Uncovering Health Patterns in Severe Pneumonia and COVID-19 Using Tensor Decomposition

Ayana Price[1], Andrew Ramirez[2], Jackson Chin[2], Aaron Meyer[2]
1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Department of Bioengineering, Samueli School of Engineering, UCLA

Pneumonia Context: COVID-19 vs. Non-COVID-19

- **Global Pneumonia Threat:** Pneumonia remains a leading global health concern with high morbidity and mortality rates. The COVID-19 pandemic has amplified the need to focus on both viral and bacterial pneumonia.
- **Demographic Trends and Epidemiological Shifts:** Post-COVID-19, shifts in public health measures have altered pneumonia patterns, particularly in Europe and Asia. These changes have disproportionately impacted vulnerable demographics, such as older adults and individuals with higher BMI, increasing their risk of severe outcomes and mortality.
- **Comparative Studies:** The significance of this study lies in its potential to uncover how pneumonia manifests differently in COVID-19 patients compared to non-COVID cases, identifying pathogen-specific differences.

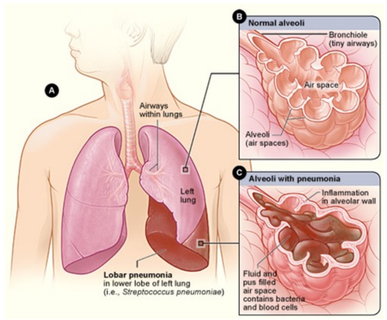


Fig. 1 Comparison of Normal and Pneumonia-Affected Alveoli in the Lungs: This image highlights how inflammation and fluid accumulation in the lungs can disrupt normal respiratory function.

Objectives

- **Objective 1:** Identify distinct biological patterns in both COVID-19 and non-COVID-19 pneumonia patients
- **Objective 2:** Correlate these patterns with clinical features such as age, BMI, and BAL cell type percentages to predict mortality

PARAFAC2

•Tensor decomposition method for high-dimensional data like scRNA-seq, preserving the structure and relationships to uncover complex patterns missed by traditional methods
•Applied to biological signatures from patient samples, yielding factors that capture key relationships among genes, patients, and cells

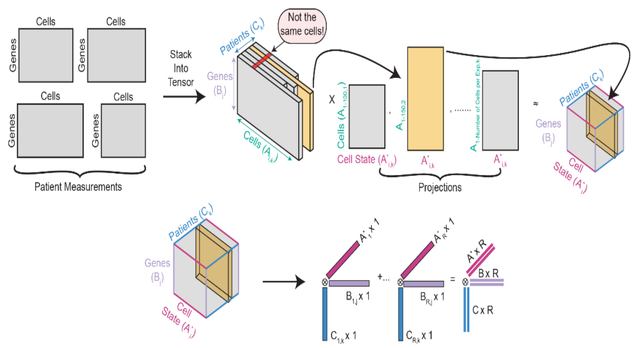


Fig. 2 Schematic of PARAFAC2 Tensor Decomposition: Patient measurements are stacked into a tensor that integrates genes, cells, and patients. The decomposition reveals key factors, allowing projections that capture the relationships between cell states and clinical parameters across different patient groups.

Data Collection

- Sampled 153 severely ill ICU patients from Northwestern, including 91 COVID-19 and 62 non-COVID-19 pneumonia patients
- Single-cell measurements collected via bronchoalveolar lavage (BAL)
- Clinical data including demographics

Methods

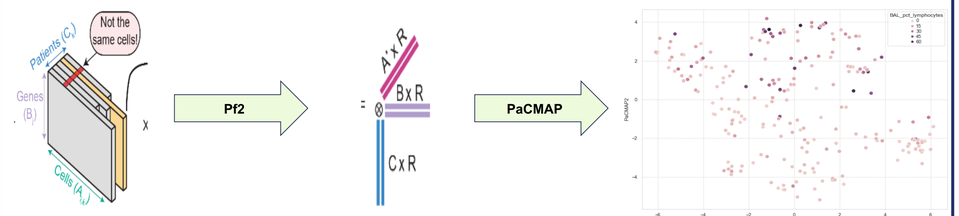


Fig. 3 PaCMAP of Patient Factor (C x R): Each feature was reduced to a two-dimensional space using PaCMAP/UMAP. This visualization highlights the association of BAL percentage of lymphocytes with the identified clusters, revealing potential patterns in immune response across different patient groups.

