Breakout Session 8: Track A

Applying Gerchberg-Saxton Algorithm on Biomedical Data to Mitigate Sampling Bias on Under-Represented Populations

Mr. Seha Ay
Graduate Student, Wake Forest School of Medicine

Developing Unbiased AI/Deep Learning Pipelines to Address Lung Cancer Health Disparities Research

Applying Gerchberg-Saxton Algorithm on Biomedical Data to Mitigate Sampling Bias on Underrepresented Populations

Seha Ay (Speaker)
Graduate Research Associate
Wake Forest School of Medicine
Biomedical Engineering PhD Program

Wei Zhang (PI)
Hanes and Willis Family Professor in Cancer, Director
Cancer Genomics and Precision Oncology
Wake Forest Baptist Comprehensive Cancer Center

Agenda

1. Project Motivation
2. Gerchberg-Saxton Algorithm
    - Previous Studies and Results
3. Project Plan
4. Current Phase and Initial Results
5. Future Directions and Expected Outcomes

Project Motivation - Challenges in Medical AI

- **PHI data regulations** poses challenges on data distributions.
- Upholding **data integrity and privacy** is paramount yet challenging.
- There is a noticeable scarcity in the **availability of public medical datasets**.

Project Motivation - Data Quality and Representation

- The imperative to improve **the quality of data generated** by individual institutions is clear.
- The **distribution among population** groups in the US is **uneven**, which is mirrored in medical data, potentially biasing AI model predictions.
- Existing AI-based solutions, such as synthetic data generation, face challenges like **inadequate source data quality or quantity**.

