# Big Data Analytics and Applied Data Science within Geisinger's Large Integrated Health System

Brandon Fornwalt, MD, PhD

Geisinger

CARDIAC IMAGING TECHNOLOGY LAB

# Disclosures

- No one works alone

Chris Haggerty PhD

# A Gamecock Then vs Now

More hair

Nicer grill

# Proudest Gamecock Moment

??? 

Correct “spur”


Circa 2001

# Truly Proudest Gamecock Moments

# Outline

- Intro Geisinger, people, data, infrastructure
- Clinical radiology informatics with deep learning
- Machine learning to optimize use of cardiac image data
- Deep learning with 12-lead electrocardiograms
- Machine learning to help manage heart failure
- Multi-modal deep learning in cardiology

# Elias Zerhouni, MD

Former Director of the National Institutes of Health and presidential science advisor

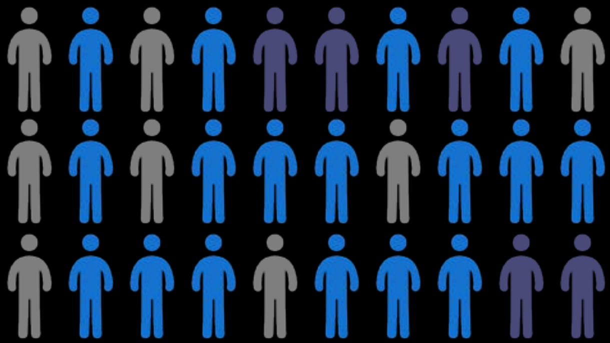
“The real winners in artificial intelligence and machine learning will be those health systems who have large databases of patients with longitudinal studies and outcomes, such as Geisinger.”

# What Do We Need to Deliver on that Expectation?

- Data – including a deep understanding of the underlying concepts
- People – diverse team
- Infrastructure – specialized compute
- Clinical implementation
  - Quantifiable impact on our patients

# Geisinger: Rich Clinical Data from Integrated Network of 13 Hospitals

Epic (since 1996)



1.9 Million patients

500 Million labs

800 Million vital signs

11 Million imaging studies

3.2 Billion rows

~140,000 whole exomes sequenced

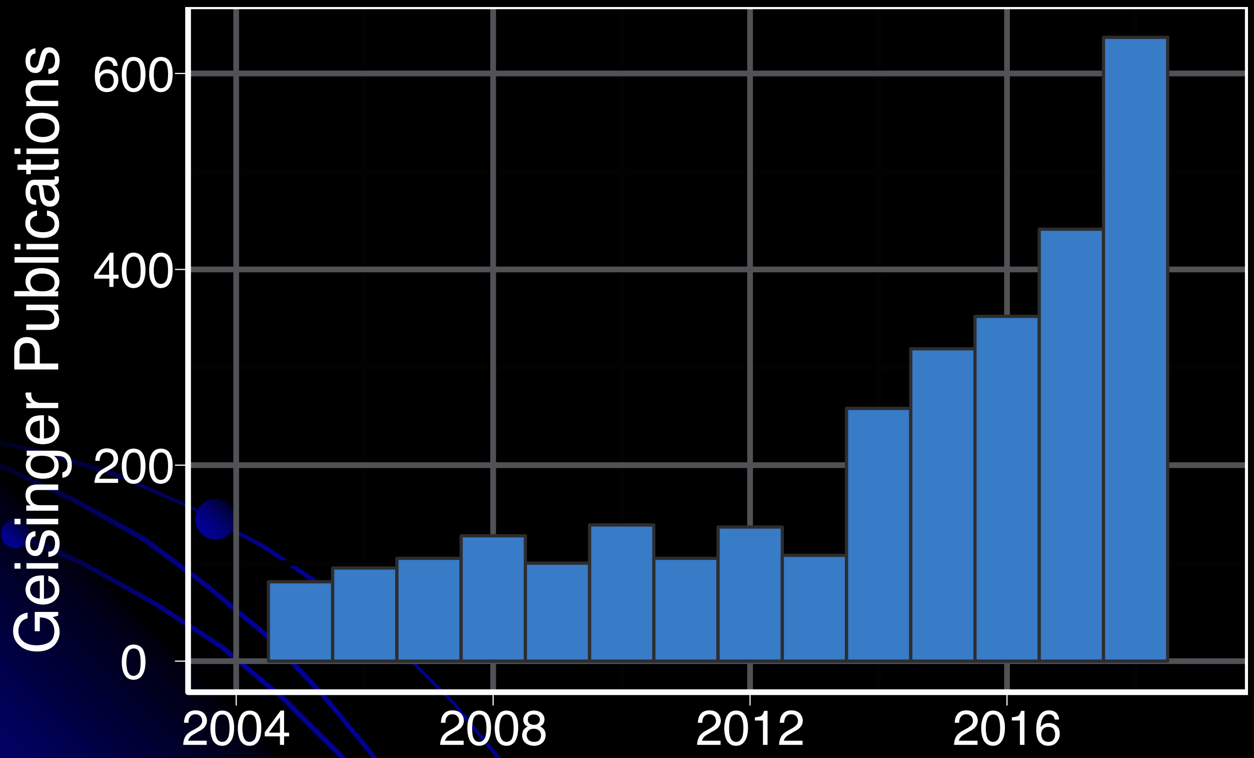
→ average **16 years** follow-up

REGENERON GENETICS CENTER™

mycode

***Large phenotype team of data analysts / modelers***
***(Joe Leader, AVP of Informatics)***

# People: Research Leadership With Institutional Re-Investment in 2010




David Ledbetter PhD
Chief Scientific Officer

# Diverse Team: Clinicians, Comp Sci, Engineers, Ops

## Ops

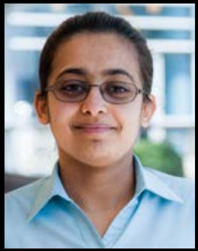

A Haggerty MBA

## Scientists


E Carruth PhD


S Fielden PhD


L Jing PhD


S Raghun-ath, PhD


C Still-well, PhD


A Ulloa PhD


X Zhang PhD

## Primary Staff


B McCarty BS


C Nevius BS


N Stoudt BS


D van-Maanen, MS

## Associated Staff


S Gazes MS


D Hartzel BS


J Leader BS


N Sauers PharmD

## Primary Faculty


El-Manzal-awy, PhD


B Fornwalt MD PhD


C Haggerty PhD

## Cardiology / Hospitalist Co-investigators


B Carry MD


C Good DO


J Pfeifer MD, MPH


G Schneider MD

# Unique Infrastructure

A Patel MD

D Ledbetter PhD

From PURESTORAGE

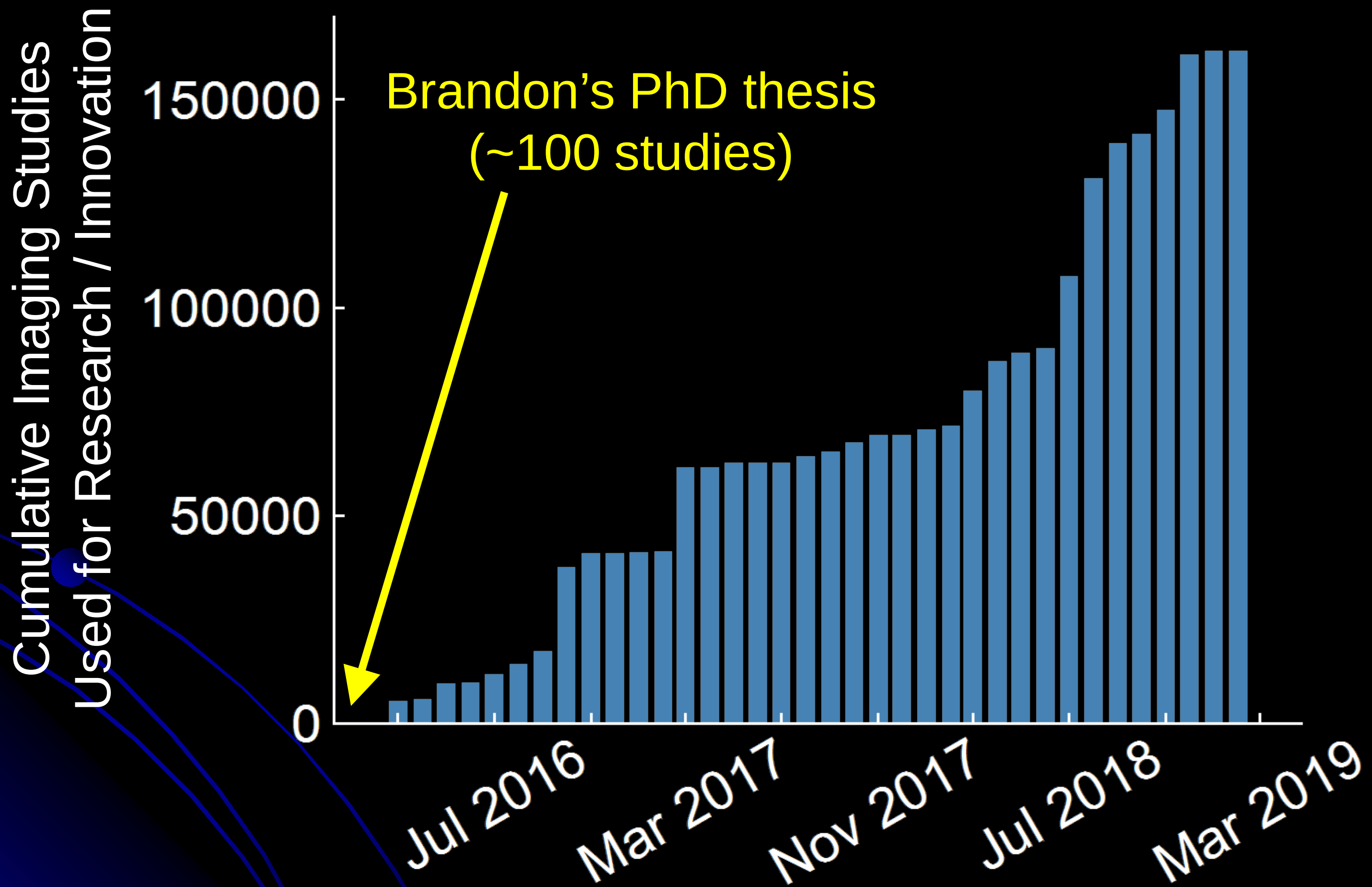
Powered by NVIDIA

2018: DGX-1 (8 GPUs) INSIDE clinical network (2nd hospital; 1st MGH)

2019: DGX-2 (16 GPUs)
+60TB FlashBlade Array

2020: 600TB Flash
+1.3PB spinning disk

# Research Image Archive (PACS) Growth



# Outline

- Intro Geisinger, people, data, infrastructure
- Clinical radiology informatics with deep learning
- Machine learning to optimize use of cardiac image data
- Deep learning with 12-lead electrocardiograms
- Machine learning to help manage heart failure
- Multi-modal deep learning in cardiology

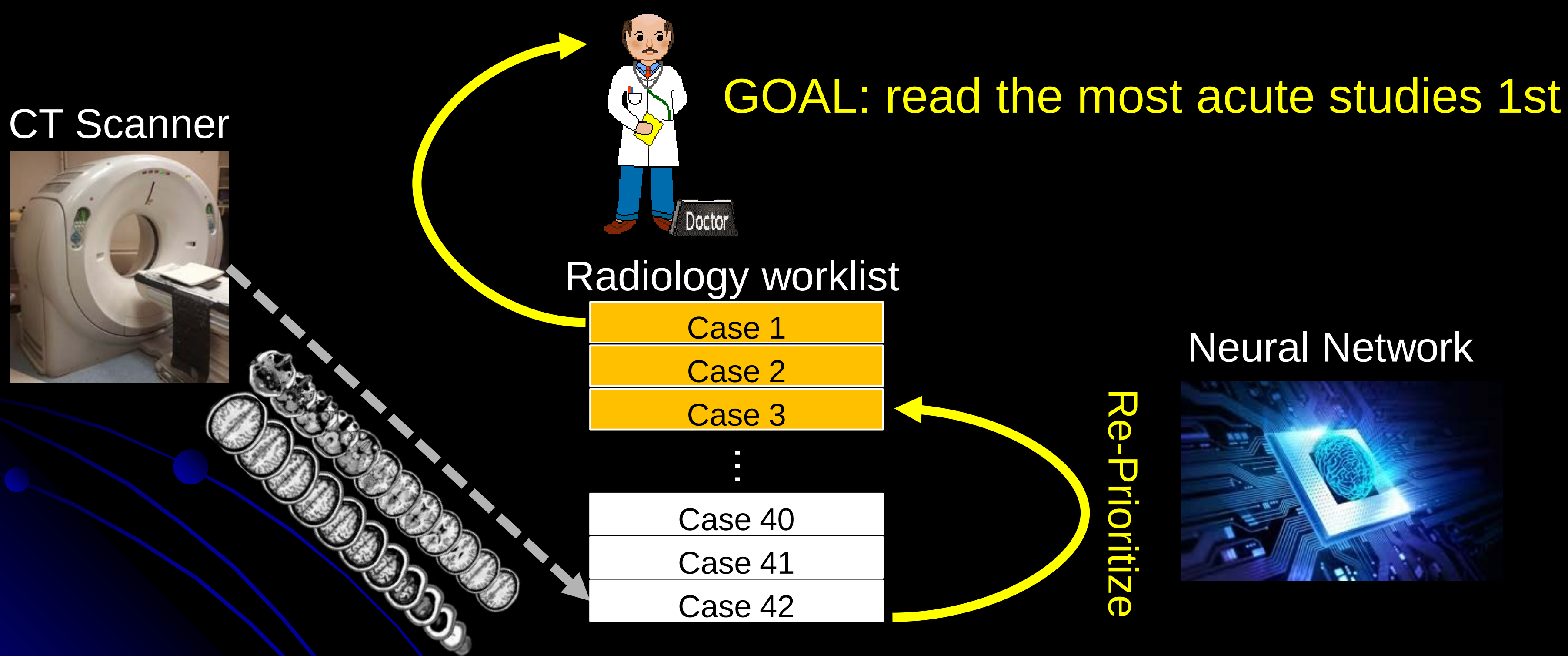
# Clinical Radiology Informatics with Deep Learning: Intracranial Hemorrhage (ICH)

- Spearheaded by Mohammad Arbabshirani, PhD

- Early and accurate diagnosis of ICH is critical to patient outcome
- Can we use machine learning to optimize radiology worklists for earlier diagnosis?
  - Quality improvement tool as initial proof of concept

M Arbabshirani et al. *Nature Partner Journal: Digital Medicine*. 2018.