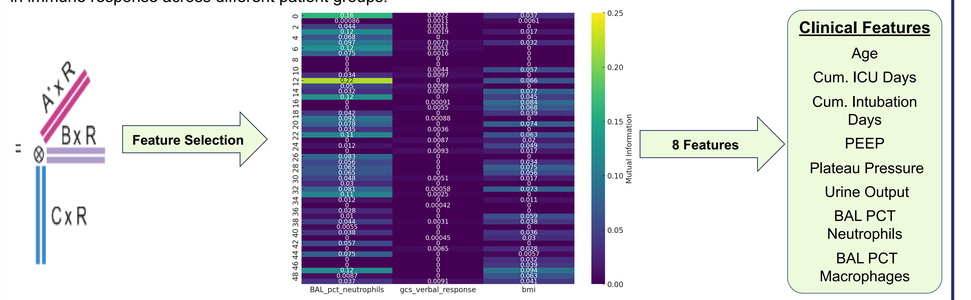


Fig. 4 Mutual Information Score Heatmap: Illustrates the strength of the relationship between clinical features and the components identified through tens or decomposition. With 62 clinical features available, it was crucial to select those with the most relevance to ensure accurate and meaningful results.

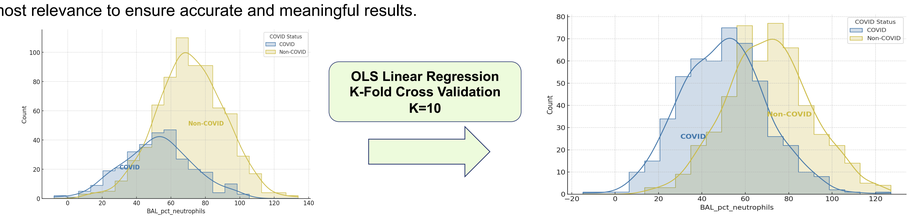


Fig. 5 Under Sampling & Oversampling: The distribution of each feature chosen differs between COVID and non COVID patients. In this graph, neutrophils for non-COVID patients show a bimodal distribution with peaks around 60-80%. While COVID patients have a more consistent distribution.

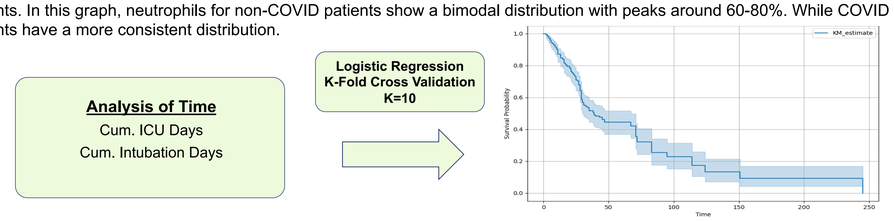


Fig. 6 Kaplan-Meier Survival Curve: This figure shows the survival probability of patients over time. The data was refined allowing only the most recent sample for each patient to be used before running the regression model. The x-axis represents a combination of cumulative intubation and ICU days. The curve illustrates how survival probability decreases as the duration of critical care increases, highlighting the significant impact that prolonged intubation and ICU stays have on patient outcomes.

References

- Edouard, S., Boughammoura, H., Colson, P., La Scola, B., Fournier, P., & Fenollar, F. (2024). Large-Scale Outbreak of Mycoplasma pneumoniae Infection, Marseille, France, 2023–2024. *Emerging Infectious Diseases*, *30*(7), 1481-1484. https://doi.org/10.3201/eid3007.240315
- Rabanser, S., Shchur, O., & Günnemann, S. (2017). Introduction to tensor decompositions and their applications in machine learning. *arXiv*. https://arxiv.org/abs/1711.10781
- Gao, C. A., Markov, N. S., Stoeger, T., Pawlowski, A., Kang, M., Nannapaneni, P., Grant, R. A., Pickens, C., Walter, J. M., Kruser, J. M., Rasmussen, L., Schneider, D., Starren, J., Donnelly, H. K., Donayre, A., Luo, Y., Budinger, G. S., Wunderink, R. G., Misharin, A. V., & Singer, B. D. (2023). Machine learning links unresolving secondary pneumonia to mortality in patients with severe pneumonia, including COVID-19. *Journal of Clinical Investigation*, *133*(12). https://doi.org/10.1172/jci170682

