Medina, V. F., Musher, D. M., Wells, G. A., Chirinos, J. A., Chen, L., & Fine, M. J. (2012). Cardiac complications in patients with community-acquired pneumonia: incidence, timing, risk factors, and association with short-term mortality. *Circulation*, 125 (6), 773-781.