# EHR Use and Academic Performance Among General Surgery Residents: An Exploratory Analysis Using Audit Logs

Catherine G. Pratt, MD[1], Roxy Huang, MSIS[2], Suguna Kotte, PharmD[2], Ralph C. Quillin III, MD[1], Danny T.Y. Wu, PhD, MSI[2,1]

[1]University of Cincinnati College of Medicine, Cincinnati, OH; [2]University of North Carolina, Chapel Hill, NC

## Introduction

- Surgical residents' engagement with the Electronic Health Record (EHR) system plays a critical role in their clinical performance and learning.
- However, EHR access patterns vary greatly among residents and are rarely analyzed at scale.
- This project aims to explore whether EHR access metrics can meaningfully cluster residents and predict their overall performance scores.

## Objective

- Identify behavioral patterns in EHR usage through clustering.
- Predict residents' performance using machine learning models.
- Evaluate model performance using AUC and regression metrics.

## Method 1 – Clustering (No Label)

**Data Collection**

- 28 surgical residents
- Over 100,000 EHR access logs
- Extracted 42 time-based features per resident: access_hour, access_week, weekend_activity, offhour_activity, etc.

**Data Processing**

- Aggregated logs per resident
- Scaled features using StandardScaler

**Data Analysis**

- Residents grouped into 4 clusters
- Hierarchical Clustering
- KMeans Clustering (k=4 chosen via elbow method)
- Cluster labels used as behavioral profiles

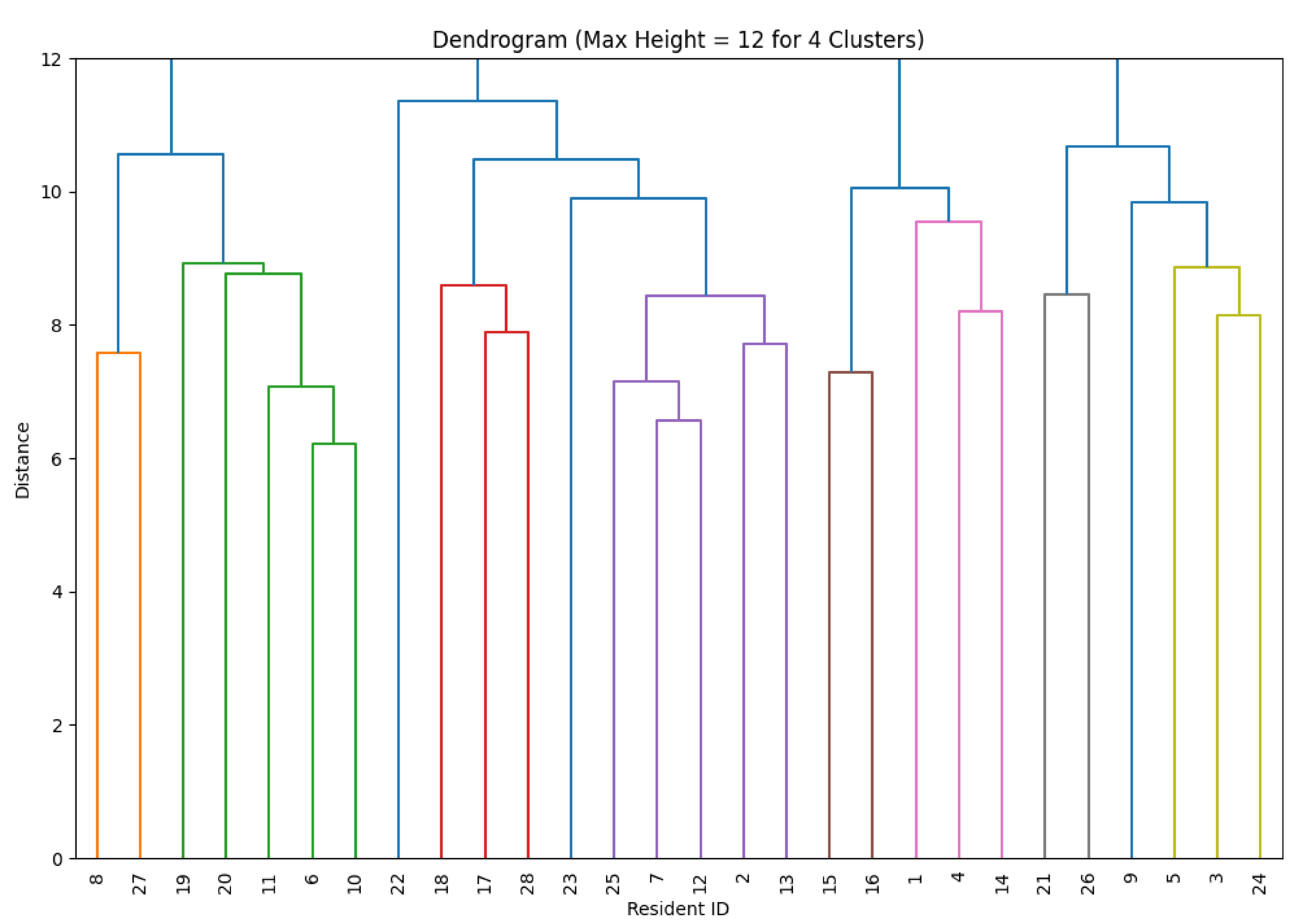


**Figure 1.** Resident Clustering Dendrogram Based on 42 EHR Access Metrics

## Results 1

- Hierarchical clustering with Ward linkage and Euclidean distance and 4 clusters performed the best, achieving a Silhouette score of 0.584 and a more balanced cluster size.
- The figure on the left shows the clustering results. The median score of the clusters 1 to 4 was -2, -2, 1, and 1, respectively.
- However, there was no statistical significance (p=.027).
- The Chi-square tests found no statistical significance either between scores and the 4 clusters ($X^2 = 8.38$, df = 6, p = 0.2119).