# Clinical Radiology Informatics with Deep Learning: Intracranial Hemorrhage (ICH)

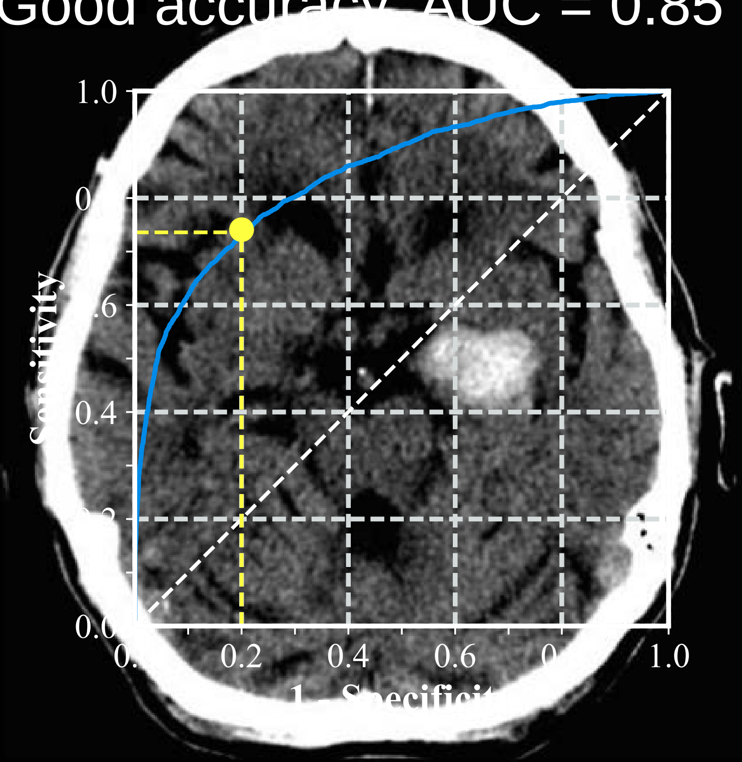



M Arbabshirani et al. *Nature Partner Journal: Digital Medicine*. 2018.

# Clinical Radiology Informatics with Deep Learning: Intracranial Hemorrhage (ICH)

- 46,583 head CTs
  - 3D→2 million images
  - collected over 10 years
- Classified into ICH or no ICH and used to train a deep neural network

Good accuracy: AUC = 0.85



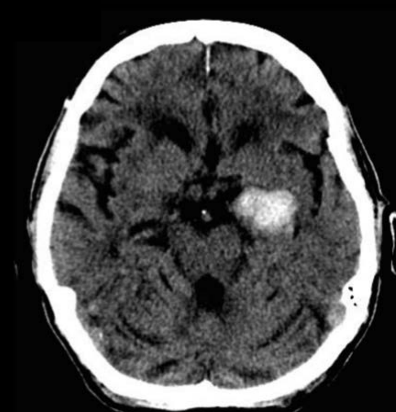

M Arbabshirani et al. *Nature Partner Journal: Digital Medicine*. 2018.

# Operational >2 years


Healthcare IT News
TOPICS

Geisinger injects machine learning into clinical workflow to find health problems faster

Hospitals executive says the algorithms can help clinicians reduce time to diagnosis of intracranial hemorrhages by 96 percent.



Clinical implementation of a machine learning algorithm reduced time to diagnosis of new outpatient cases of intracranial hemorrhage by 96%

Approximately 10% of "false positives" likely have subtle hemorrhage

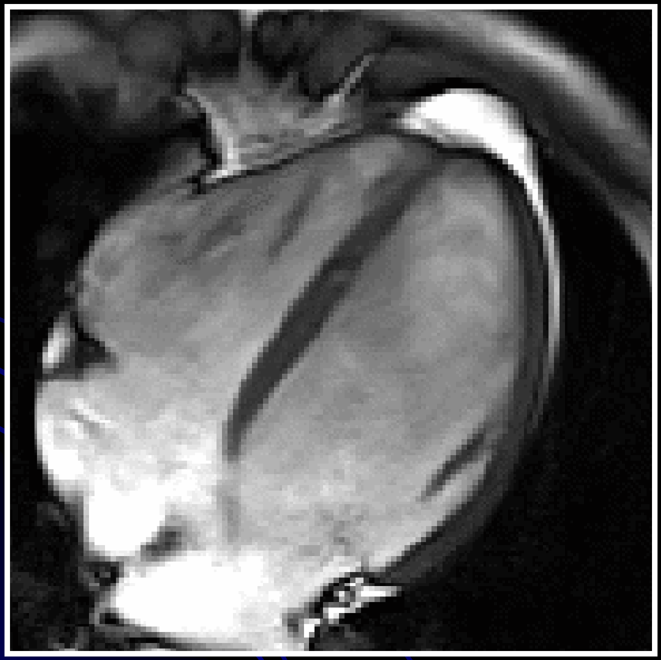
M Arbabshirani et al. *Nature Partner Journal: Digital Medicine*. 2018.

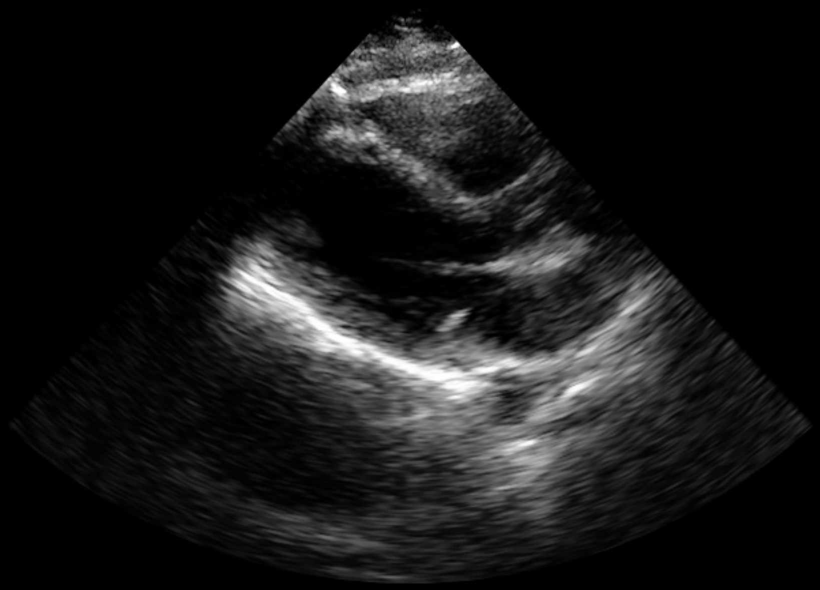
# Outline

- Intro Geisinger, people, data, infrastructure
- Clinical radiology informatics with deep learning
- Machine learning to optimize use of cardiac image data
- Deep learning with 12-lead electrocardiograms
- Machine learning to help manage heart failure
- Multi-modal deep learning in cardiology

# Cardiac Imaging

- Imaging forms the cornerstone for diagnosis, prognosis, and management of heart disease

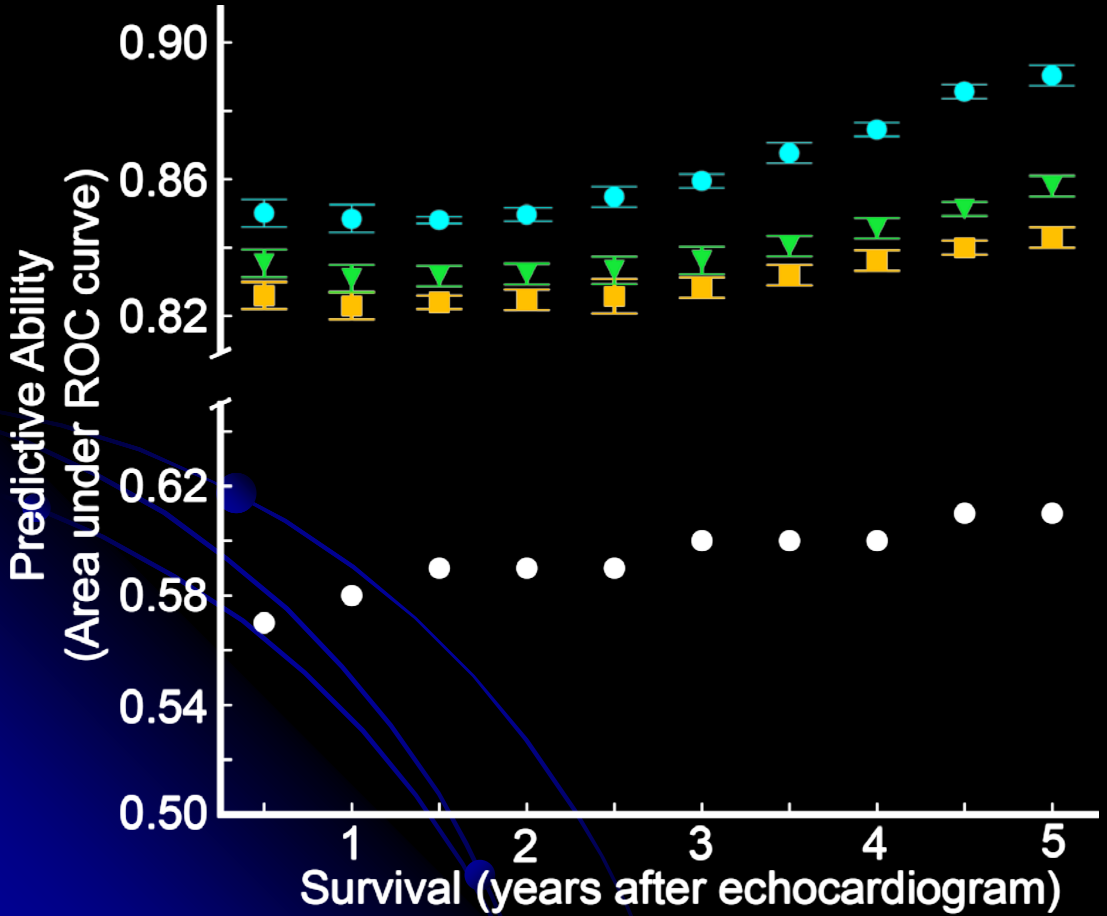




# Can we improve mortality prediction after imaging using machine learning?

171,510 patients (331,317 echocardiograms)

M Samad PhD




Clinical + all image measures

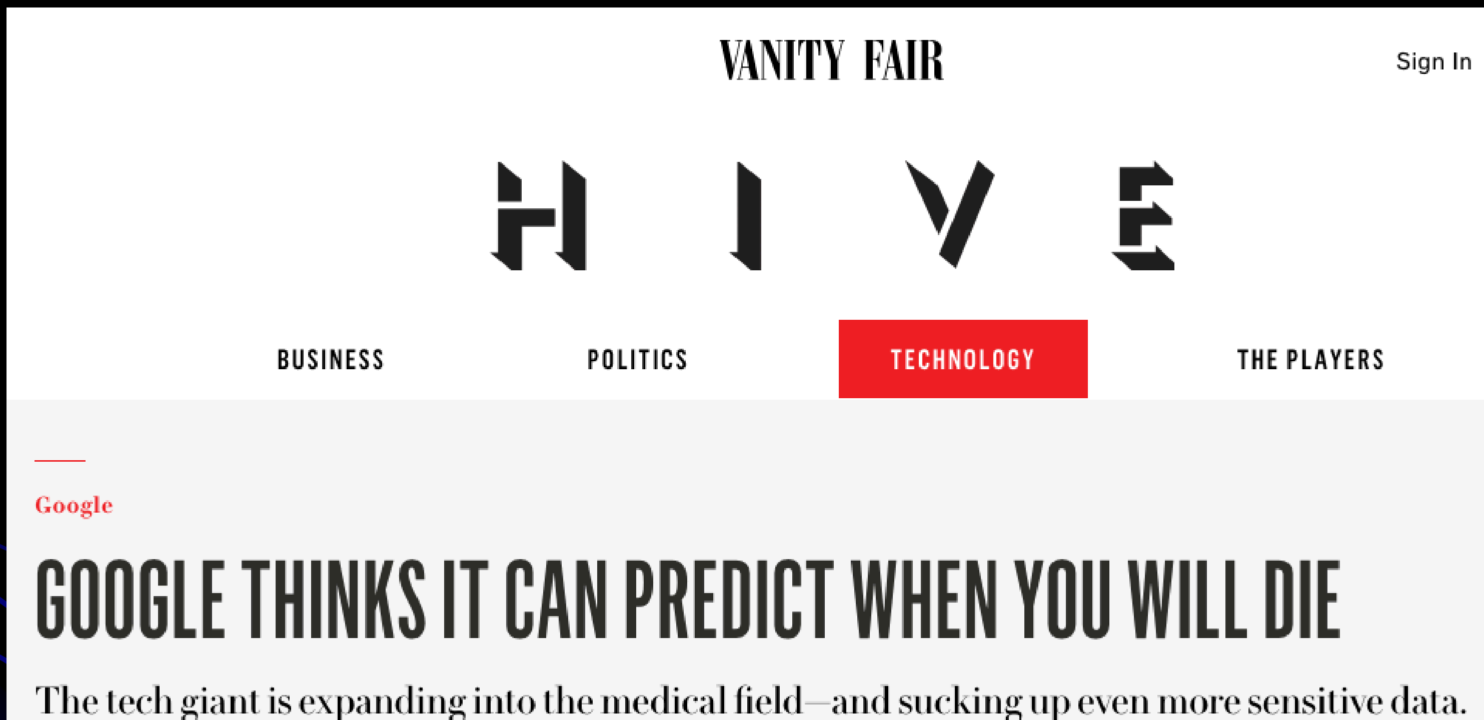
Clinical + ejection fraction

Clinical electronic health records data

Framingham risk score = best current clinical metric

M Samad et al. *JACC: Imaging*. 2018.

# Sidenote: Comparison to Landmark Results




Google AUC ~ 0.93 for in-hospital mortality in 216,221 patients

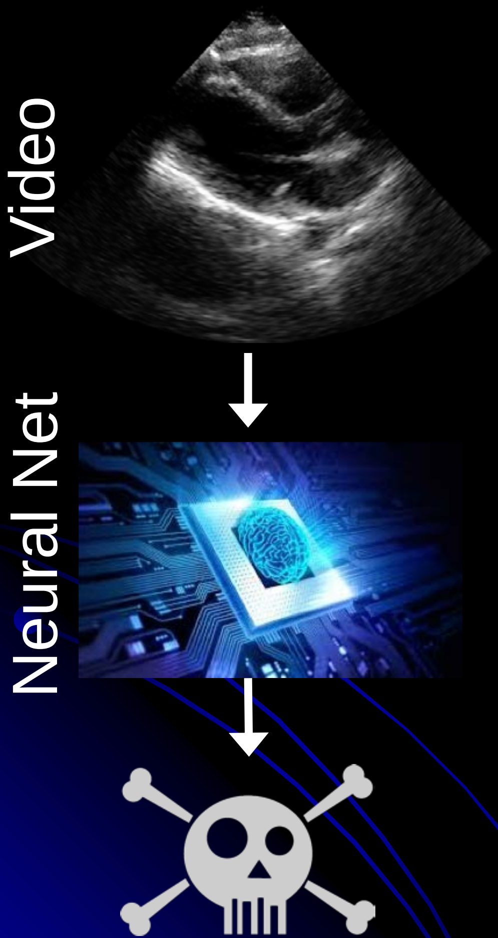
Geisinger AUC ~ 0.91 for 5-year mortality in 171,510 patients

A Rajkomar et al. *NPJ Digital Med*. 2018.

