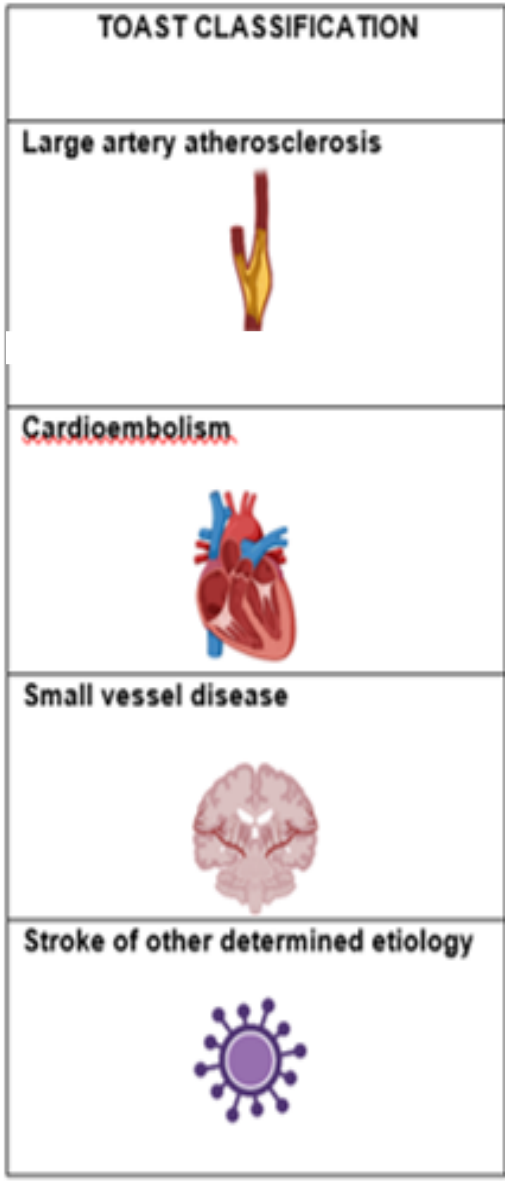
StrokeClassifier:

An Automated, EHR-based Ischemic Stroke Etiology Classification System

Richa Sharma, MD, MPH
Ho-Joon Lee, PhD
Yale School of Medicine
Department of Neurology

Introduction

- ~ 676,000 ischemic strokes annually in the US
  - ¼ are recurrent
- Unlike heart attacks, causes (etiologies) of stroke can be diverse:
  - Large artery atherosclerosis
  - Cardioembolism
  - Small vessel disease
  - Other, rare determined causes
  - Undetermined or cryptogenic cause
    - Currently upto 40% of strokes are determined to be cryptogenic
    - Recurrent stroke rate is upto 20%



The Clinical Gap: An Unmet Need

- Though we have evidence-based therapies to target specific stroke etiologies to reduce recurrent stroke risk, they are **underutilized**.

Dual blood thinners under prescribed after minor or warning stroke, especially in women

June 04, 2019; 92 (23) ARTICLE

Lower carotid revascularization rates after stroke in racial/ethnic minority-serving US hospitals

Anticoagulation under-utilization in atrial fibrillation patients is responsible for a large proportion of strokes requiring endovascular therapy

Underuse of statins in patients with atherosclerotic ischemic stroke in China

Reasons for Secondary Stroke Prevention Therapy Under-Utilization

1. Diagnostic uncertainty

   Diagnosis requires synthesis of large amounts of patient data collected during a stroke hospitalization

2. Lack of local stroke expertise

   Only **1 in 6 stroke patients** is treated by a board-certified vascular neurologist

3. Therapeutic inertia

Solution: *StrokeClassifier* Clinical Decision Support Tool

Data Source: 2 academic, Comprehensive Stroke Centers (n=2,039)

Features: 2,027

Model: Ensemble model of 9 classifiers

Validation:

5-fold cross-validation

External validation in MIMIC-3 data

Data: Discharge summaries from stroke EHR
YNHH and MGH for 4 non-cryptogenic TOAST etiology classes

Feature extraction/engineering
: *MetaMap* UMLS CUIs

Feature extraction/engineering
: regular expression, imputation, discretization, etc.