## Method 2 – Classification

**Performance Labeling**

- Total Performance Score provided by faculty (range: -5 to +5)
- Top 25% labeled as **"High performers"**, rest as **"Low performers"**

**Modeling**

- **Logistic Regression** used to classify high vs low performers
- Compared **L1 (Lasso), L2 (Ridge),** and **ElasticNet (L1+L2)**
- 3-fold **Stratified Cross-Validation**
- Repeated testing 100 times
- AUPRC (precision-recall) as primary evaluation metric due to imbalanced data

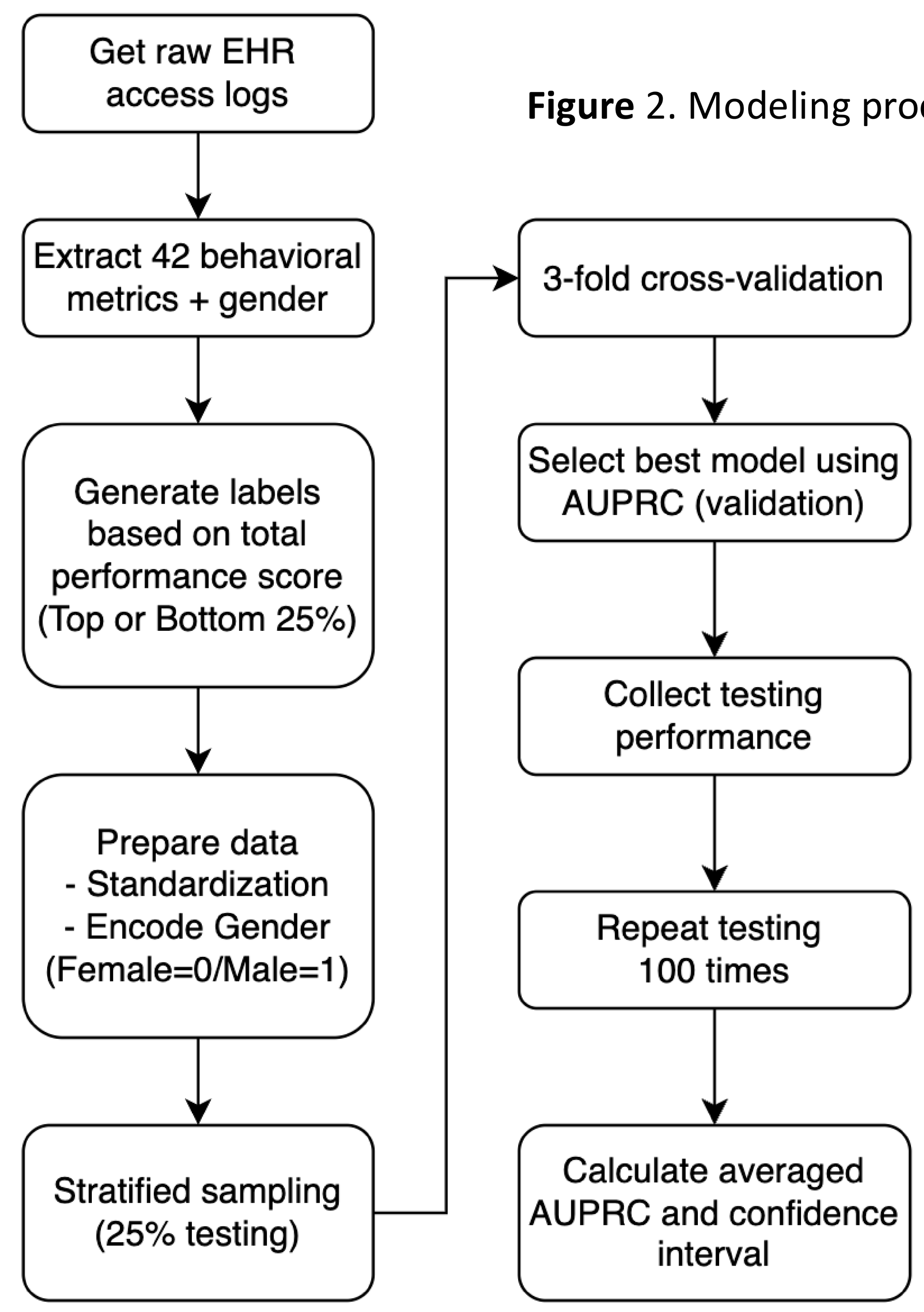


**Figure** 2. Modeling process.

## Results 2

| Target | Algorithm | Validation AUPRC | 95% CF | Testing AUPRC | 95% CF |
|---|---|---|---|---|---|
| Top 8 | LR-L1 | 0.397 | (0.378, 0.416) | 0.380 | (0.346, 0.413) |
| Top 8 | LR-L2 | 0.408 | (0.385, 0.431) | 0.407 | (0.368, 0.447) |
| Top 8 | LR-Elastic | 0.387 | (0.369, 0.405) | 0.354 | (0.326, 0.381) |
| Bottom 8 | LR-L1 | 0.468 | (0.445, 0.491) | **0.485** | (0.444, 0.527) |
| Bottom 8 | LR-L2 | 0.456 | (0.430, 0.481) | 0.457 | (0.418, 0.495) |
| Bottom 8 | LR-Elastic | **0.482** | (0.455, 0.508) | 0.477 | (0.435, 0.519) |

## Conclusion

- Clustering results did not correlate with performance scores.
- Classification using only EHR audit logs achieved moderate performance.
- The model performed better when when predicting low-performance residents than high-performance residents.
- Future work: retrain model with baseline scores and more detailed workflow patterns.