Meta-Prediction and Machine Learning for Glaucoma Disease Risk

Lleyton Emery[1], Shreya Musini[1], Shaun Chen[1], Ahmed Khattab[1], Salvatore Loguercio[1]

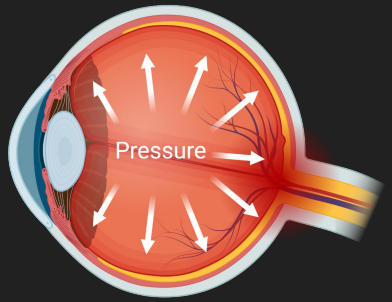
[1]Torkamani Lab, Scripps Research Translational Institute, La Jolla, CA

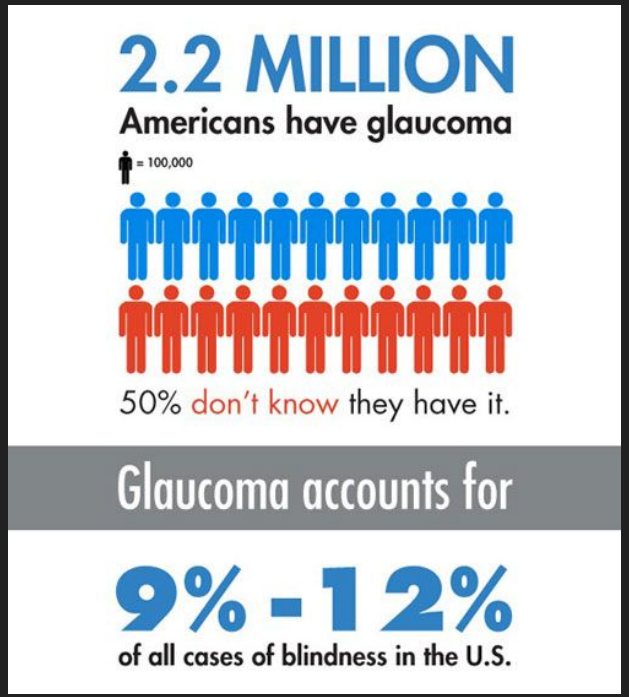
Scripps Research
Translational Institute

Introduction

What is glaucoma?

- Damage to the optic nerve
- Leading cause of irreversible blindness
- Early stages are asymptomatic
- Difficult to diagnose

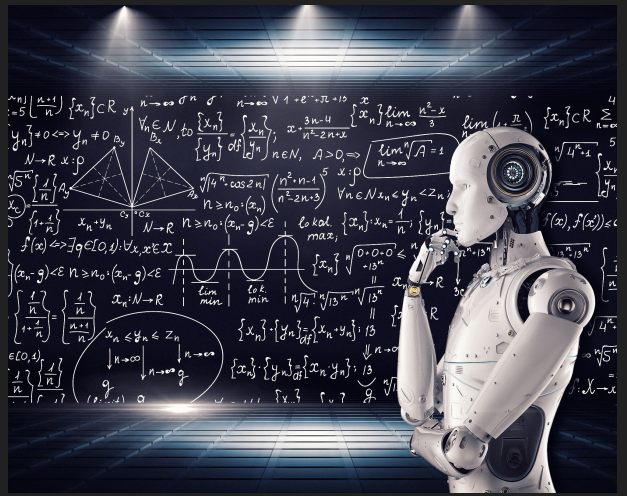




Scripps Research
Translational Institute

How can data science help?