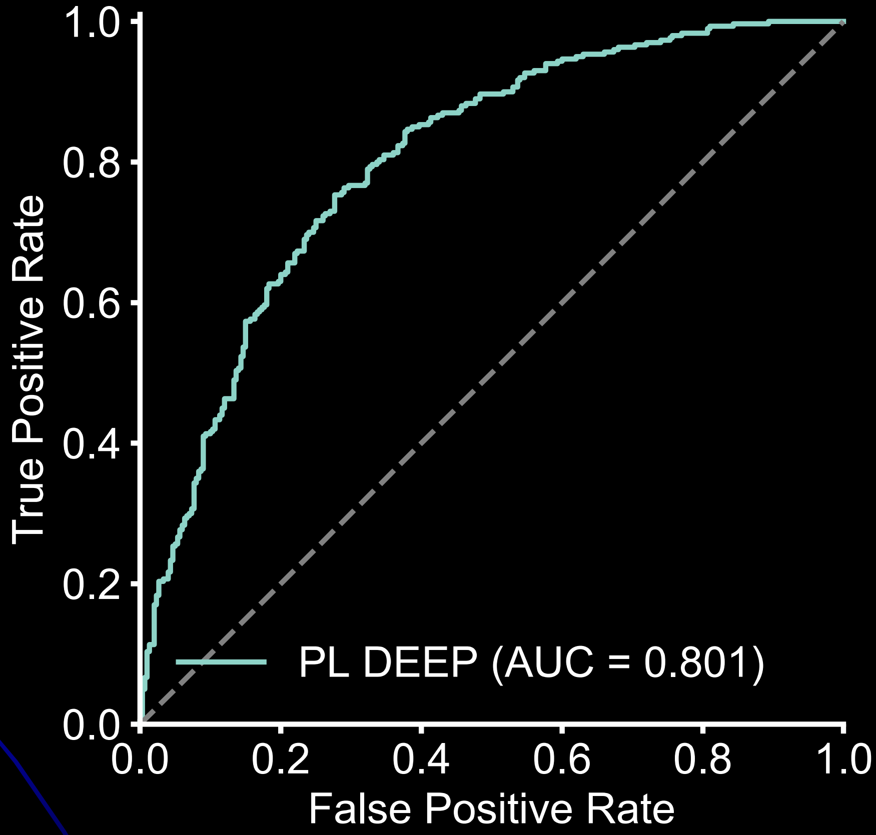
M Samad et al. *JACC: Imaging*. 2018.

# Next: Fully Automated Echocardiography Analysis

A Ulloa PhD




Predict Survival



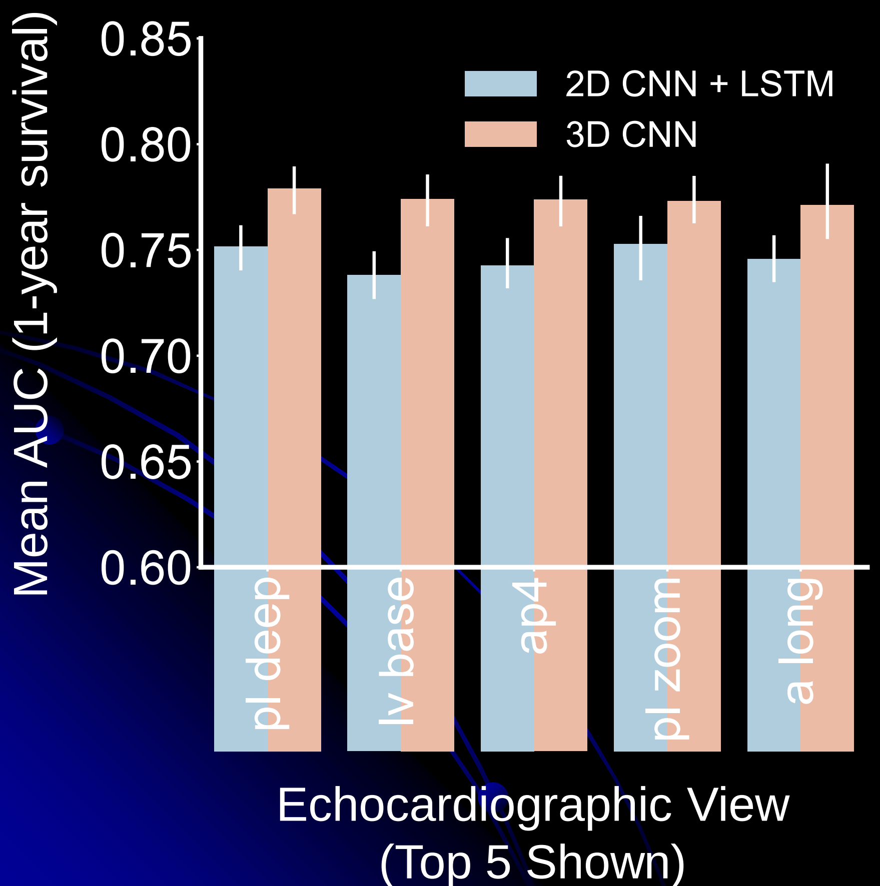

Comparison:
Clinical Score
(Framingham)
AUC = 0.62!!

Note: one of many (>20) videos and clinical data not yet added

https://arxiv.org/abs/1811.10553

# Echocardiography Video Model Selection

723,754 videos (~45 million images)




22 views, 5 folds, 4 models,
[3, 6, 9, 12] months = 1760 Fits

1 Fit ~ 1.5 hours

→ 2,640 GPU hours ~ 3.7 months

= 2 weeks on DGX1

https://arxiv.org/abs/1811.10553

# Can’t my doctor already do that??

## CARDIOLOGIST

- 17 years of practice
- 10 years of indentured servitude resulting in ~$200k of loans
- Interpreted ~35,000 echocardiograms in career
- Needs sleep

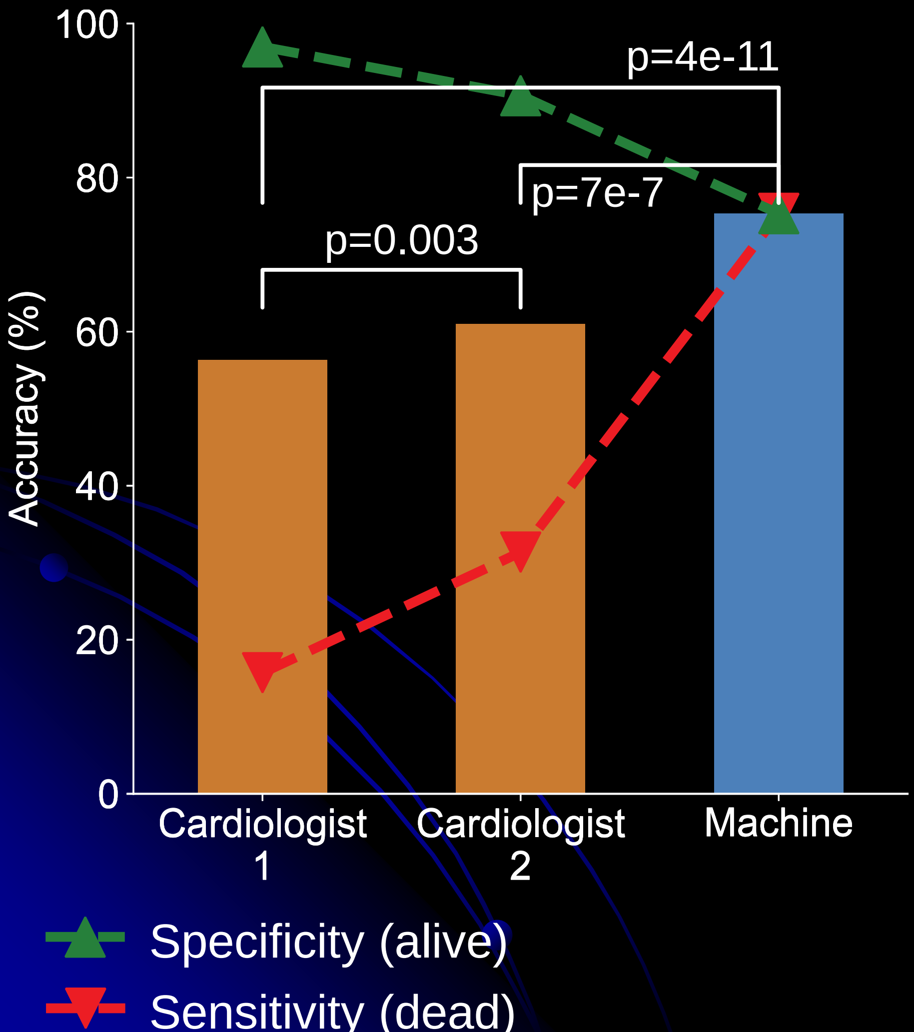
vs

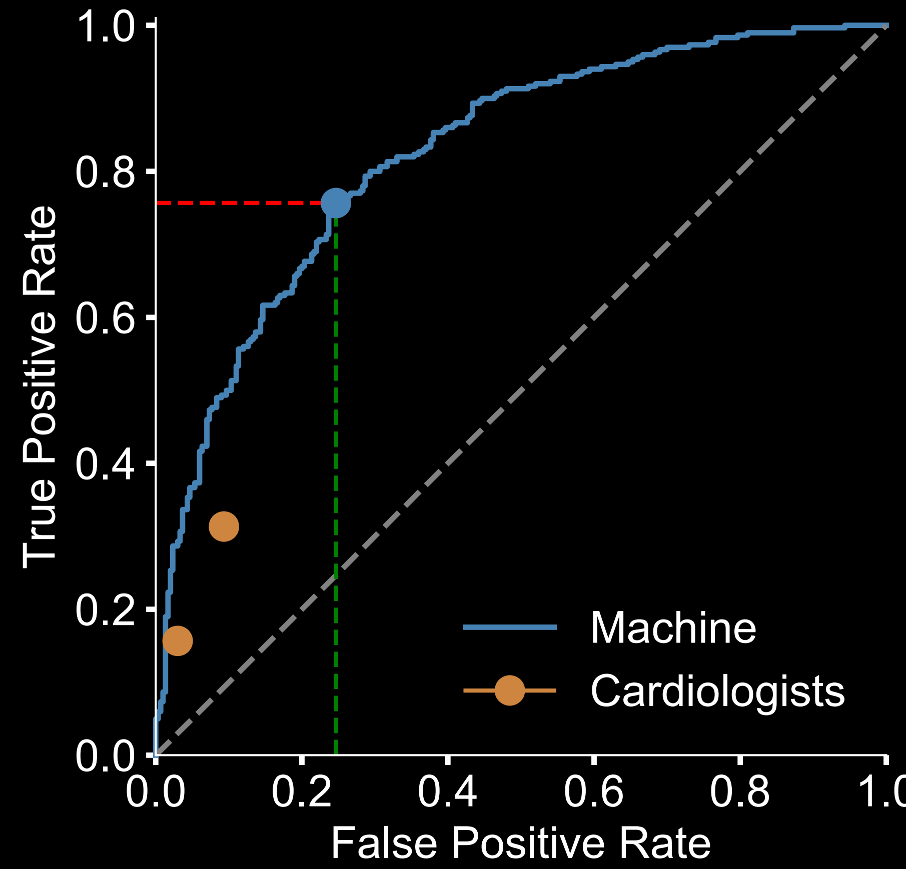
## MACHINE

- Trained on ~30,000 echocardiograms in a week
- “Wants more data”
- Never sleeps

https://arxiv.org/abs/1811.10553

# Machine Beats Cardiologist (n=600)



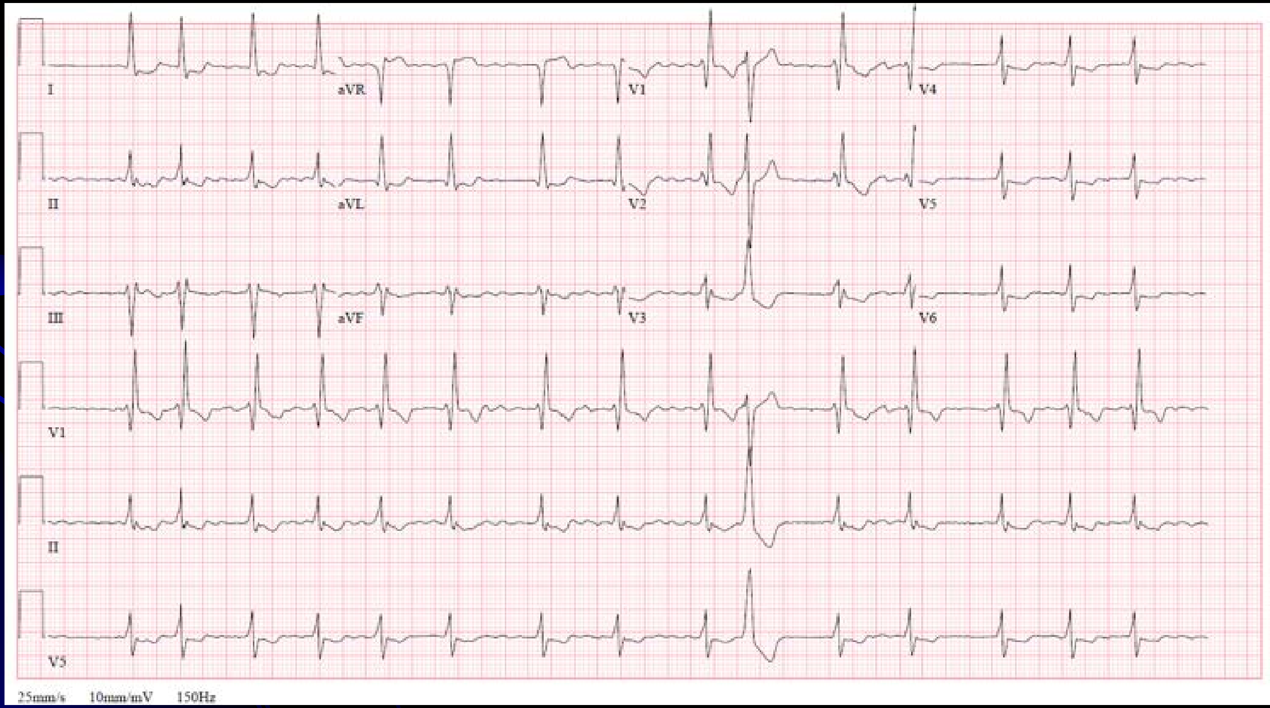

https://arxiv.org/abs/1811.10553

# Outline

- Intro Geisinger, people, data, infrastructure
- Clinical radiology informatics with deep learning
- Machine learning to optimize use of cardiac image data
- Deep learning with 12-lead electrocardiograms
- Machine learning to help manage heart failure
- Multi-modal deep learning in cardiology

# 12-Lead ECG

- Ubiquitous medical test
- ~150 million acquired annually



# 12-Lead ECG

- Deriving *diagnoses* with deep learning

nature medicine

FOCUS | LETTERS

https://doi.org/10.1038/s41591-018-0268-3

Corrected: Publisher Correction

**Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network**

Awni Y. Hannun[1,6*], Pranav Rajpurkar[1,6], Masoumeh Haghpanahi[2,6], Geoffrey H. Tison[3,6], Codie Bourn[2], Mintu P. Turakhia[4,5] and Andrew Y. Ng[1]

Single-lead wearable
91k train
328 test
Beat cardiologists at rhythm dx

**Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram**

Zachi I. Attia[1], Suraj Kapa[1], Francisco Lopez-Jimenez[1], Paul M. McKie[1], Dorothy J. Ladewig[2], Gaurav Satam[2], Patricia A. Pellikka[1], Maurice Enriquez-Sarano[1], Peter A. Noseworthy[1], Thomas M. Munger[1], Samuel J. Asirvatham[1], Christopher G. Scott[3], Rickey E. Carter[4] and Paul A. Friedman[1*]

12-lead
36k train
53k test
Identify ”asymptomatic” LV dysfn

ELSEVIER

Contents lists available at ScienceDirect

Journal of Electrocardiology

journal homepage: www.jecgonline.com

JOURNAL OF Electrocardiology

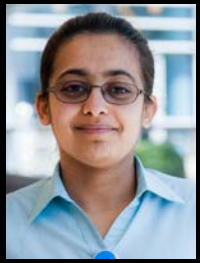
A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation

Check for updates

Stephen W. Smith, MD[a,b,*], Brooks Walsh, MD[c], Ken Grauer, MD[d], Kyuhyun Wang, MD[f], Jeremy Rapin, Ph.D.[h], Jia Li[h], William Fennell, M.D.[e], Pierre Taboulet, M.D.[h,g]

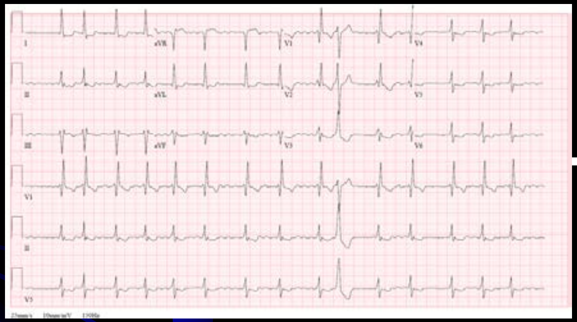
12-lead
Emergency room
100k train
1500 test
Beat old algorithm

# 12-Lead ECG



S Raghunath
PhD

- But what about predictions?