U.S. Population by Race and Hispanic Origin, 2019

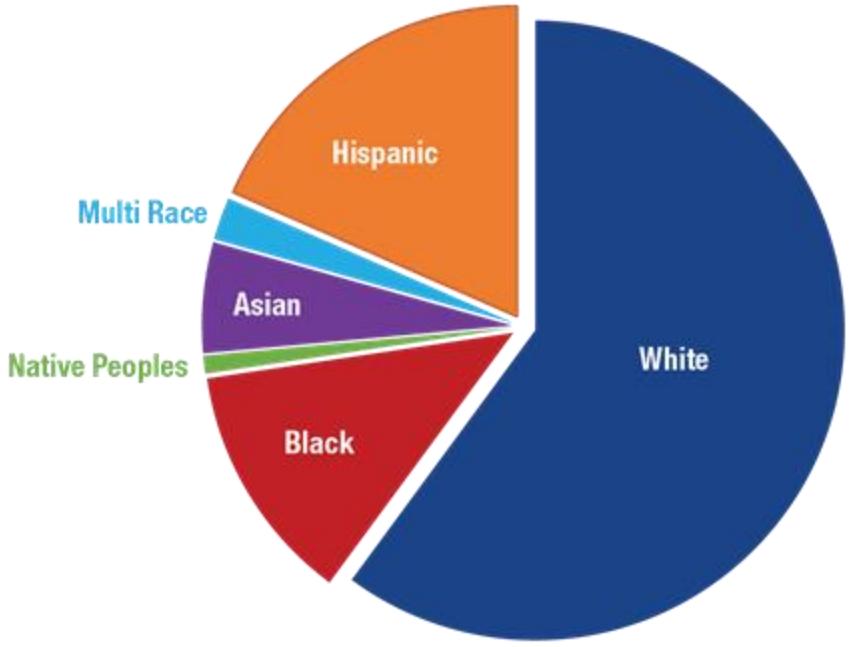


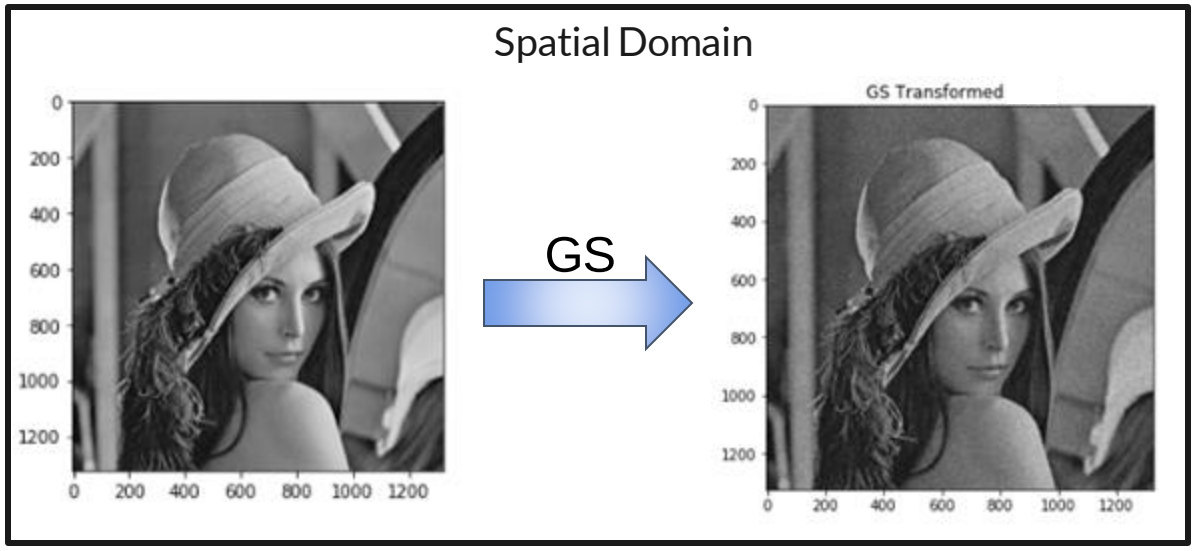
Image Source: U.S Census Bureau, Population Estimates, 2019

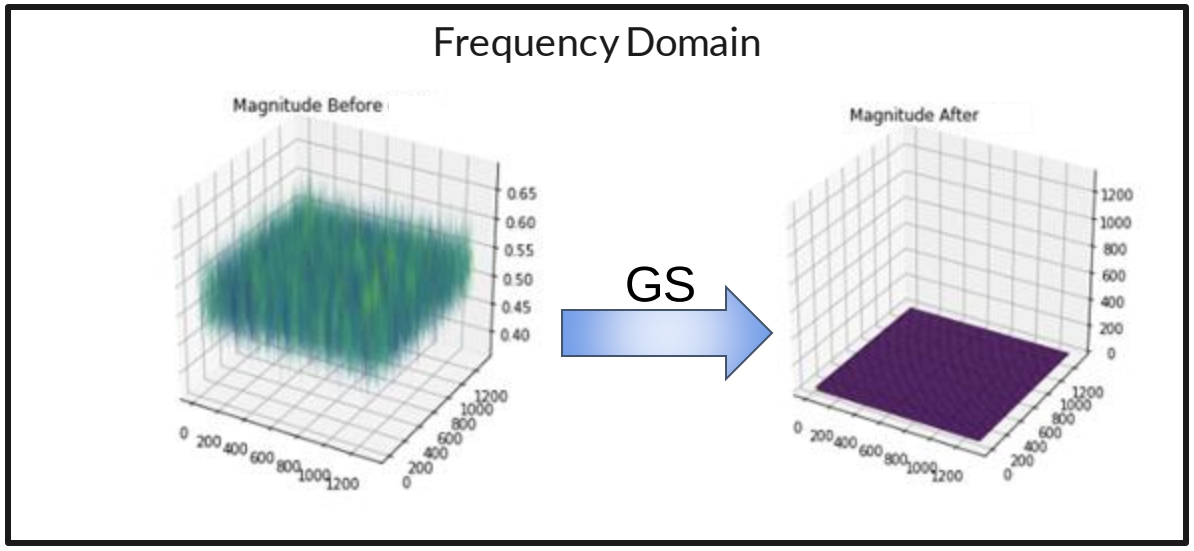
Project Motivation - Innovative Solution

- We propose an innovative data transformation approach utilizing the **Gerchberg-Saxton algorithm** to address data quality and representation issues.
- This algorithm transforms data in the frequency domain, balancing intensity components while preserving phase information, thereby **enhancing data uniformity**.
- The transformed data will more uniformly represent each population group, aiming for **fairer results in machine learning applications**.