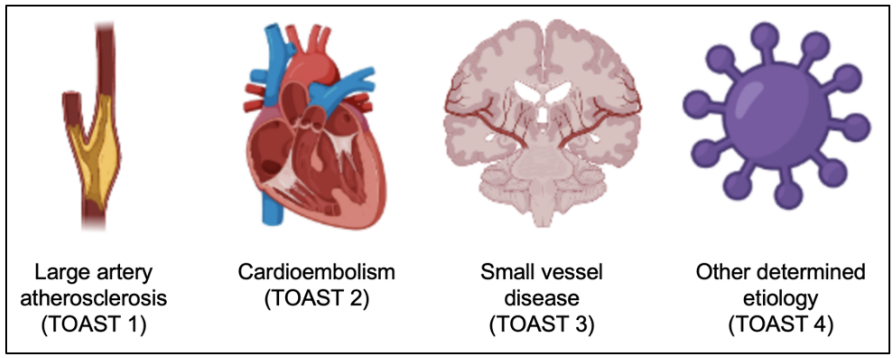
Etiology classification by StrokeClassifier, an ensemble consensus model based on multiple classifiers and comprehensive CV

Test of model generalizability
YNHH+MGH to MIMIC; YNHH to MGH; MGH to YNHH

Etiology prediction of non-cryptogenic samples

Etiology prediction of cryptogenic samples

Provisional patent filed



Automated Placement of *StrokeClassifier* Output as a Progress Report in the EHR in real-time