Electrocardiogram

Neural Net



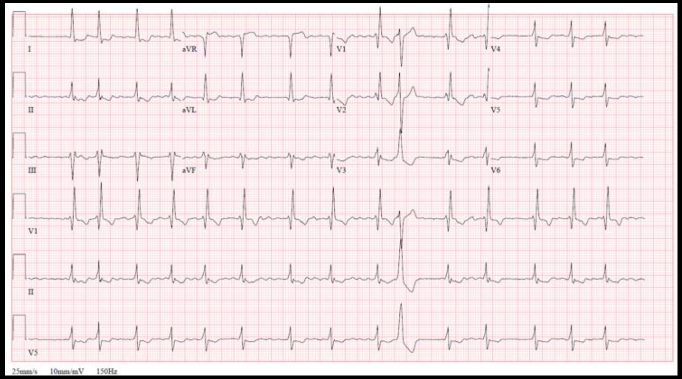
Predict

Raghunath et al. *Nature Med*. In press.

# 12-Lead ECG

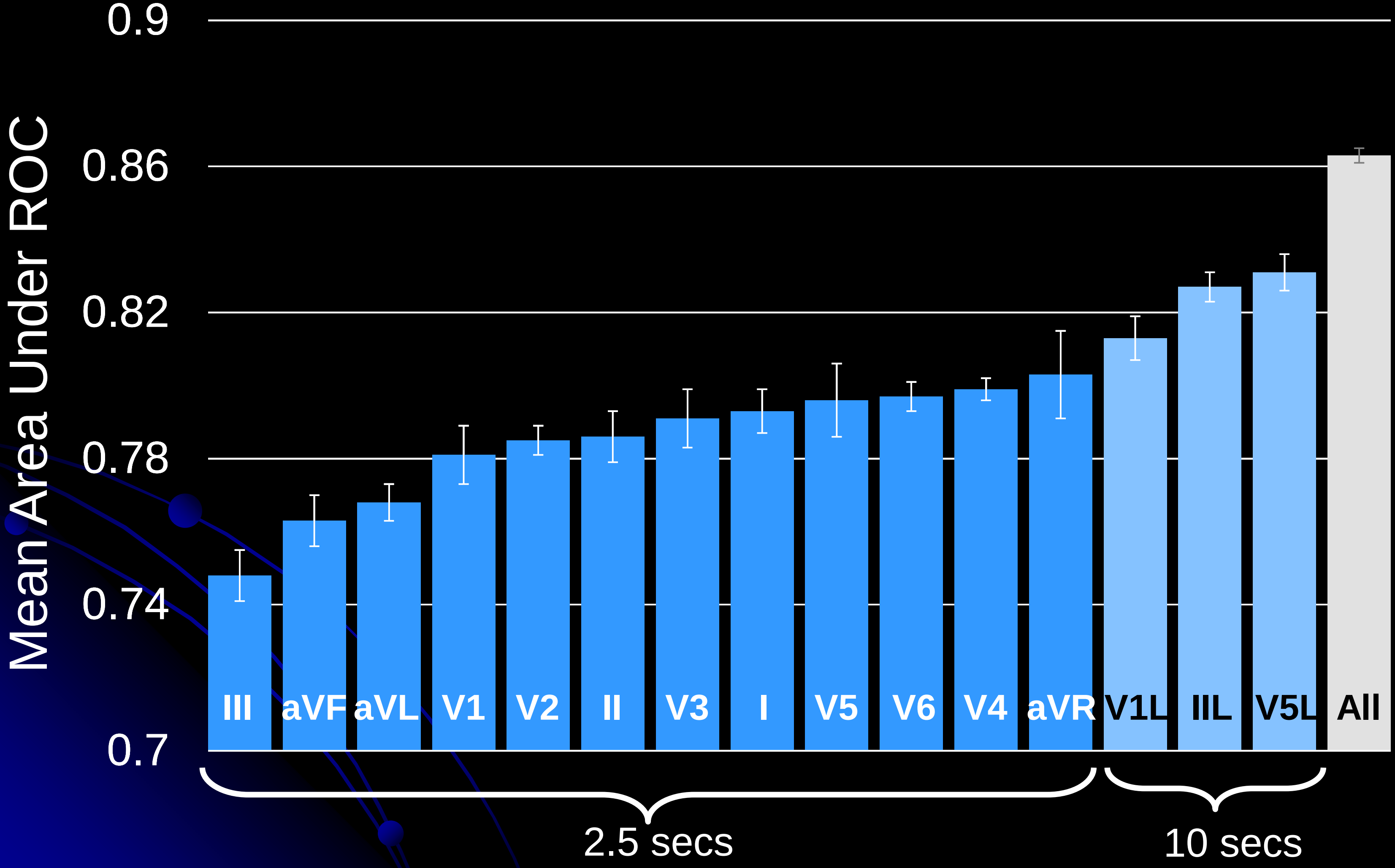
Geisinger ECG dataset

- 1.8 million 12-lead ECGs over 38 years
- ~398k patients
- 250-500Hz raw data
- Linked to outcomes (death registries, clinical events, etc)



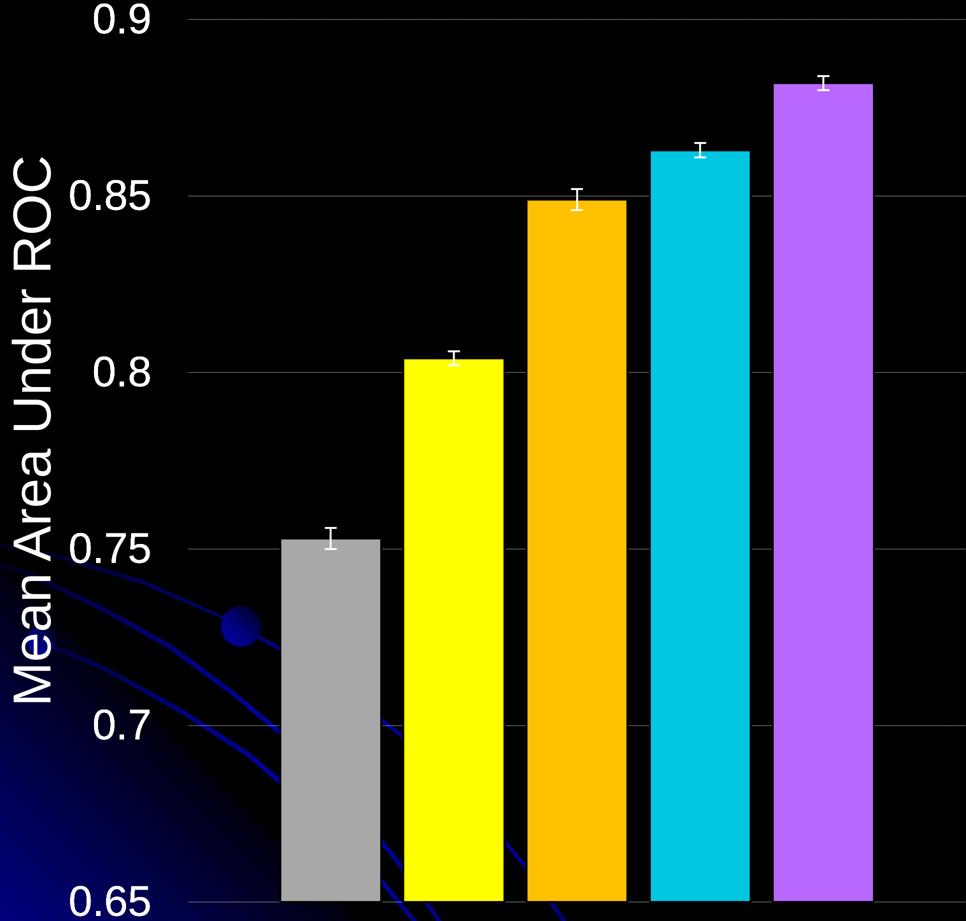

Raghunath et al. *Nature Med*. In press.

# Predicting 1-year Mortality from 1.8 Million ECGs




Raghunath et al. *Nature Med*. In press.

# Predicting 1-year Mortality from 1.8 Million ECGs




ECG measures:

9 computed measures

- QRS duration, PR interval, ventricular rate …

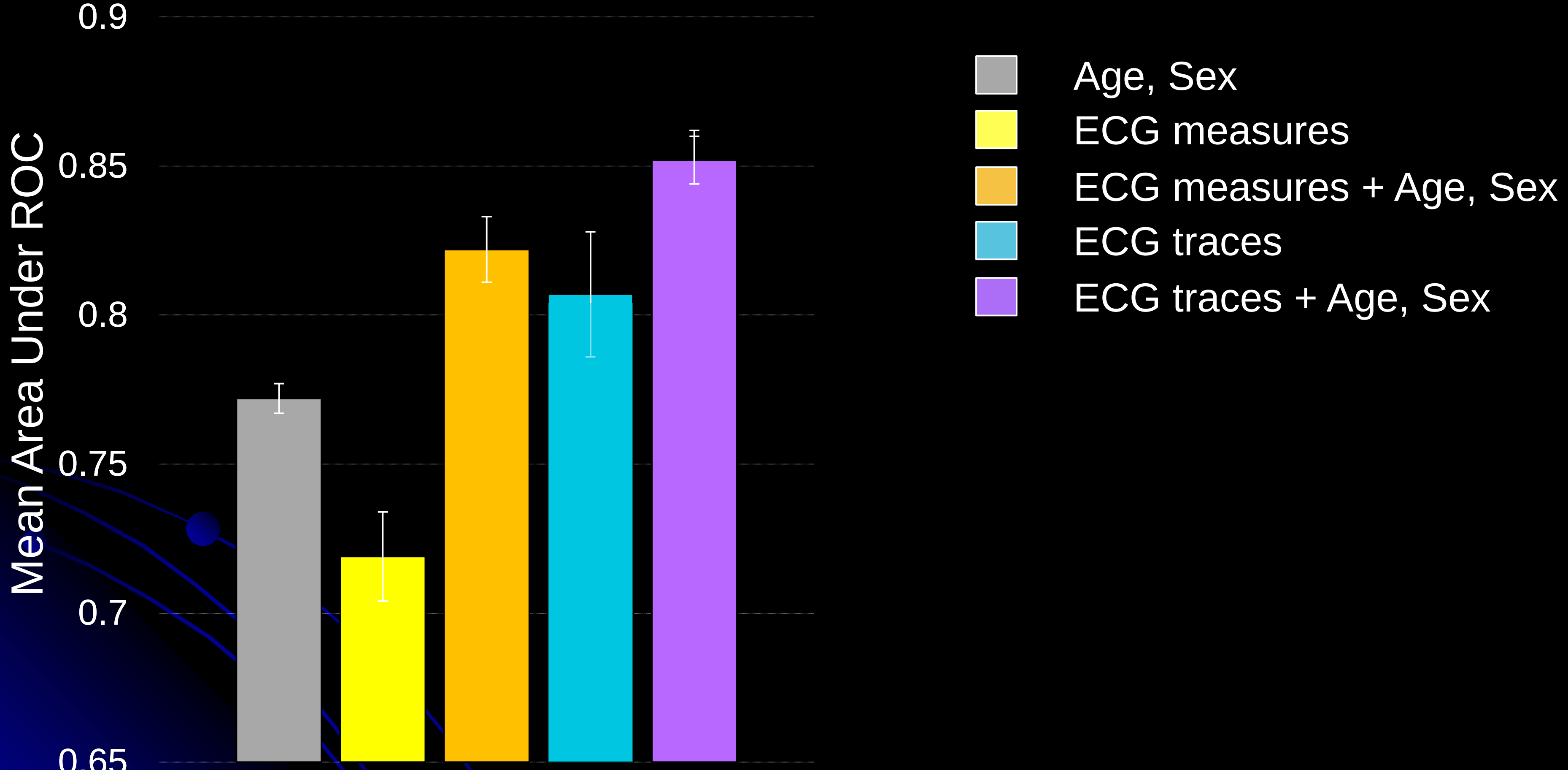
30 pattern labels

- Atrial fibrillation, left bundle branch block …

*Error bars: standard deviation of 5 folds*

Raghunath et al. *Nature Med*. In press.