Introduction to Gerchberg-Saxton

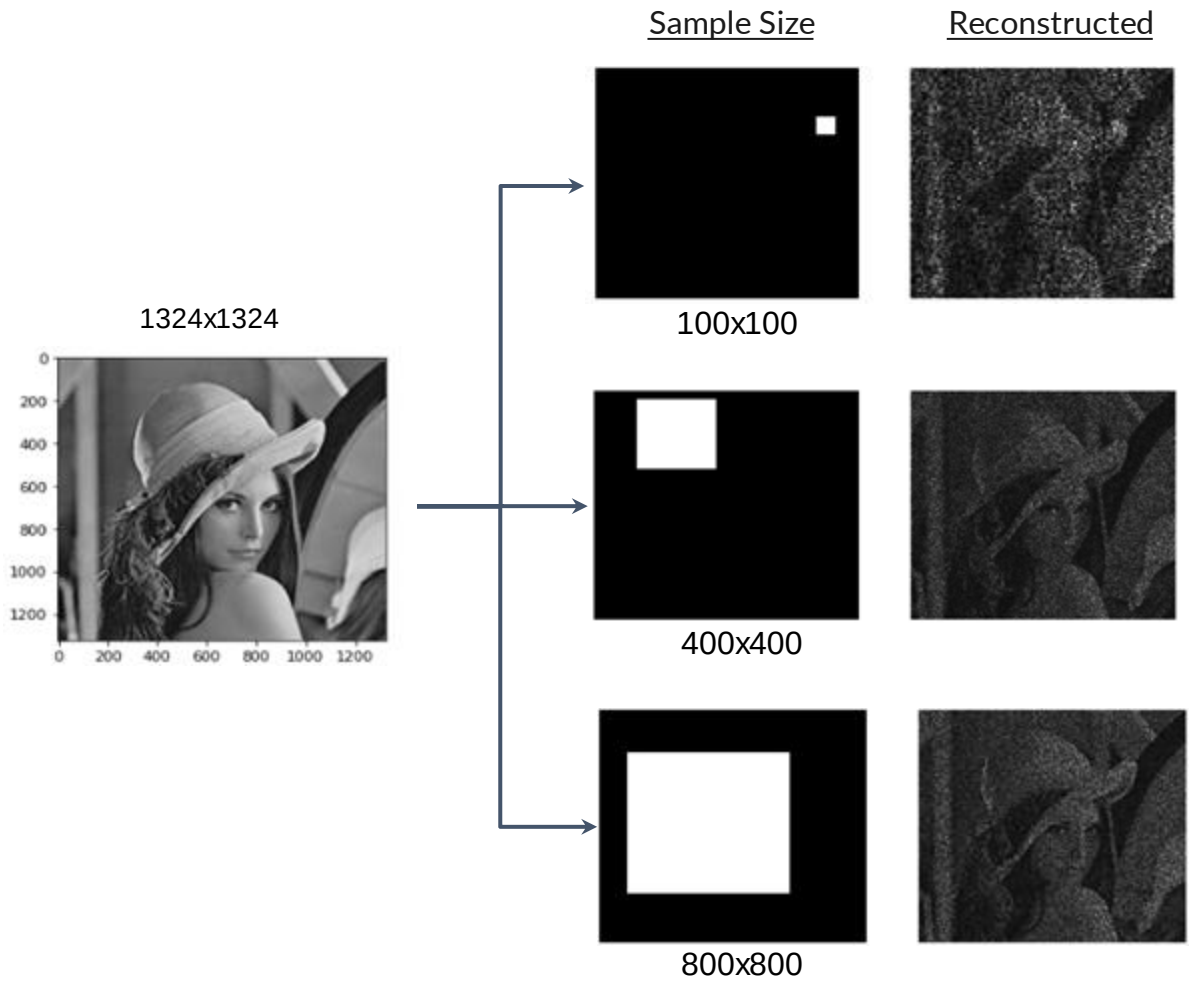
- The Gerchberg-Saxton algorithm is a cornerstone in holographic imaging, enabling the creation of **holographic representations of images in digital environments**.
- We leverage the Gerchberg-Saxton algorithm to utilize two crucial **characteristics of holographic images** to enhance medical data analysis.
  - *Holographic Divisibility*
  - *Information Distribution*