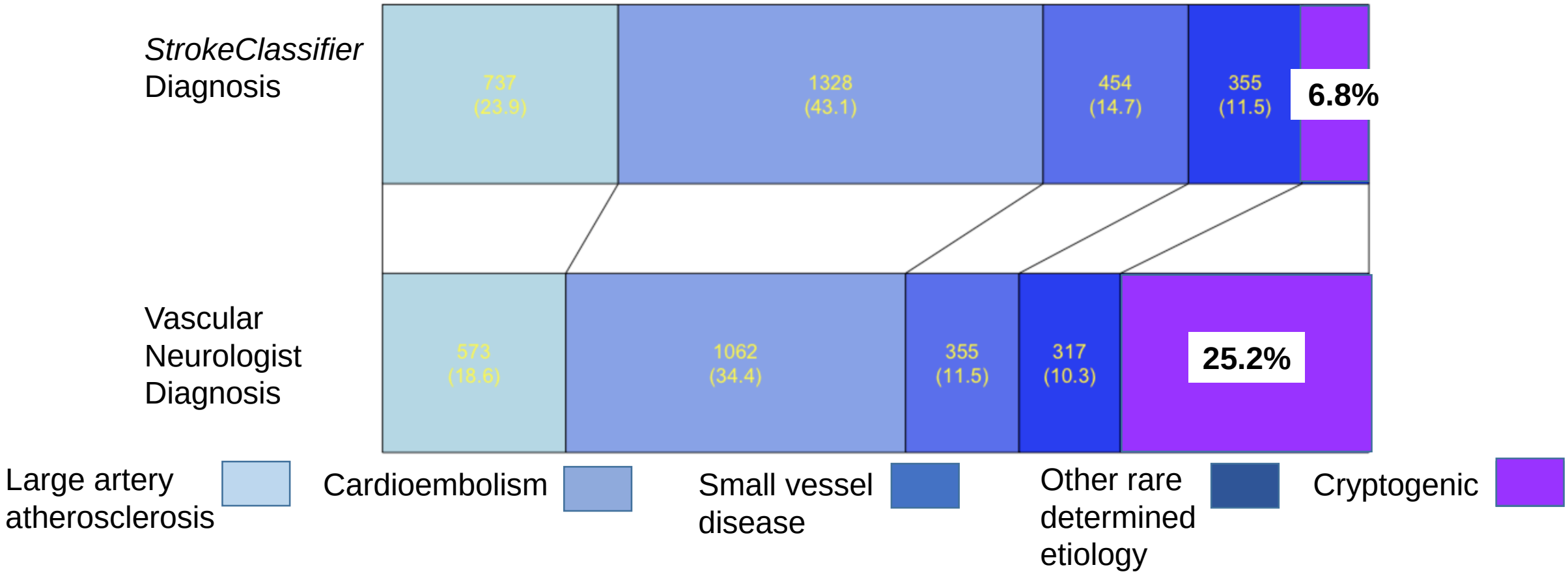
StrokeClassifier Recommendation Report

Target	Guideline-Recommended Therapy	Evidence
Predicted Stroke Etiology		
Atrial fibrillation	Anticoagulation with DOAC	URL
Other Features		
LVEF 38%	Quadruple medical therapy for heart failure	URL
LDL 129 + statin allergy	PCSK9 inhibitor	URL
Hypertension	ACE inhibitor	URL
BMI 42	Mediterranean diet; nutrition referral	URL
Tobacco use	Nicotine replacement; Referral to smoking cessation program	URL
Tests Still Necessary		
Carotid ultrasound	-	URL
Stroke Prevention Clinical Trial Eligibility		
NCT_____		

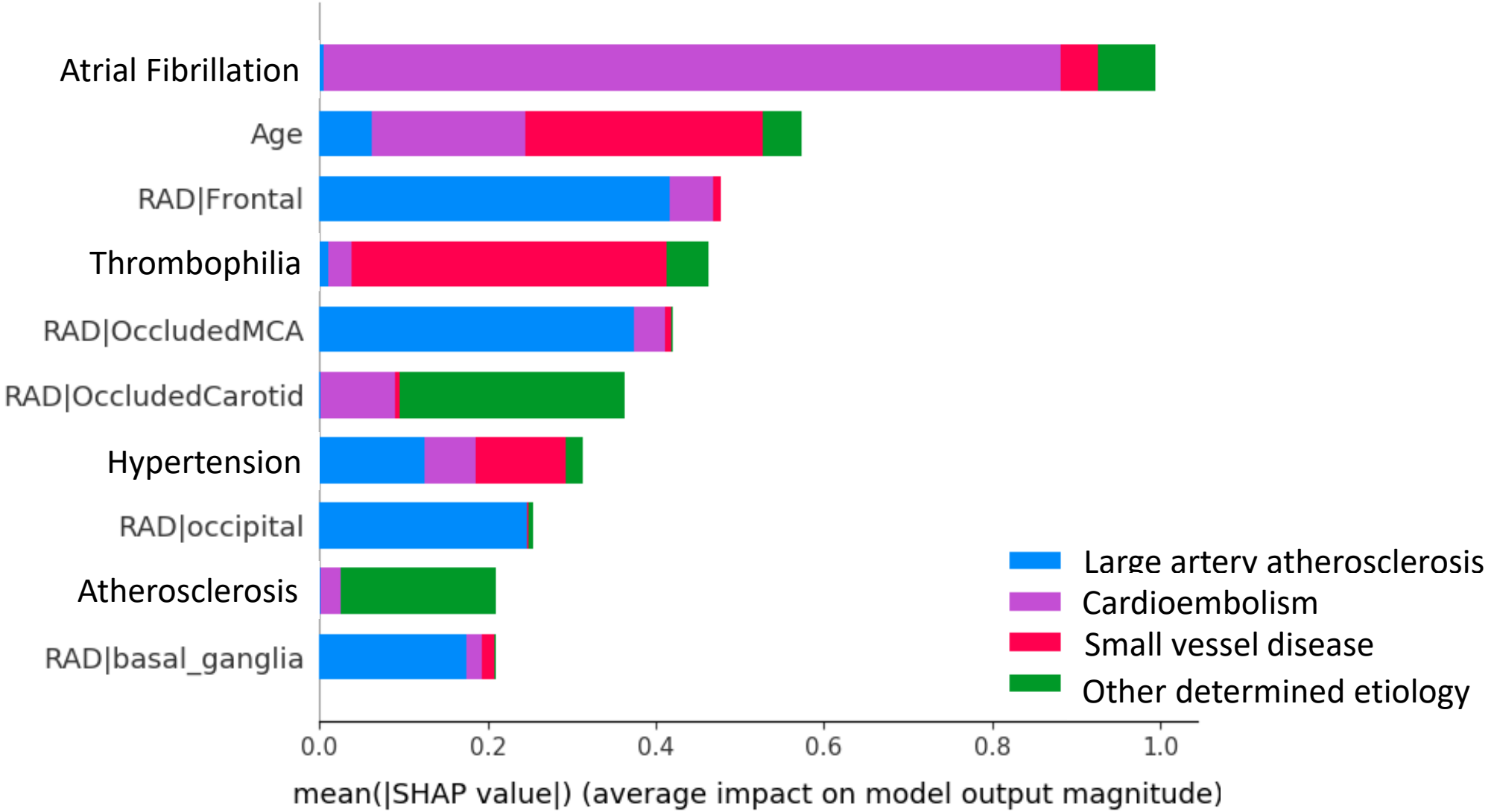
Benchmarks: Added Value of *StrokeClassifier*