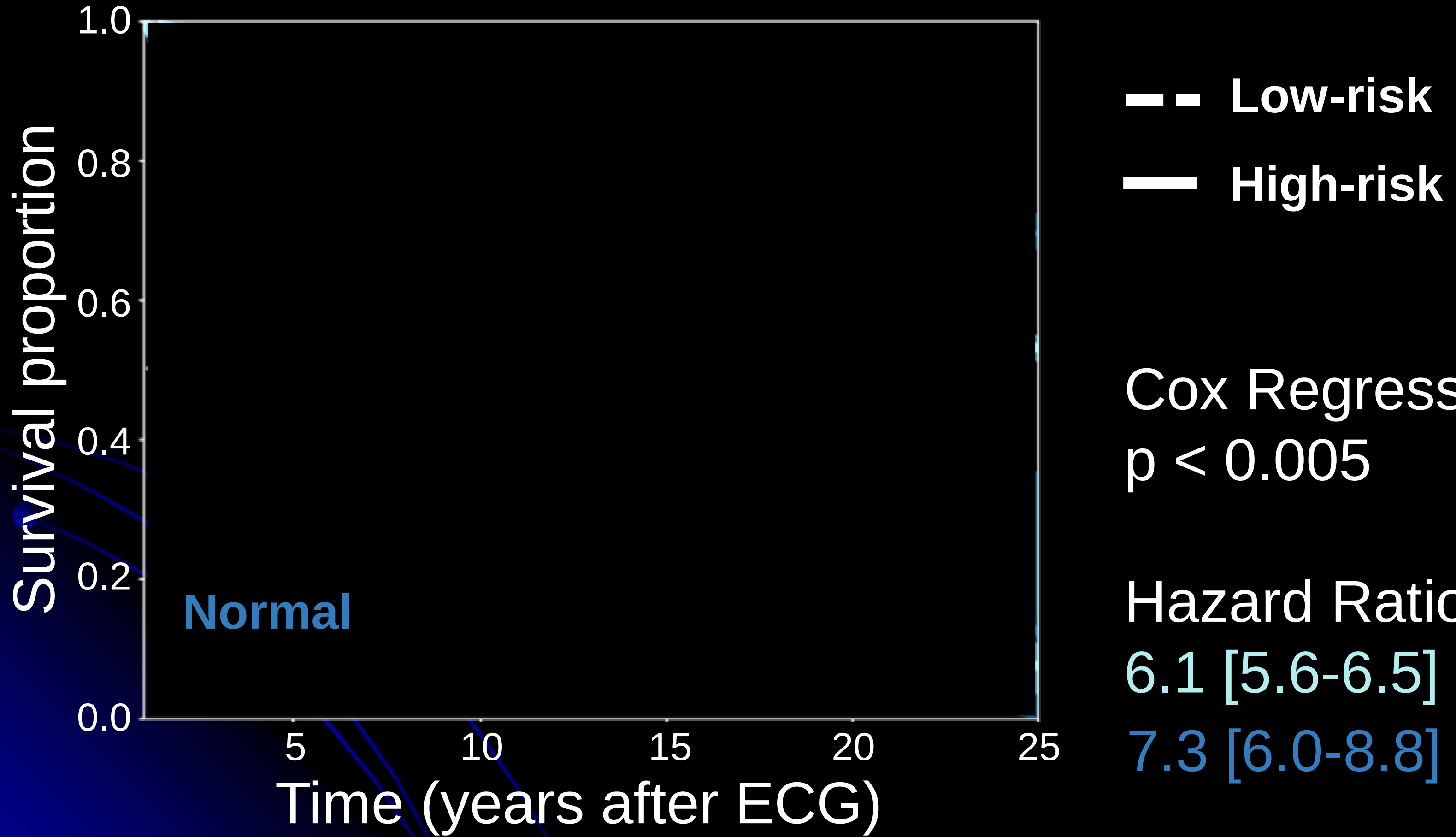
# Can a Cardiologist Do This? 100k "*NORMAL*" ECGs




*Error bars: standard deviation of 5 folds*

Raghunath et al. *Nature Med*. In press.

# Stratification of Predicted Groups in 1.8 Million ECGs



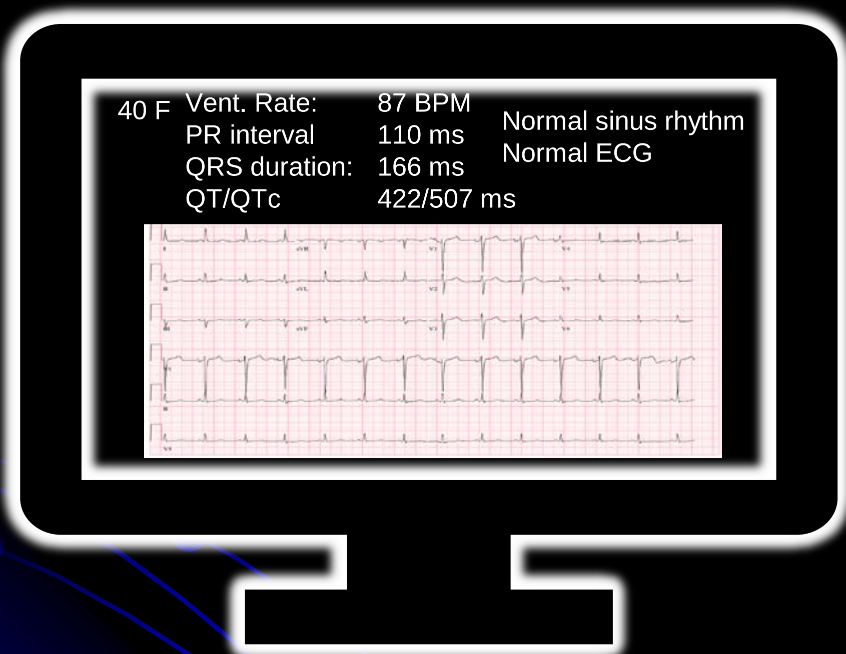

Cox Regression
$p < 0.005$

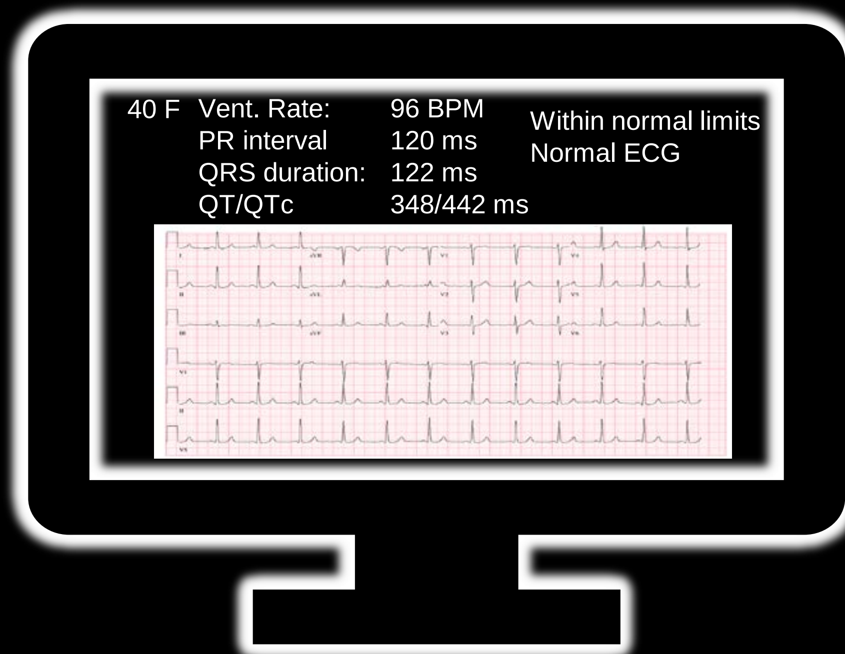
Hazard Ratios:
6.1 [5.6-6.5] (All)
7.3 [6.0-8.8] (Normal)

Note: Prediction is for 1-year mortality!

Raghunath et al. *Nature Med*. In press.

# Surely My Cardiologist Can Find Features Predictive of Mortality?

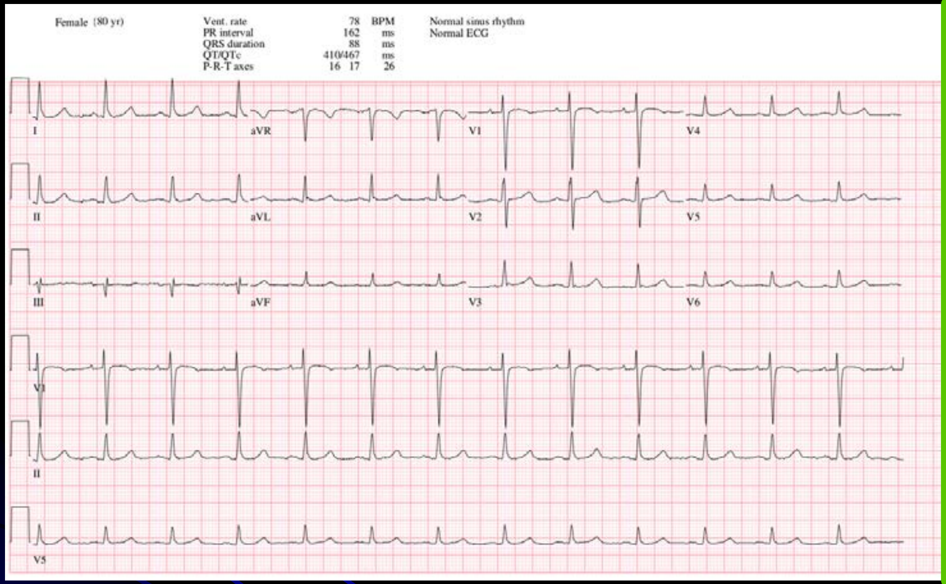


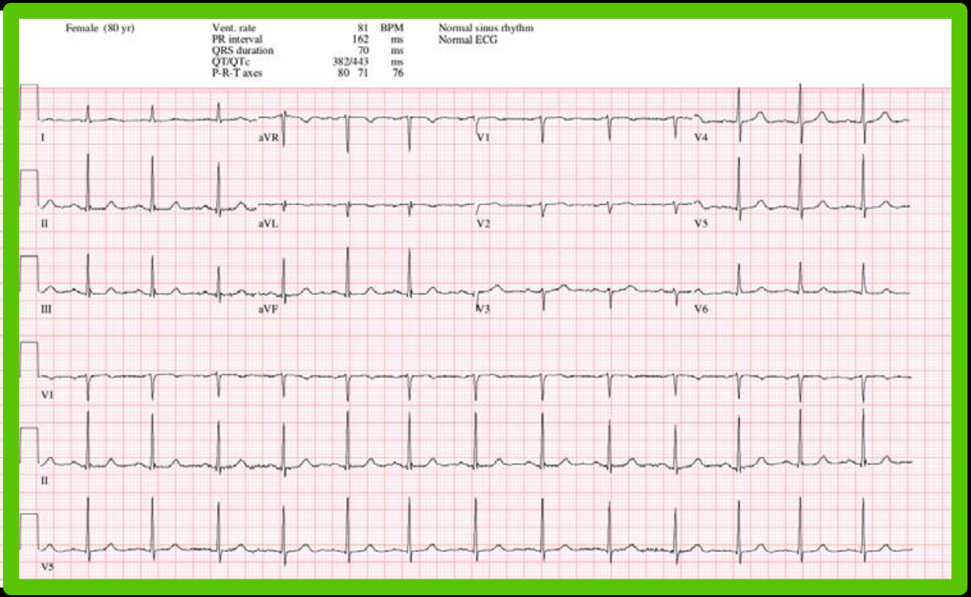




**Blinded survey: Which patient will survive >1 year?**
**Paired sets: one true positive, one true negative.**

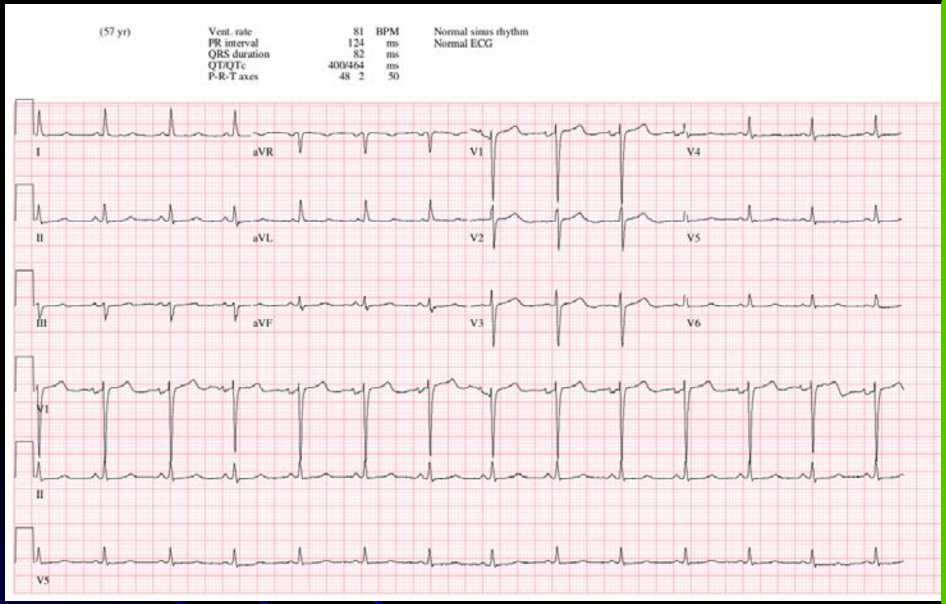
Raghunath et al. *Nature Med*. In press.

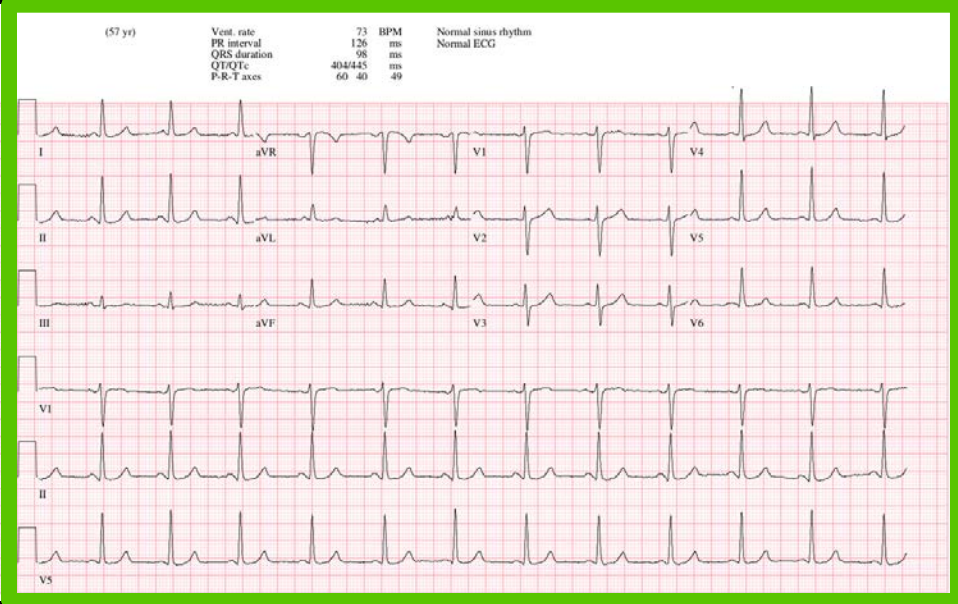
# Which One Survived?