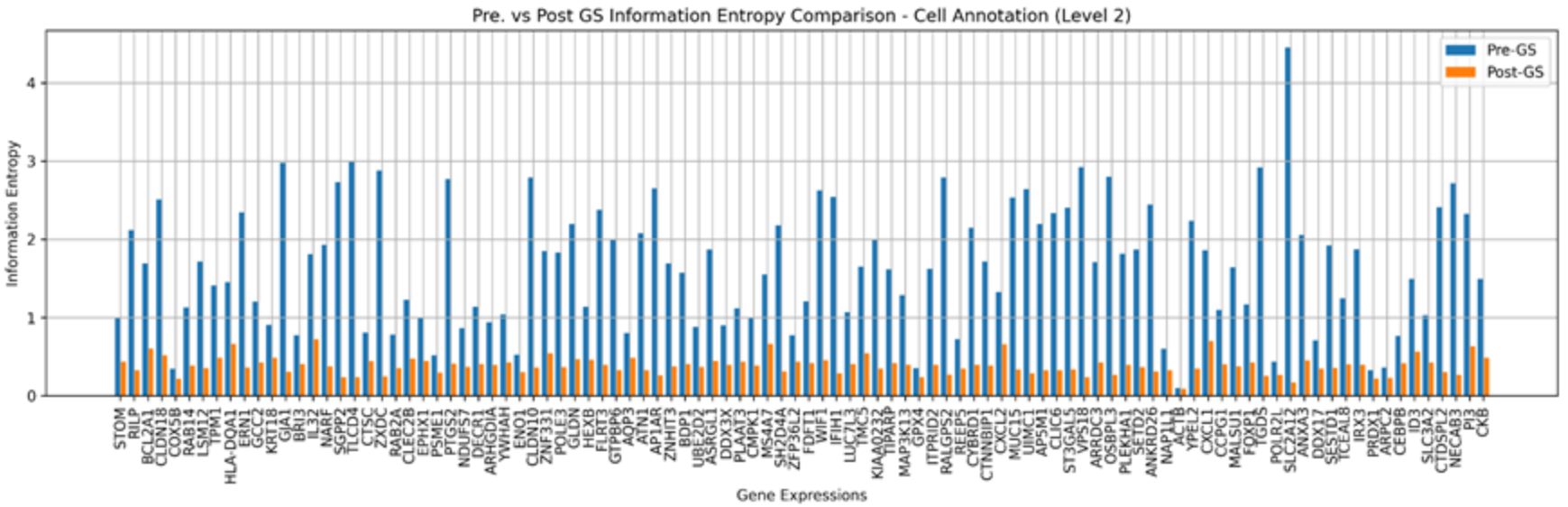
Holographic Divisibility

- Each data sample encapsulates information representative of the entire dataset, echoing the **holistic nature of holograms**.
  - Small Sample, Whole Dataset
- Equitable insights and analysis across diverse patient data.