Stroke Etiology	Stroke Etiology Classification Compared with Adjudicated Etiology (Internal Accuracy)	External Validation in MIMIC-3 (External Accuracy)
StrokeClassifier	***~75%***	***~69%***
Trained GWTG data abstractors at YNHH	~57%	-
Non-vascular neurologists at YNHH	~56%	-

StrokeClassifier Identified a Cause of Stroke in 73% Deemed Cryptogenic by Vascular Neurologists



Top 10 Features Contributing to Stroke Etiology Predictions



Major Competitors

viz.ai

RAPIDAI

BRAINOMIX

qure.ai

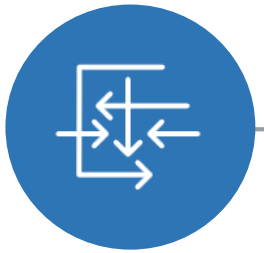
aidoc

CAIDE

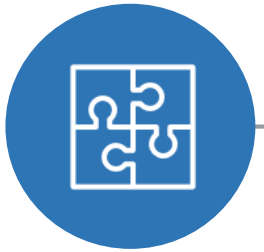
CAIDE Systems

StrokeClassifier is the <u>only</u> automated tool for secondary stroke prevention.

Personalized Diagnostics



Comprehensive Inputs and Validated Algorithm



Easy, User-Friendly Integration

Hospitals Currently Pay Annual Fees for Stroke Diagnostics

Stroke Diagnostics	Annual Cost Per Hospital
RapidAI	$25,000
Telestroke	$5,193
Vascular Neurologist	$195,471

* There are 5533 hospitals listed in the U.S.

Project Timeline

Milestone	Year 1				Year 2			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
1. **Integration of *StrokeClassifier* in YNHH Epic**		**BLAVATNIK FUNDING WOULD SUPPORT DATA SCIENTIST EFFORT**						
2. Beta-testing *StrokeClassifier* in YNHH								
3. Qualitative survey to YNHH neurologists								
4. Integration of StrokeClassifier in YNHH Telestroke Network Hospitals (n=5)								
5. Cluster randomized controlled trial (n=6)								
6. External Validation in Veterans Affairs Medical Centers								