Results

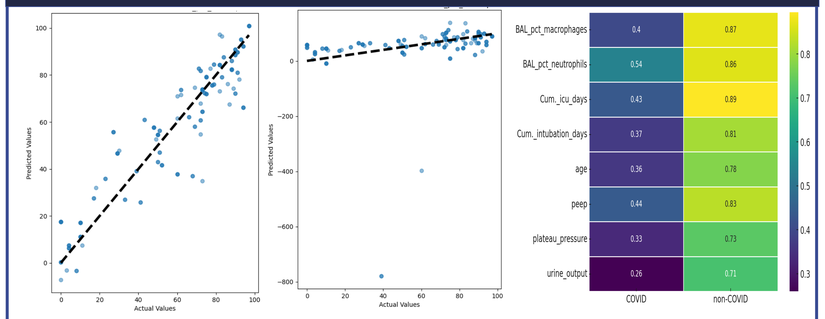
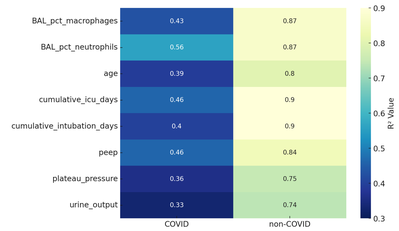


Fig. 7 OLS Linear Regression Model & Metrics: The scatterplots display the regression model of BAL percentage neutrophils, comparing COVID and non-COVID patients. The left graph illustrates the model's better performance in non-COVID patients, while the middle graph demonstrates the model's lower performance in COVID patients. The heatmap illustrates the R^2 values for all eight features highlighting differences in their predictive power.

Fig. 8 K-Fold Metrics: This heatmap illustrates the consistency and strength of associations, with higher R^2 values in the non-COVID group. The input for the linear regression model were the patient factors derived from the Pf2 analysis. All eight features from the metadata, highlight the model's performance between COVID and non-COVID.



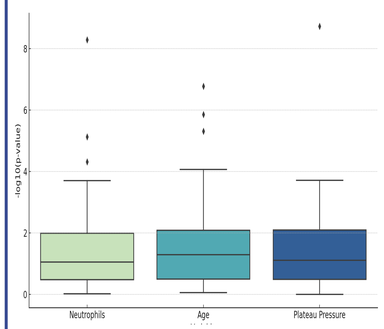


Fig. 9 Boxplot of P-Values: Illustrates the statistical significance of associations between selected variables and the outcome of interest. The input for this analysis were the patient factors obtained from the Pf2 model, which were then used in the regression model to determine the p-values. Neutrophils show greater variability in p-values compared to age and plateau pressure. This variability indicates a heterogeneous response among patients, highlighting the diverse role neutrophils may play in influencing clinical outcomes.

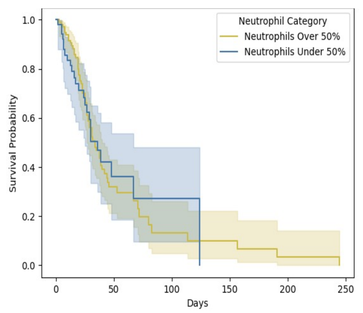


Fig. 10 Kaplan-Meier Survival Curve: Illustrates patients with neutrophil levels over 50% exhibit a lower survival probability over time compared to those with neutrophil levels under 50% . The shaded areas represent the confidence intervals for each group, highlighting the variability in survival outcomes. This suggests that higher neutrophil percentages are associated with poorer survival outcomes during extended ICU stays.

Conclusion

Our analysis reveals all eight clinical parameters used not only correlate with mortality but also align with specific biological pathways involved in pneumonia. By linking these clinical features to underlying biological mechanisms, our findings offer a more nuanced understanding of how these factors drive disease progression and impact patient outcomes.

Acknowledgements

I would like to thank the UCLA Bruins in Genomics Program for providing this valuable research opportunity. I am also deeply thankful to the Meyer Lab, particularly my mentors, Andrew Ramirez, Jackson Chin, and Dr. Aaron Meyer, for their unwavering dedication, guidance, and support throughout every stage of this project.