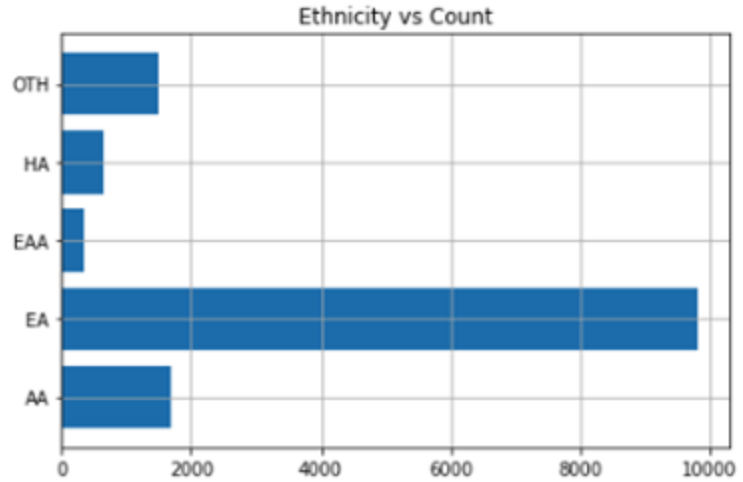
Uniform Information Distribution

- The Gerchberg-Saxton algorithm works in the frequency domain, **balancing intensity components** and preserving phase information during data transformation.
- Such transformation ensures a more **uniform distribution of information across datasets**.
- This uniformity is critical for obtaining **fairer results** from machine learning models and simplifying the **identification of feature model contributions**.



Previous Study - Mortality Prediction

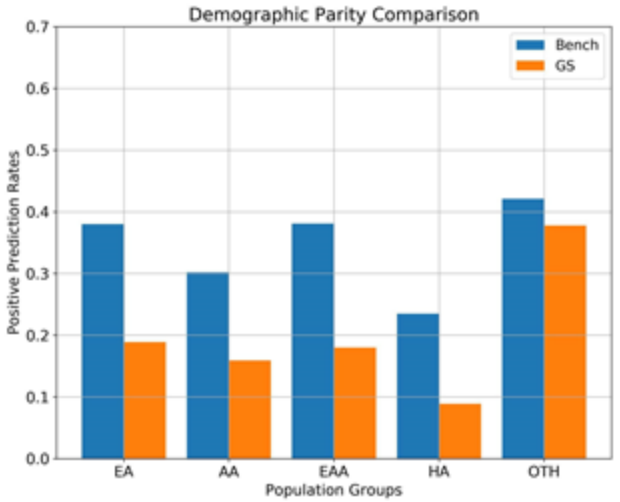
- Study Focus
  - Evaluation of mortality rates among ICU patients, with a particular emphasis on **detecting and correcting bias across diverse population groups**.
- Database Selection
  - The **MIMIC-III database** was chosen for its comprehensive data
  - **Unbalanced population distribution** across different racial groups.