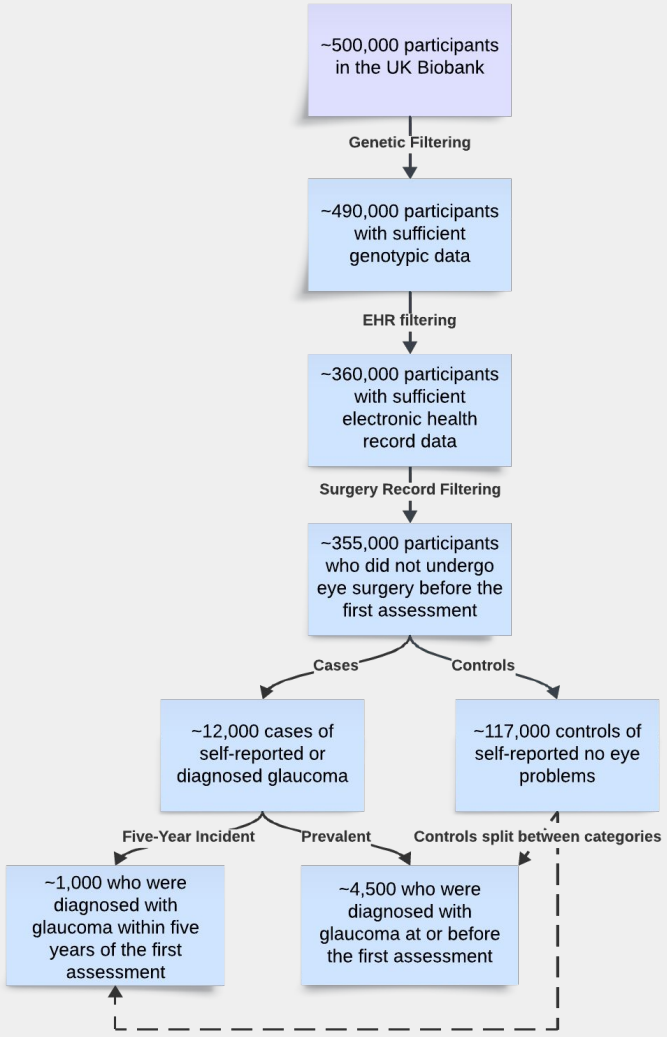
- Large amounts of phenotypic/genotypic data
  - Use machine learning to recognize patterns
- Predict five-year incident glaucoma risk
- Personalized intervention



Scripps Research
Translational Institute

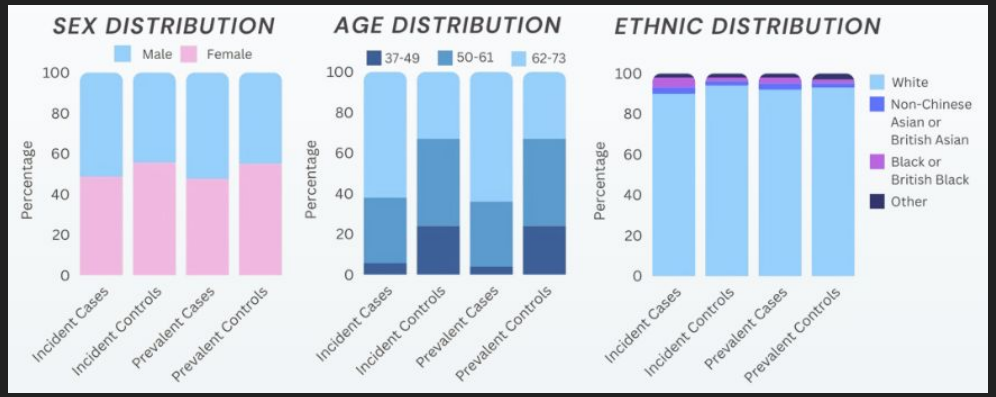
Methods

Creating our Cohorts



Phenotypic/genotypic data from British participants

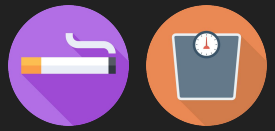
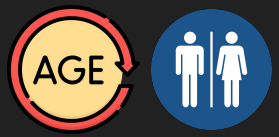
Demographic Distributions Across our Dataset Splits



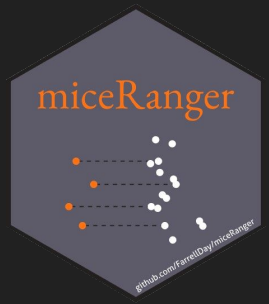
Scripps Research Translational Institute

Data Preprocessing

Initial Feature Pool

Phenotypes (~250)		Genotypes (~1,000)
Modifiable	Unmodifiable	Polygenic Risk Scores
		 Weighted Sum of SNPs → PRS

Imputation



- Data missingness
  - Over 80% in eye-specific variables!