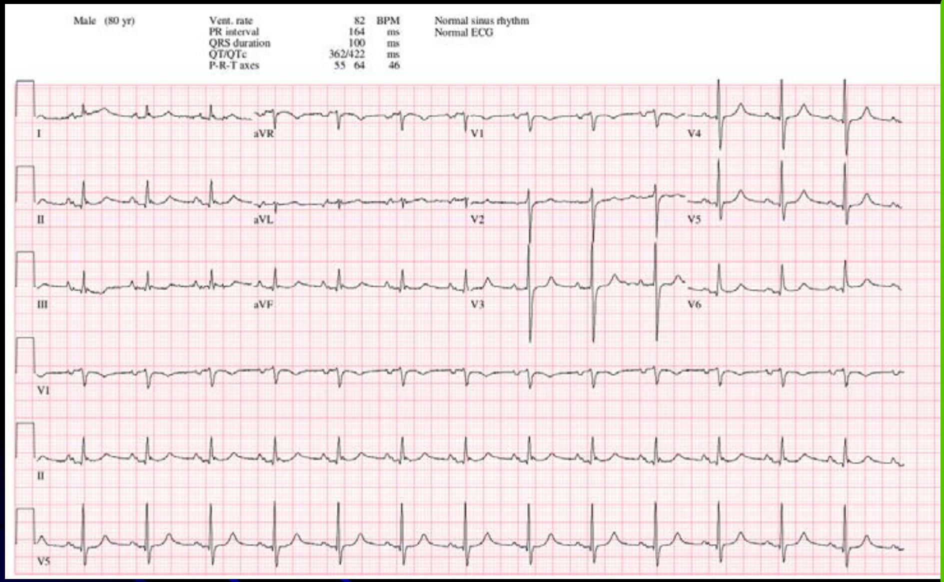



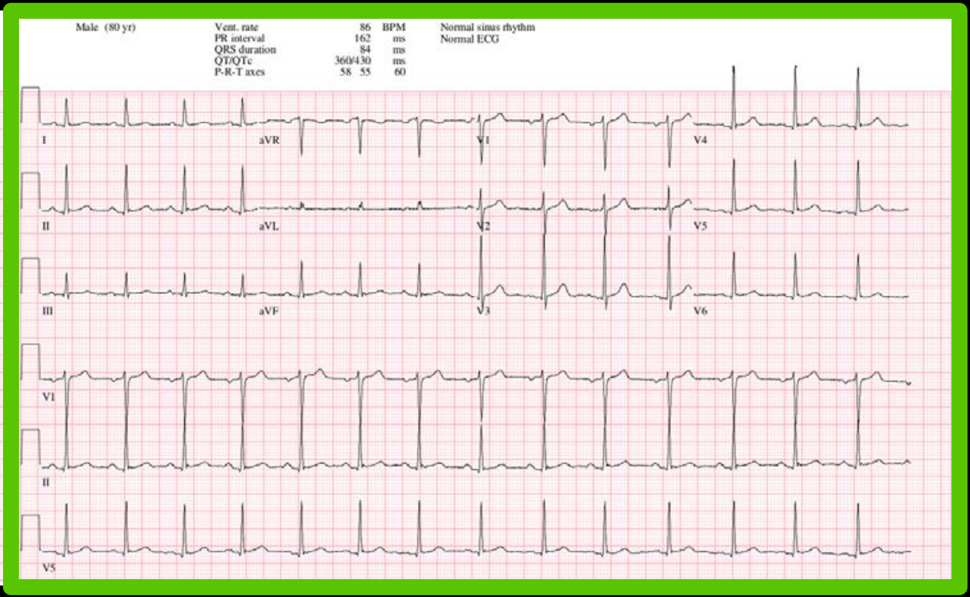
# Which One Survived?



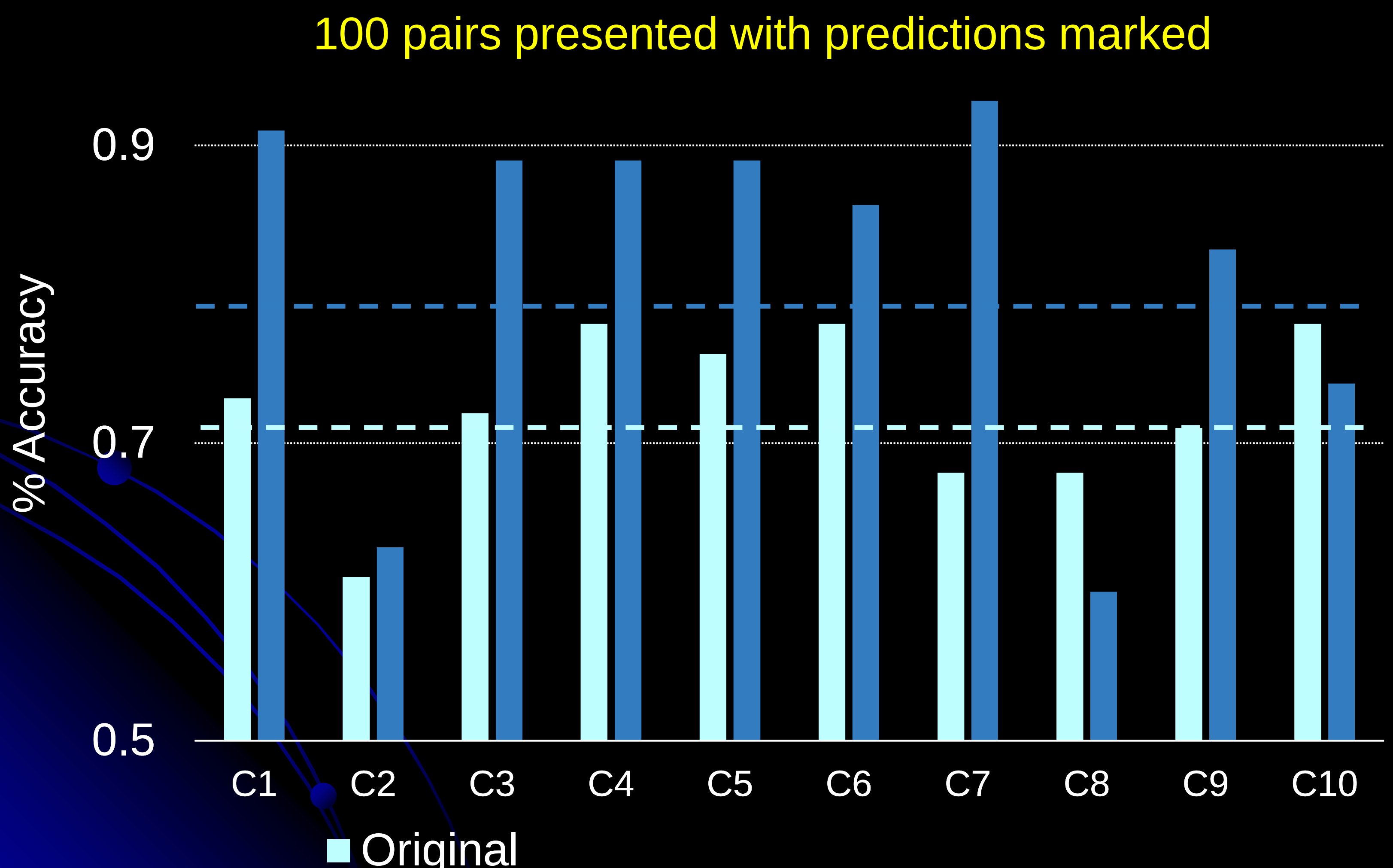



# Which One Survived?



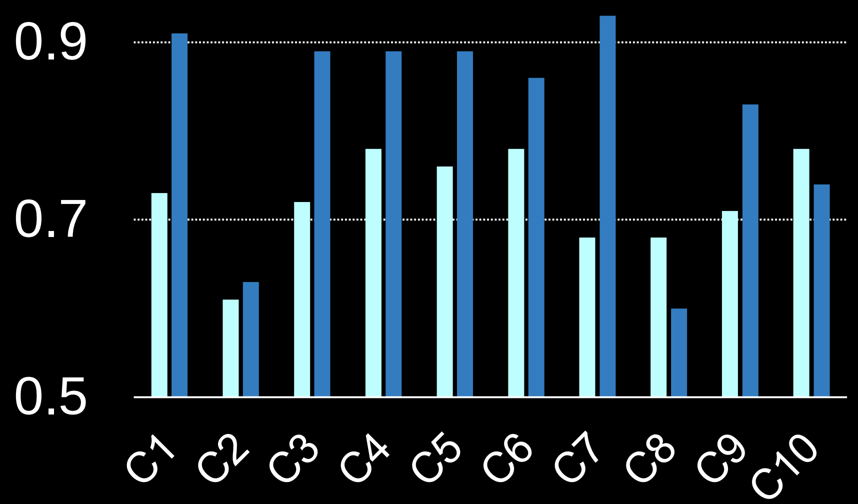



# Can 10 Different Cardiologists Learn from the Model?



# Can 10 Different Cardiologists Learn from the Model?

- Cardiologists can see some features predictive of 1-yr mortality (less than half)
- Can learn from the model
  - 8 /10 cardiologists improved
- Reported features learned:
  - Higher heart rate
  - Poor ECG baseline
  - Slight left atrial enlargement



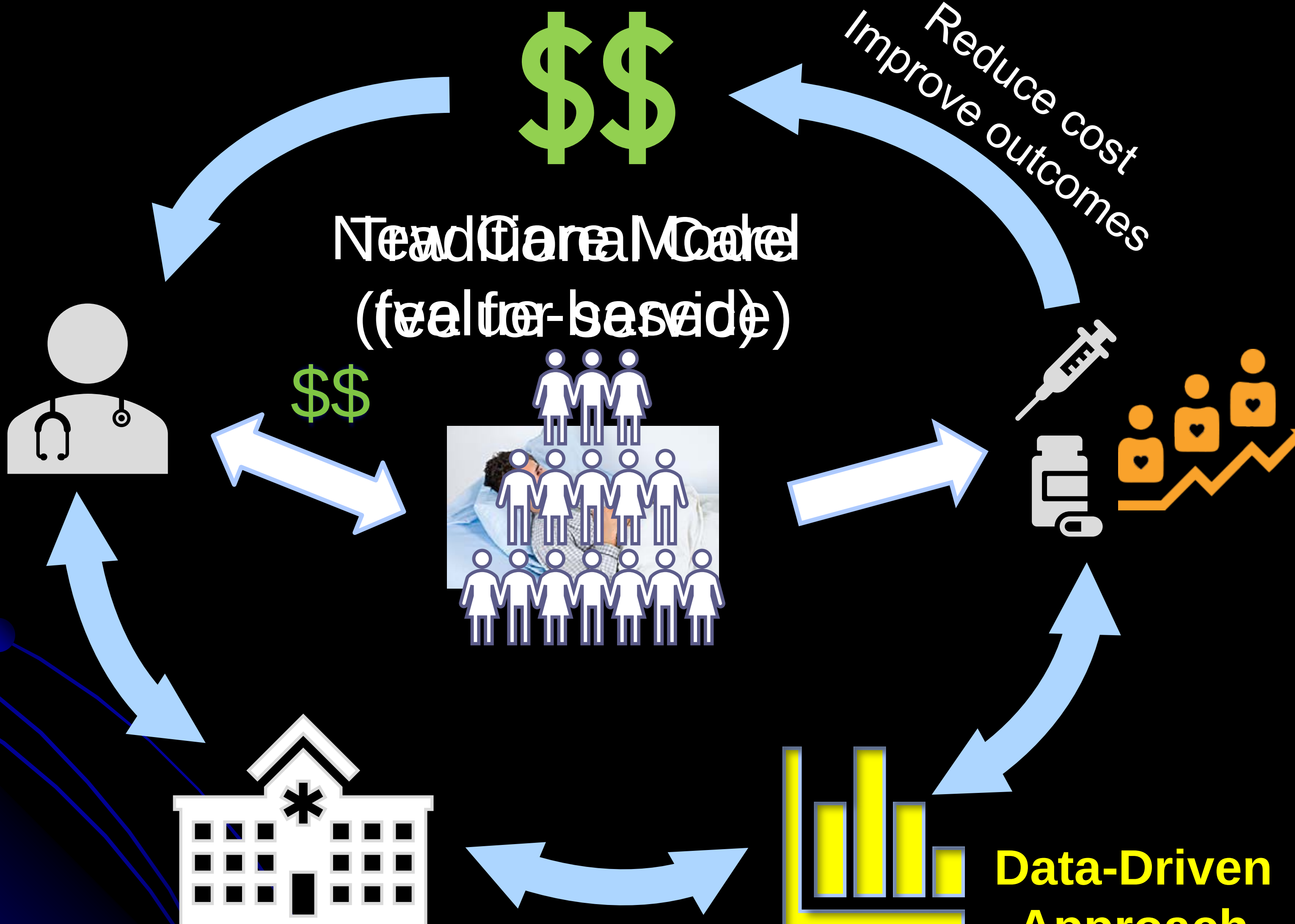

Raghunath et al. *Nature Med*. In press.

# Outline

- Intro Geisinger, people, data, infrastructure
- Clinical radiology informatics with deep learning
- Machine learning to optimize use of cardiac image data
- Deep learning with 12-lead electrocardiograms
- Machine learning to help manage heart failure
- Multi-modal deep learning in cardiology

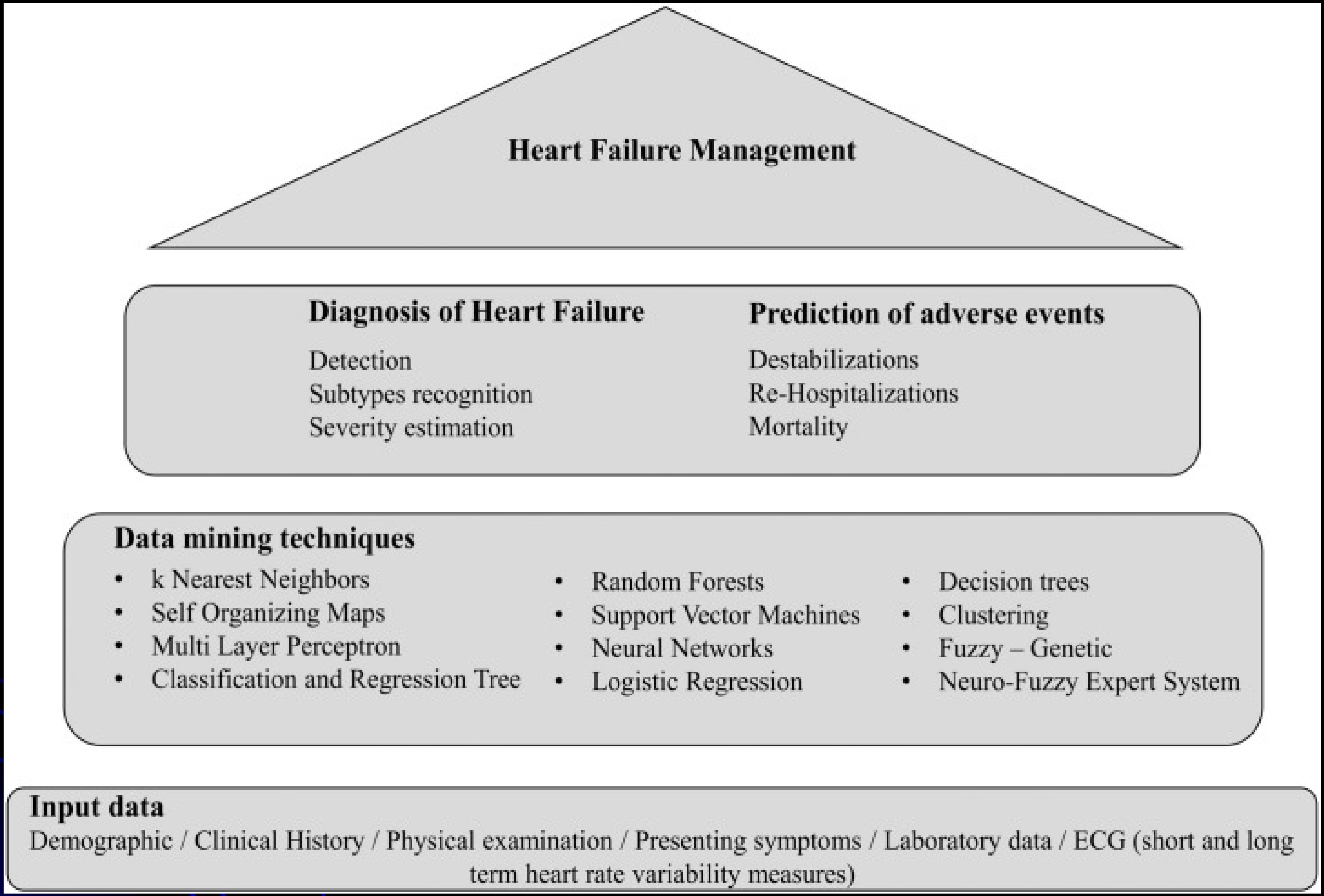
# Heart Failure Population Management

L Jing
PhD




Geisinger

Jing et al. *JACC: HF*. In press.