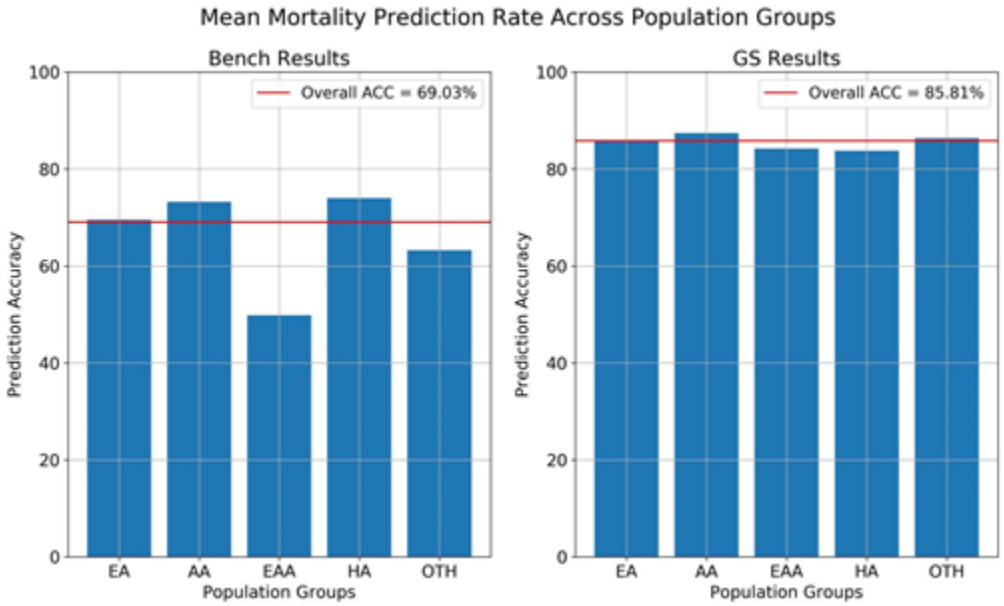


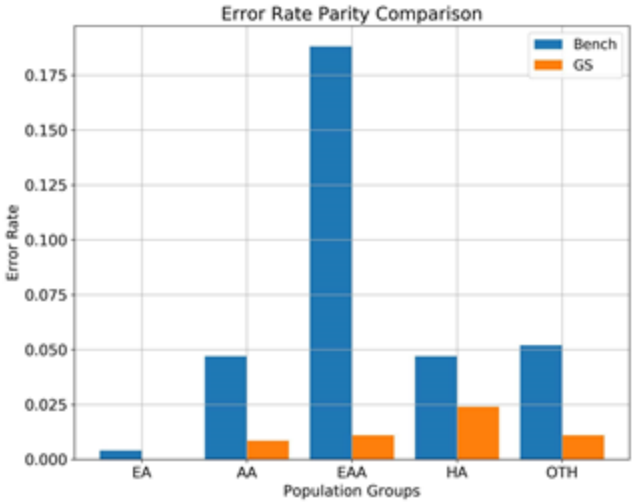
Ay, S., Cardei, M., Meyer, AM., Zhang, W., Topaloglu, U. "*Improving Equity in Deep Learning Medical Applications with the Gerchberg-Saxton Algorithm*". J Healthc Inform Res (2024). https://doi.org/10.1007/s41666-024-00163-8

Previous Study - Results

- Implementation of GS transformations on the dataset **significantly reduced the bias,** enhancing model predictions across different population groups.
- The improvements were quantitatively supported by **demographic parity** and **error rate parity**, demonstrating more uniform model prediction rates across demographics.





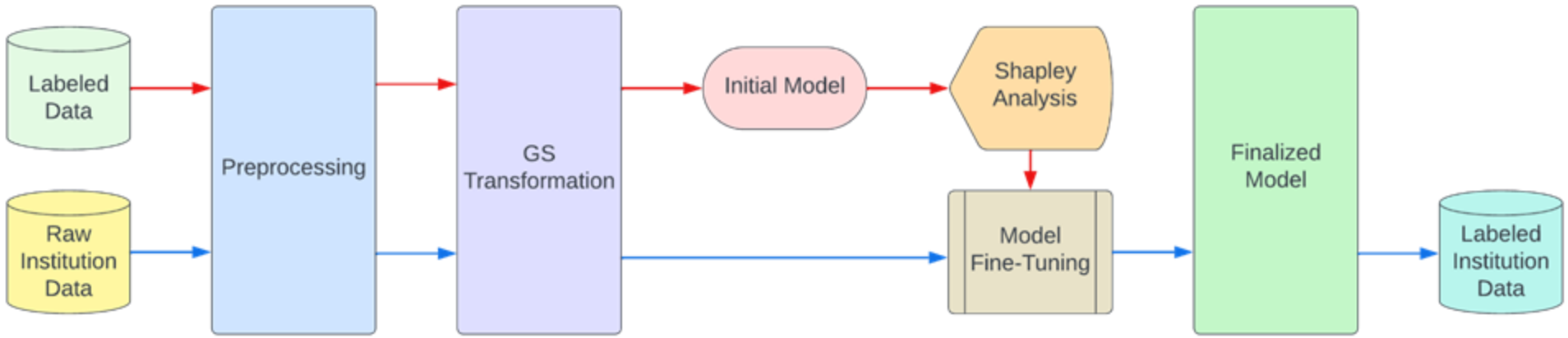


Project Objectives and Approach

- **Goal:** Enable fair data representations for single-cell cancer studies, improving healthcare decision-making.
- **Strategy:** Develop a pipeline to transform sc-RNA sequencing data into an AI-ready format, ensuring fair and uniform data representation through GS transformations.
    - Data Preprocessing and GS transformation
    - Model Development
    - Testing and Implementation