Scripps Research Translational Institute

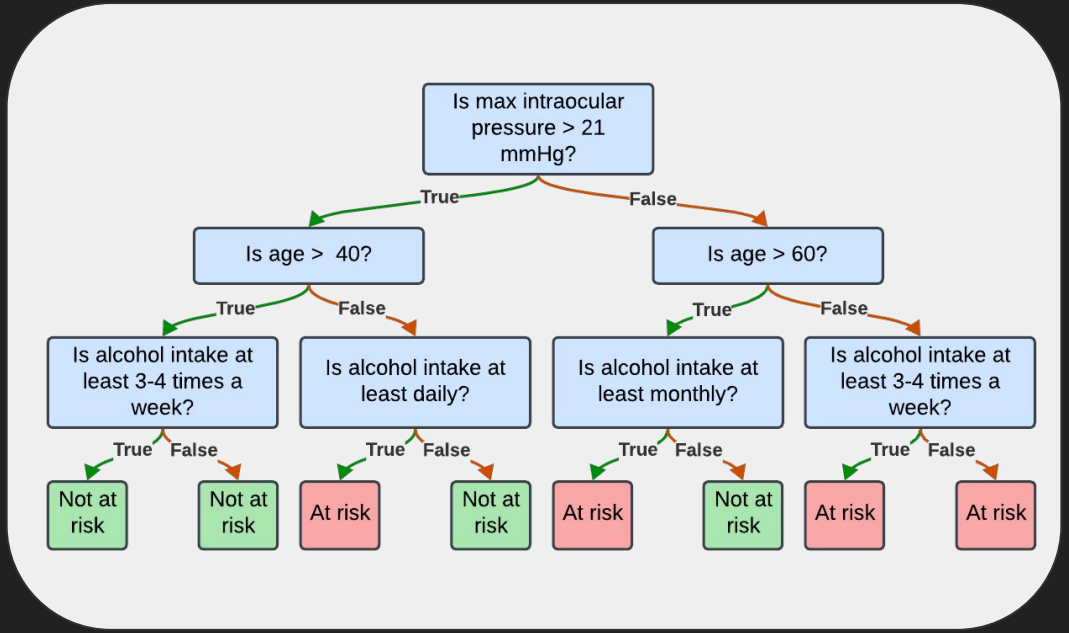
Modeling with Machine Learning

- XGBoost for binary classification
- Ensemble of decision trees

Article

Meta-Prediction of Coronary Artery Disease Risk

Ali Torkamani, Shang-Fu Chen, Sang Eun Lee, Hossein Sadaei, Jun-Bean Park, and 4 more

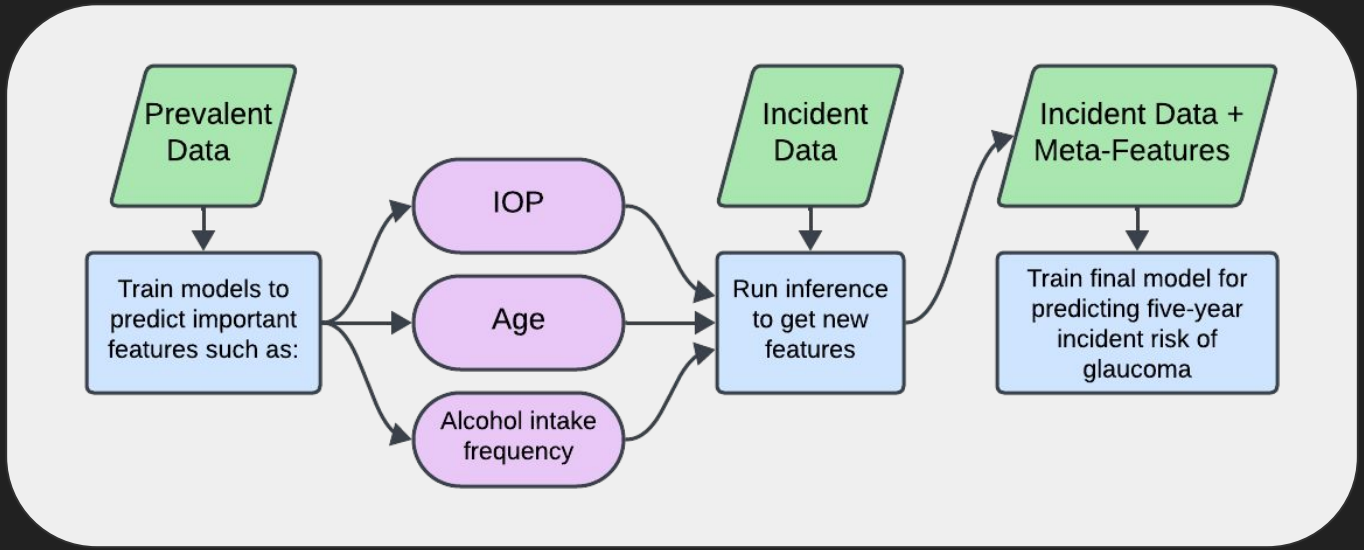


XGBoost

Scripps Research
Translational Institute

Incorporating Meta-Prediction

- We are already using incident data
- Can we use our prevalent data?