- Previous studies using machine learning show promising results (AUC ~0.6 – 0.9)

Tripoliti *et al, Comput Struct Biotechnol J.* 2017; 15: 26–47.

# Question: How do we drive action with machine learning models?

Answer:

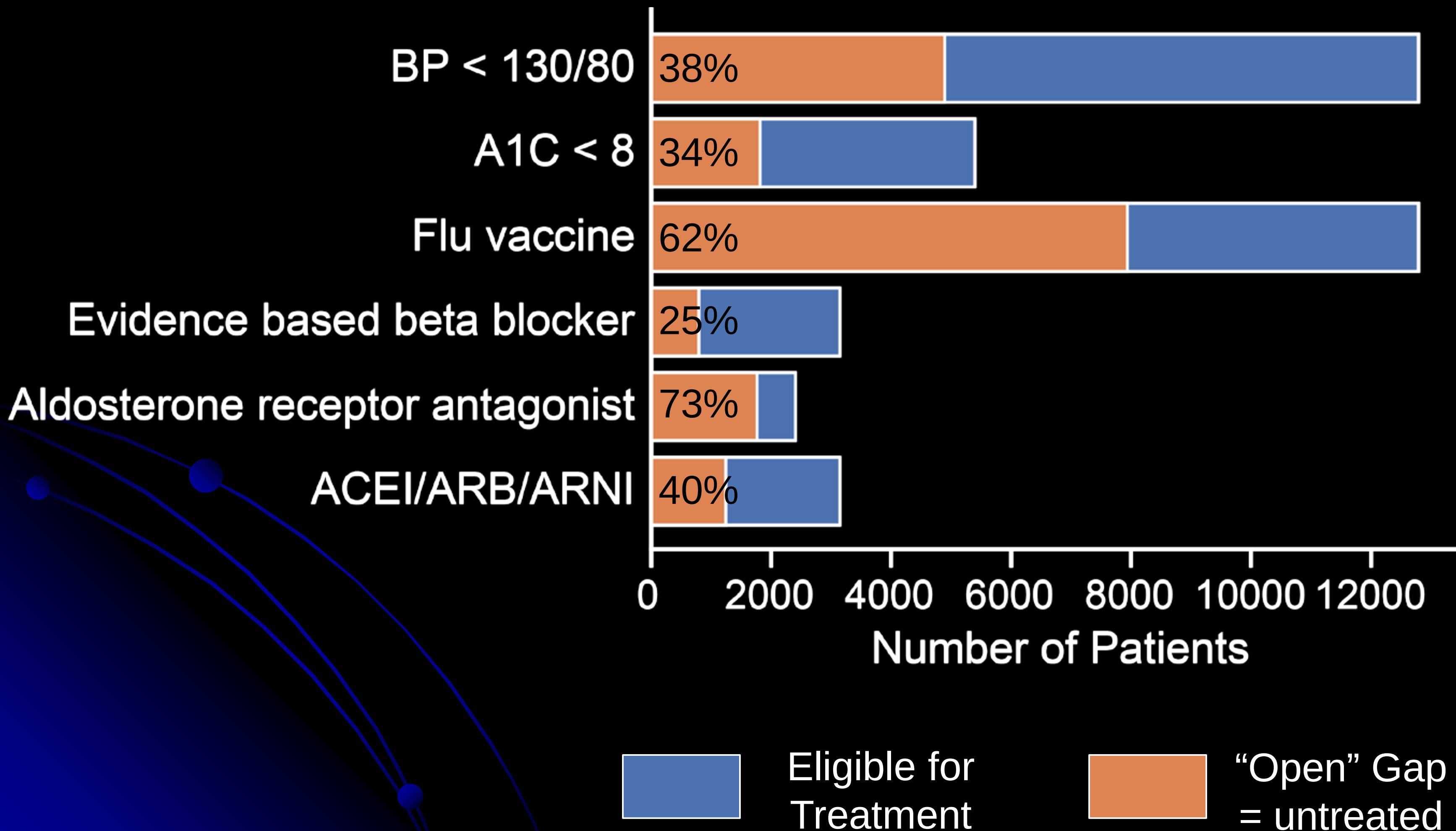
1. Add actionable “care gap” variables
   - Example: Flu shot given

2. Model effect of care gaps in retrospect

3. Predict effect of closing gaps prospectively

Jing et al. *JACC: HF*. In press.

# How Common are the "Open" Care Gaps?




Jing et al. *JACC: HF*. In press.

## 268,096 episodes from ~1.5 million encounters in 26,524 HF patients from Geisinger Electronic Health Records

Demographics (5)

Vital signs (3)

Diagnostic codes (89)

Labs (15)

Medications (2)

Care gaps (6)

Echocardiographic measurements (44)

ECG measures (9) and patterns (32)

Machine Learning Classifiers

- Logistic Regression (LR)
- Random Forest (RF)
- XGBoost (XGB)

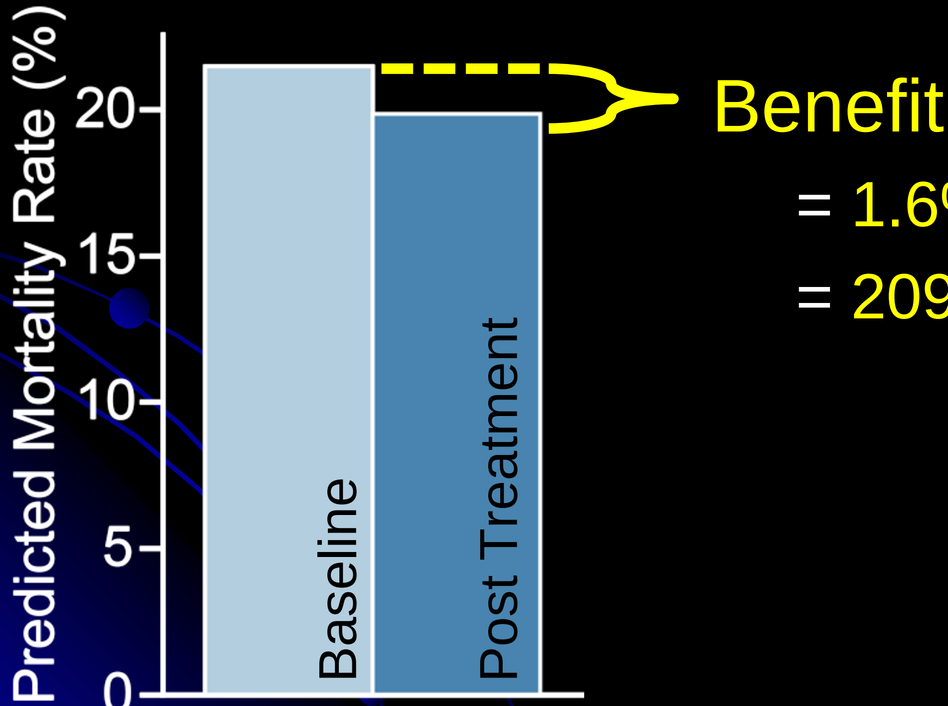
Train-by-year cross-validation

One-year All-cause mortality

**Good model performance**: area under the curve (AUC) = 0.77 and 0.78 (hold-out set)

# Predicting Effect of Closing Care Gaps

- “Treat” patients by closing care gaps via simulation
  - Re-calculate risk

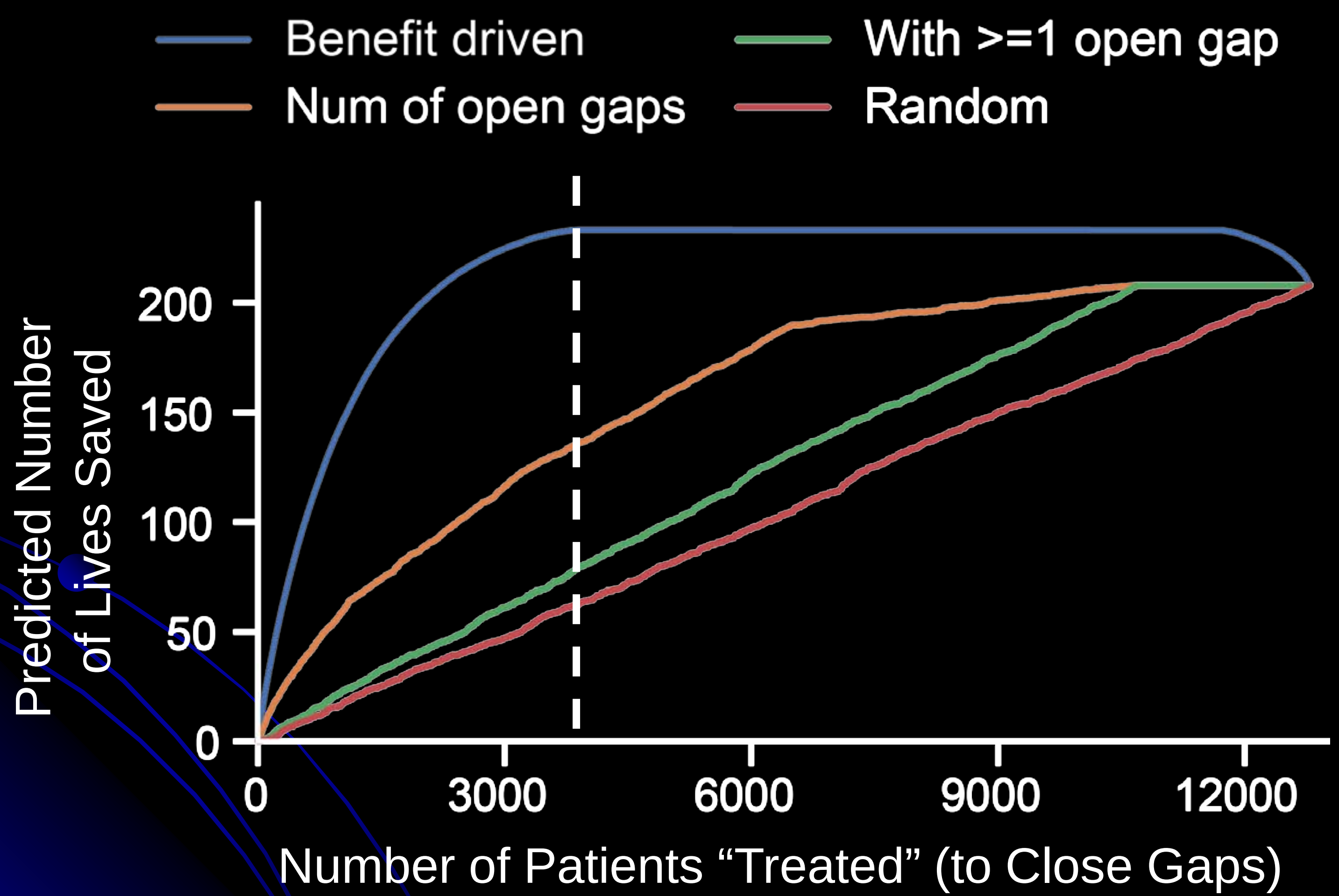



= 1.6% mortality risk reduction

= 209 lives saved

Jing et al. *JACC: HF*. In press.

# Optimized Care Gaps Team Deployment




Jing et al. *JACC: HF*. In press.

# AIM HI Trial:
# Artificial Intelligence Managed Heart Failure Intervention

*ClinicalTrials.gov Identifier: NCT03804606*

| | MTM | Predicted Risk | Predicted Benefit |
|---|---|---|---|
| **Group 1** | No | High | High |
| **Group 2** | Yes | High | High |
| **Group 3** | Yes | High | Low |

a (Group 1 vs Group 2)
b (Group 2 vs Group 3)

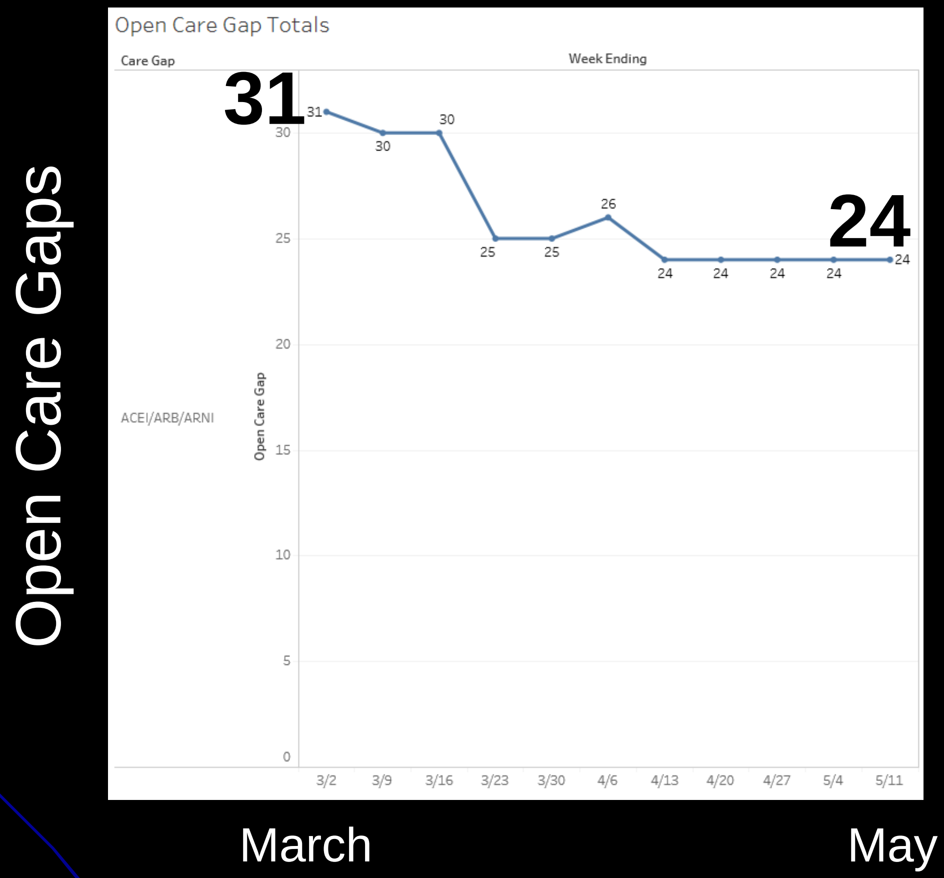
MTM = Medication Therapy Management Pharmacist

a Effect of MTM intervention for patients predicted to be high risk and high benefit?

b Can the model discriminate between patients with high vs low benefit?