Pipeline

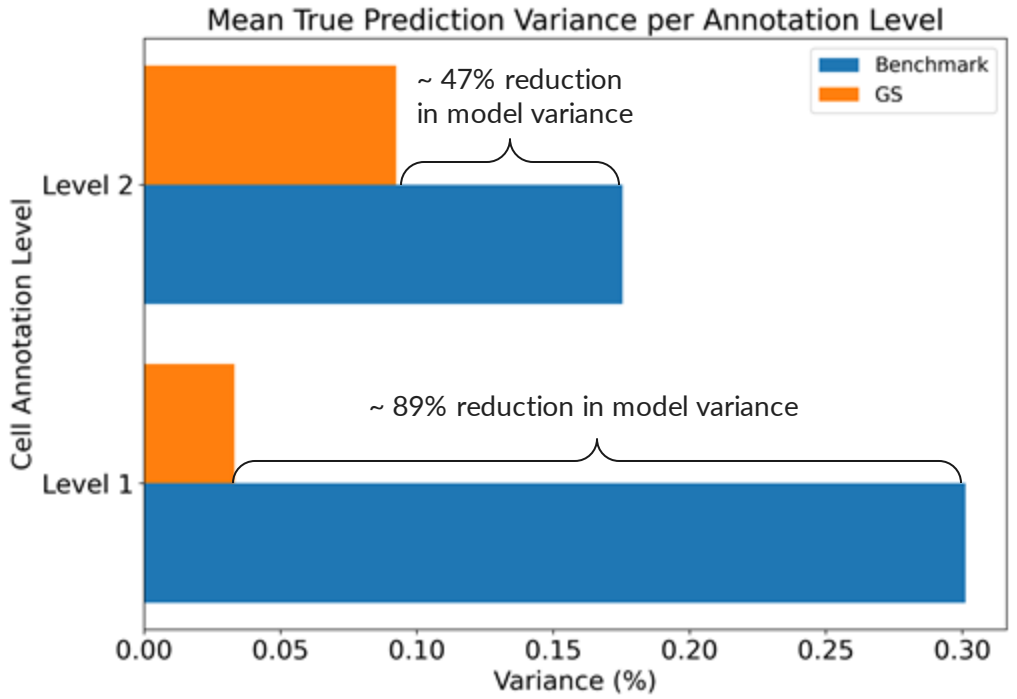
- Raw data **preprocessing** and **GS transformation** for fairness.
- Train models, then refine using **Shapley analysis** to identify and retain only **features that significantly contribute to model performance**.
- Validate model with **labeled public data subsets**.
- Apply to **institutional data** for enhanced decision-making in healthcare.



Initial Results - Model Prediction Uniformity

- Human Lung Cell Atlas, featuring **over 2 million single cell profiles from lung tissue**, with detailed cell annotations.
- Selected 20,000 cells **across diverse populations**, focusing on annotation levels 1 **(39 cell types)** and 2 **(61 cell types)**.

	Pre-GS Accuracy	Post-GS Accuracy
Level 1 (39 cell types)	90.62 %	89.48 %
Level 2 (61 cell types)	89.67 %	88.72 %



Future Directions and Expected Outcomes

- Pipeline Enhancement: Integrating a secondary model for tumor stage prediction to **distinguish between normal and cancer cells**, facilitating tumor microenvironment analysis.
- Comprehensive Analysis: Combination of models will **improve understanding** of *gene regulatory networks*, *cell-to-cell interactions*, and *therapeutic pathways*.
- Impact on Healthcare: Aims for enhanced, **equitable decision-making in healthcare** through better insights into cancer biology and treatment pathways.

Thank you

say@wakehealth.edu