Scripps Research
Translational Institute

Results

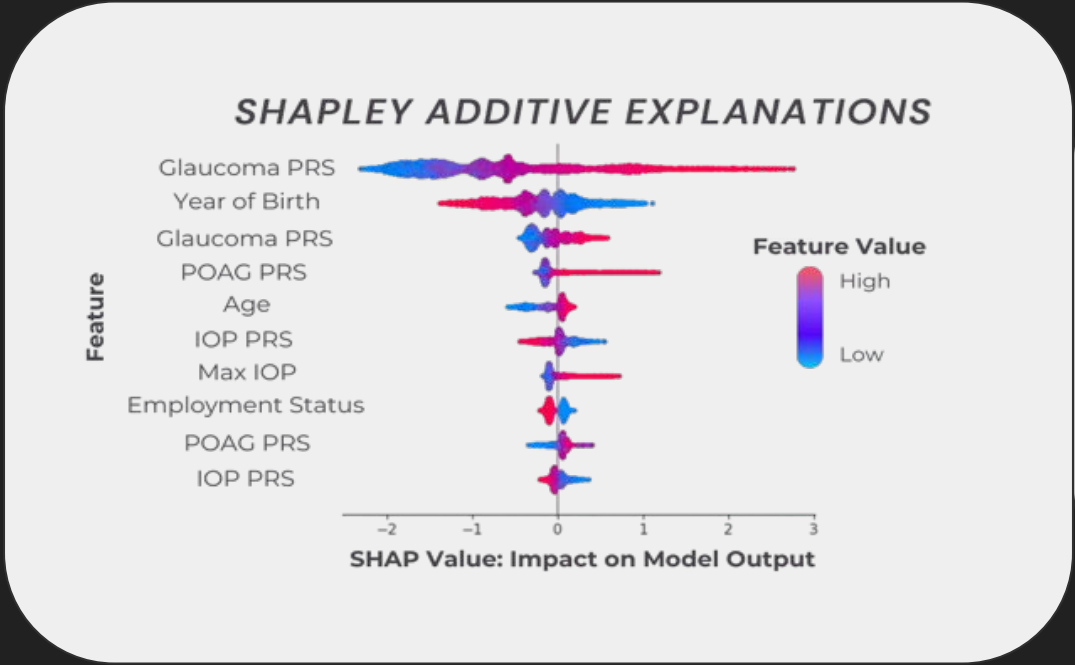
Feature Selection

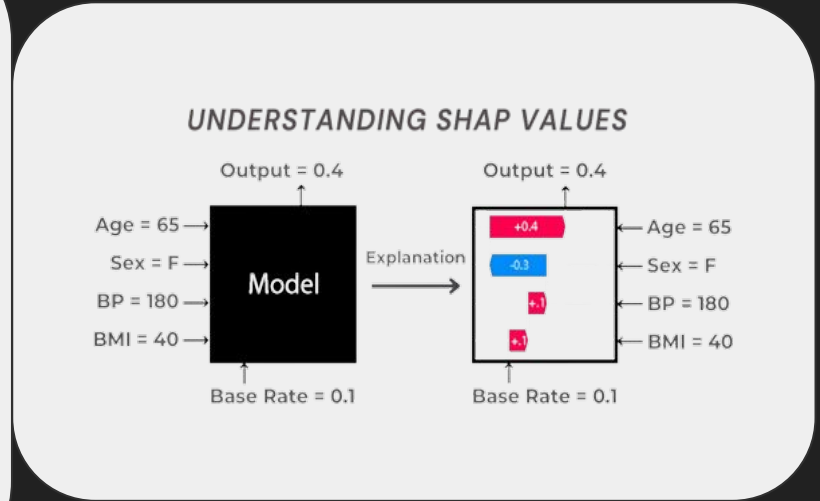
- Accuracy vs simplicity

Scripps Research
Translational Institute

SHAP Values

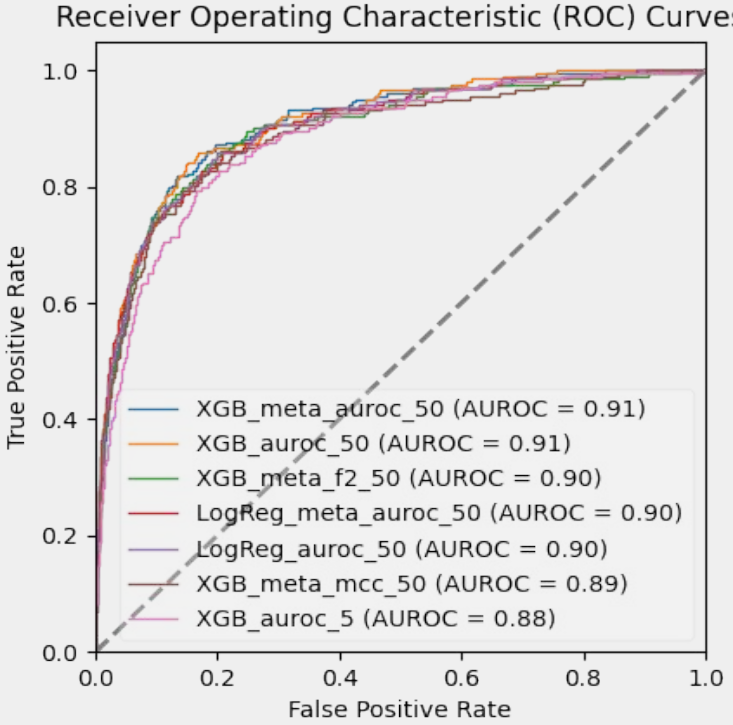
- Model explainability is key for intervention





Scripps Research
Translational Institute

Model Evaluation



CONFUSION MATRICES

Maxing AUROC

Actual \ Predicted	0	1
0	4691	48
1	143	60

Maxing F2 Score

Actual \ Predicted	0	1
0	4439	300
1	72	131

Maxing MCC

Actual \ Predicted	0	1
0	4694	45
1	131	72

$$F_2 = \frac{5}{\frac{4}{Precision} + \frac{1}{Recall}}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Scripps Research
Translational Institute

Conclusion

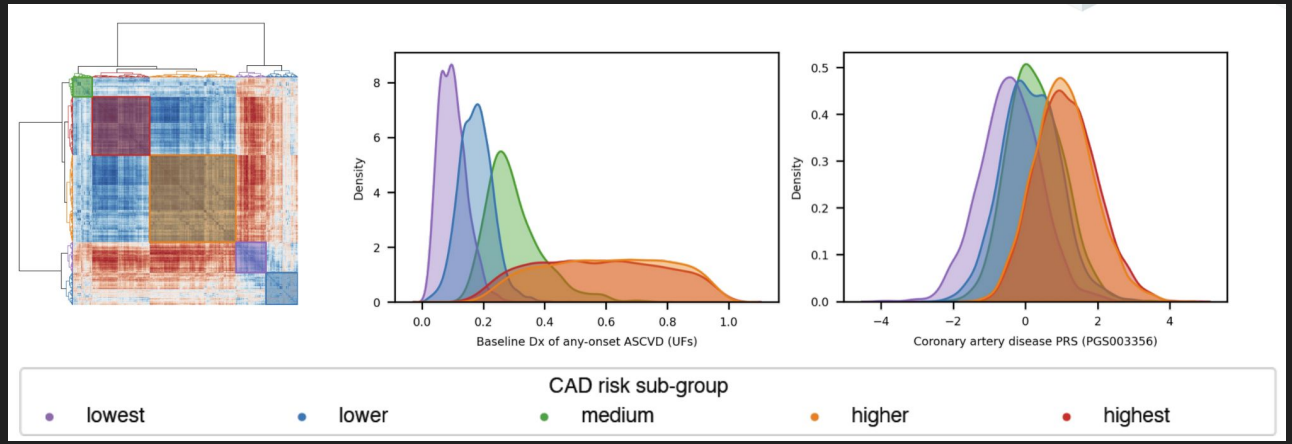
Was our model a success?

- Our models produce very good AUROC scores, however…
- Minority class classification can be improved
- Narrowly out-perform the baseline models

Scripps Research
Translational Institute

What's next?

- Can we acquire more data?
- Is our model generalizable?
  - External validation on UCSD African American glaucoma cohort
- Risk stratification



Intervention

Scripps Research Translational Institute

Acknowledgments

- SRTI intern program heads: Theresa Hill, Dr. Laura Nicholson, Dr. Jaiswal
- Mentor: Dr. Salvatore Loguercio
- Glaucoma experts: Dr. Anthony Khawaja, Dr. Je Hyun Seo
- Torkamani Lab: Dr. Ali Torkamani, Dr. Shaun Chen, Ahmed, Kaushik, Austin, Shreya

Scripps Research
Translational Institute

References

[1] Torkamani, A., Chen, S.-F., et al. Meta- Prediction of Coronary Artery Disease Risk, 20 Dec. 2023, PREPRINT (Version 1) available at Research Square.

[2] Craig, J.E., et al. Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. Nat Genet 52, 160–166 (2020).

Thank you

Questions?