Current Team and Status

Ho-Joon Lee, PhD

Associate Research Scientist in Genetics

Richa Sharma, MD, MPH

Yale Vascular Neurologist funded by NIH NINDS K23

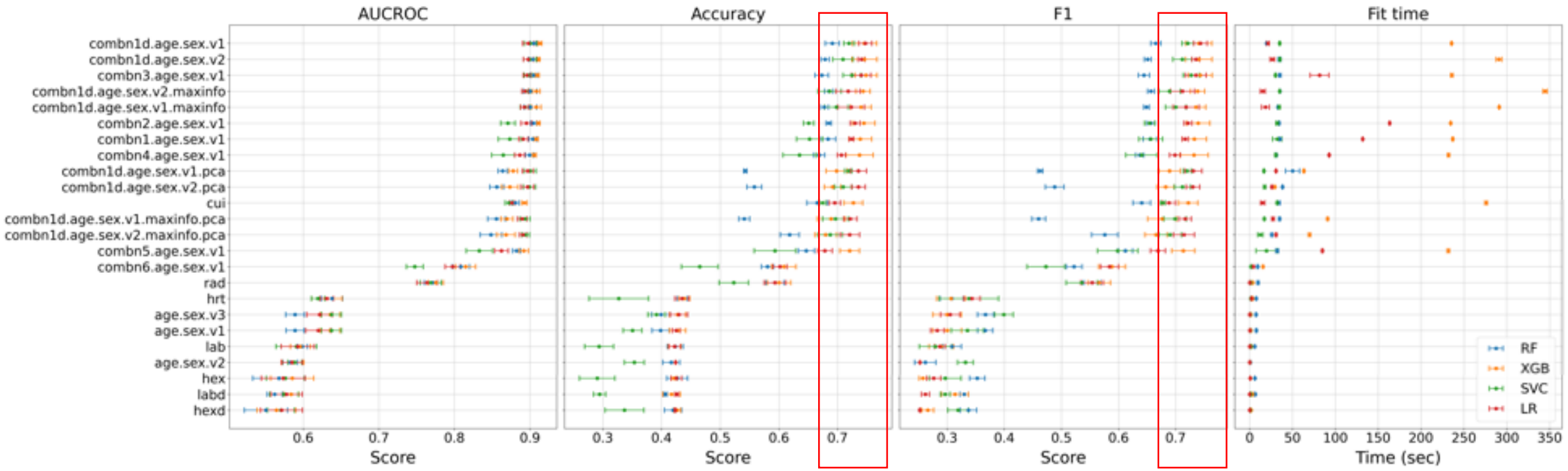
- Working with Robert Migliorini at Yale Ventures
  - Provisional patent filed

*<u>Grant Mentors</u>

Dr. Lee **Schwamm**, Dr. Harlan **Krumholz**, Dr. Lauren **Sansing**, Dr. Cynthia **Brandt**, Dr. Hongyu **Zhao**

Thank You

Performances of each optimized model for each feature group by 5-fold CV



Development and Validation of *StrokeClassifier*

- ***StrokeClassifier* Algorithm**: Ensemble consensus model using 9 supervised machine learning-based classifiers

Model	AUCROC	ACCURACY	F1	KAPPA
CLF1	0.898±0.008	0.747±0.012	0.744±0.013	0.632±0.019
CLF2	0.900±0.009	0.728±0.007	0.725±0.008	0.607±0.011
CLF3	0.887±0.009	0.719±0.010	0.721±0.011	0.606±0.014
CLF4	0.913±0.003	0.746±0.023	0.741±0.025	0.627±0.035
CLF5	0.905±0.005	0.691±0.013	0.665±0.010	0.523±0.021
CLF6	0.907±0.006	0.736±0.009	0.736±0.010	0.626±0.012
CLF7	0.907±0.007	0.743±0.009	0.740±0.010	0.628±0.014
CLF8	0.912±0.005	0.750±0.008	0.748±0.009	0.640±0.011
CLF9	0.910±0.005	0.749±0.009	0.746±0.010	0.636±0.014
StrokeClassifier	***NA***	***0.744±0.009***	***0.740±0.010***	***0.629±0.014***

***StrokeClassifier* Characteristics (5-fold CV)**

StrokeClassifier Characteristics for each stroke etiology

Physician Diagnosis	*StrokeClassifier* Accuracy
Large artery atherosclerosis (1)	0.836±0.015
Cardioembolism (2)	0.829±0.014
Small vessel disease (3)	0.909±0.010
Other determined (4)	0.913±0.006

Carotid stenosis: Carotid artery revascularization (recurrent stroke risk reduction of 17%)

Intracranial stenosis: Dual antiplatelet therapy for 3 months (HR 0.66)

Atrial fibrillation: Anticoagulation (OR 0.56)

Endocarditis: Antibiotics (stroke risk reduction: 65%)

Patent foramen ovale: Closure (HR 0.23)

Dual antiplatelet therapy for 21 days (0.68)
Hypercoagulability of malignancy: Anticoagulation (HR 0.26)