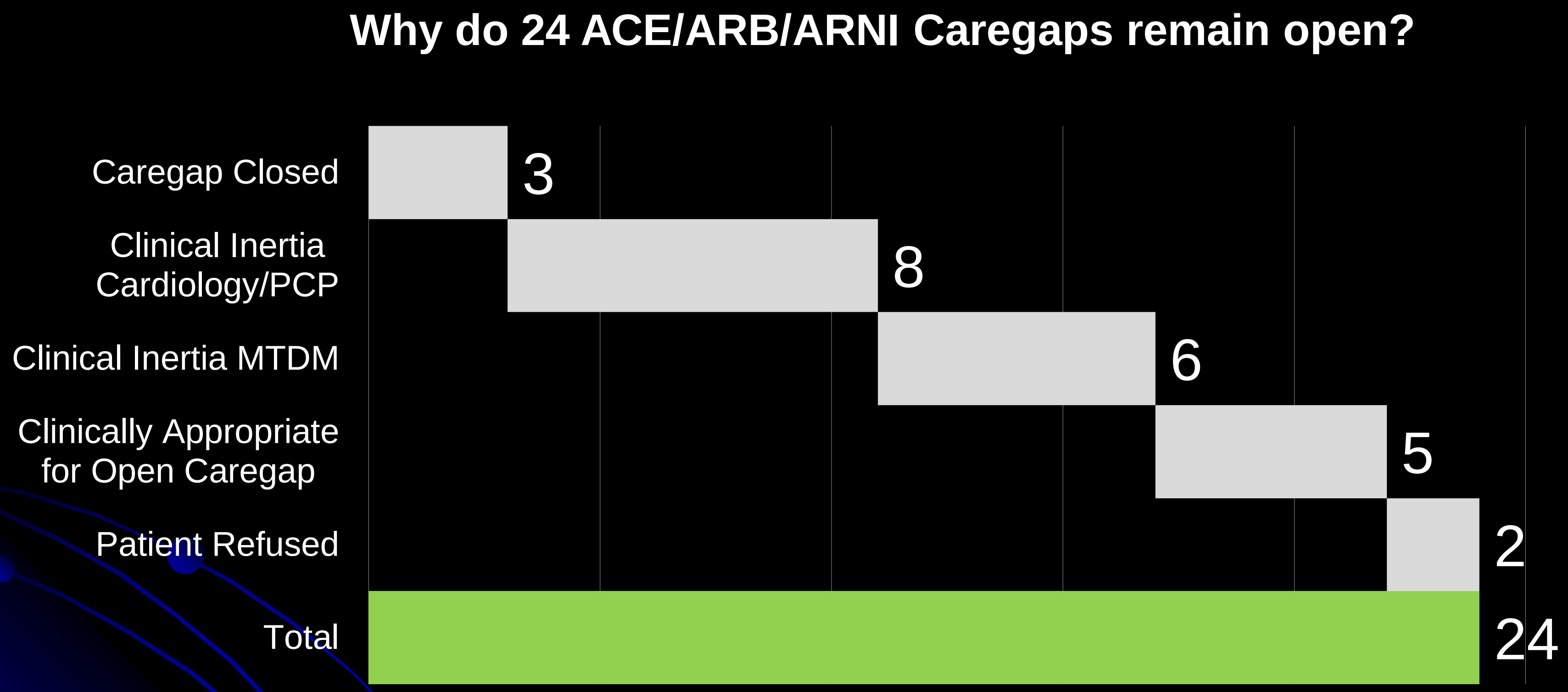
# First 100 Patients ~March 2019: It’s Not Easy.

# Care Gap Closure: ACE/ARB/ARNI




**Eligibility -** Heart Failure with LVEF < 40%
**Contraindications -** No due to Pregnancy, History of angioedema, Hypotension - SBP < 100 mm Hg (ave. of up to last 5 SBP within last 6 months), Serum creatinine > 2 (in any of preceding 3 readings), Potassium > 5 (in any of preceding 3 readings), Bilateral renal artery stenosis, Hemodialysis, Listed allergy, ACEI/ARB contraindicated on the problem list.

# Care Gap Closure: ACE/ARB/ARNI

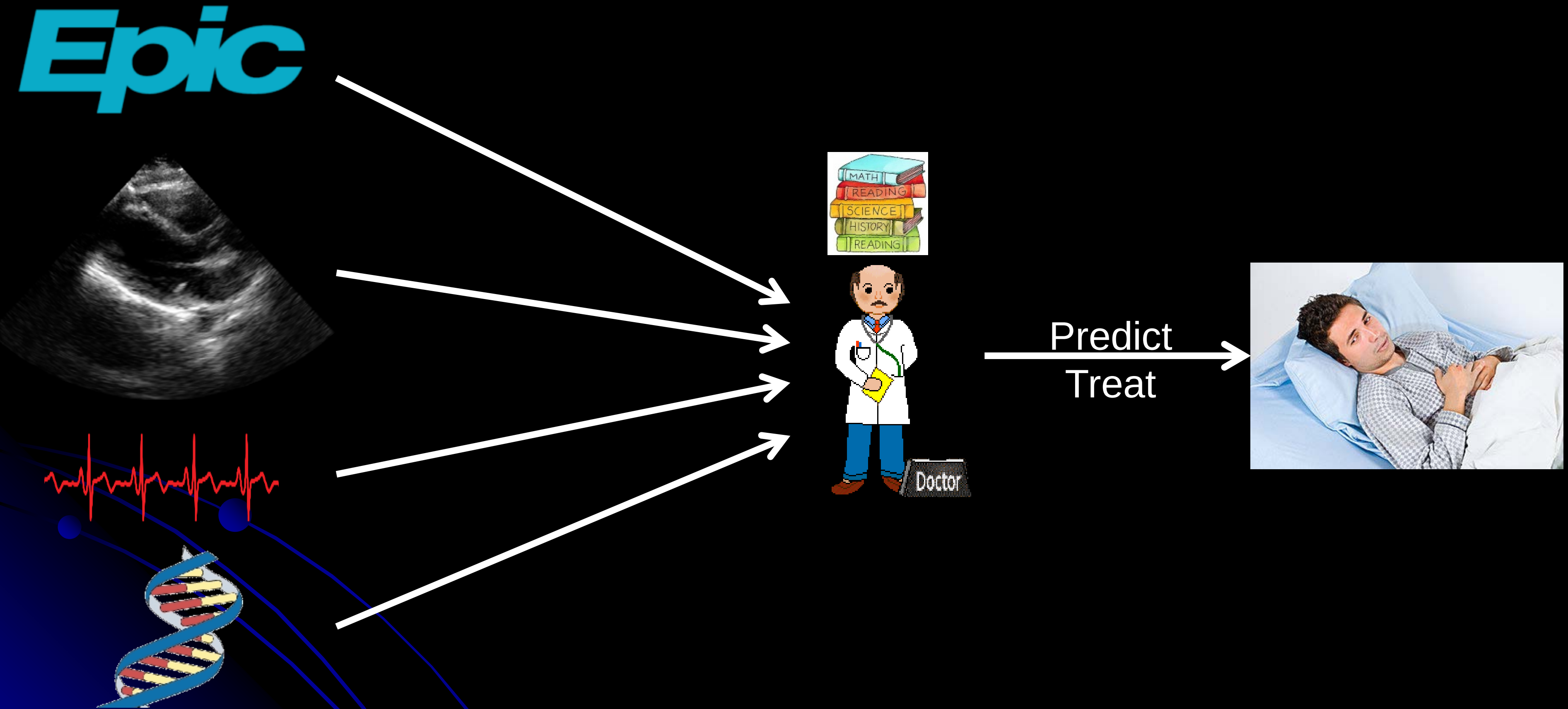



**Eligibility -** Heart Failure with LVEF < 40%
**Contraindications -** No due to Pregnancy, History of angioedema, Hypotension - SBP < 100 mm Hg (ave. of up to last 5 SBP within last 6 months), Serum creatinine > 2 (in any of preceding 3 readings), Potassium > 5 (in any of preceding 3 readings), Bilateral renal artery stenosis, Hemodialysis, Listed allergy, ACEI/ARB contraindicated on the problem list.

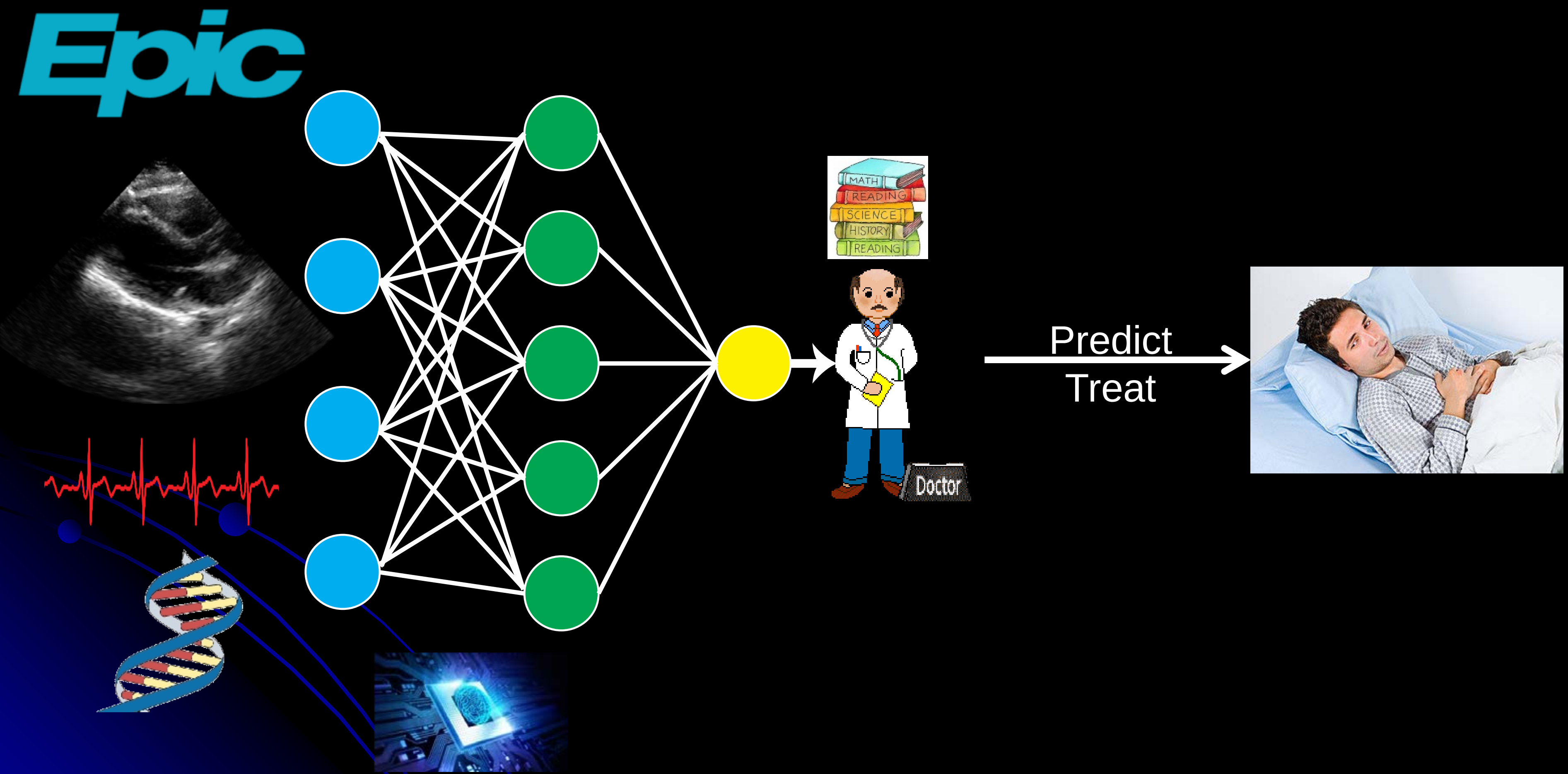
# Outline

- Intro Geisinger, people, data, infrastructure
- Clinical radiology informatics with deep learning
- Machine learning to optimize use of cardiac image data
- Deep learning with 12-lead electrocardiograms
- Machine learning to help manage heart failure
- Multi-modal deep learning in cardiology

# Cardiology in the PAST



# Cardiology in the FUTURE

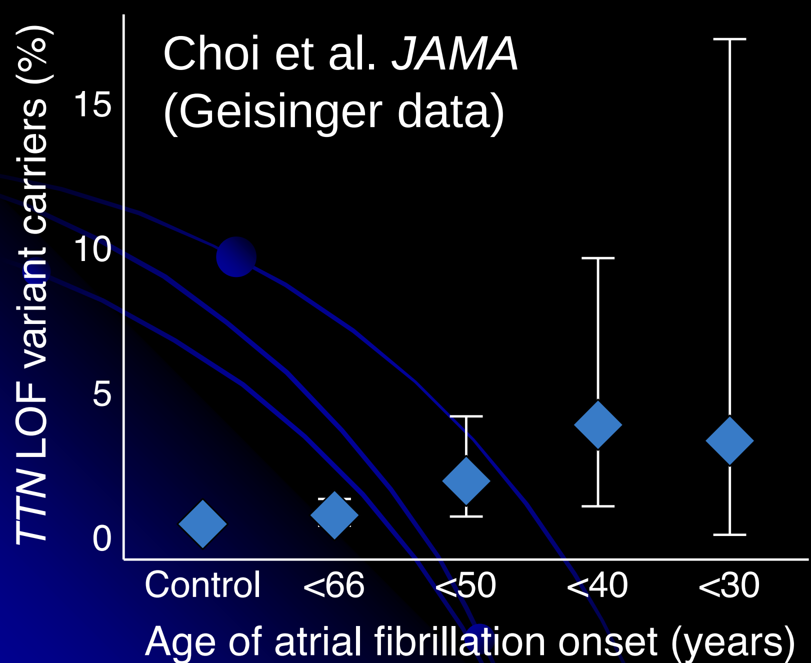



Neural Network

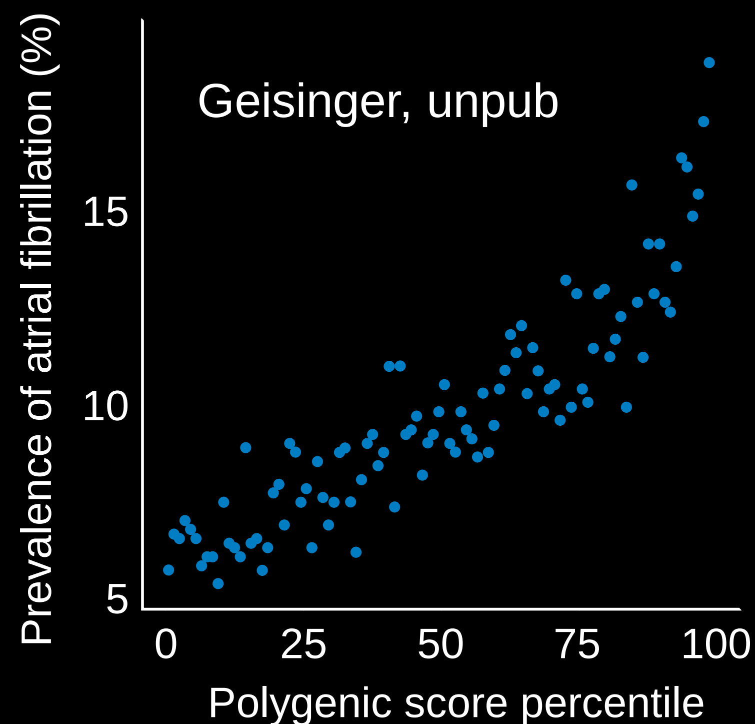
# Genomic Inputs to Machine Learning Models

C Haggerty
PhD

## Rare Variants



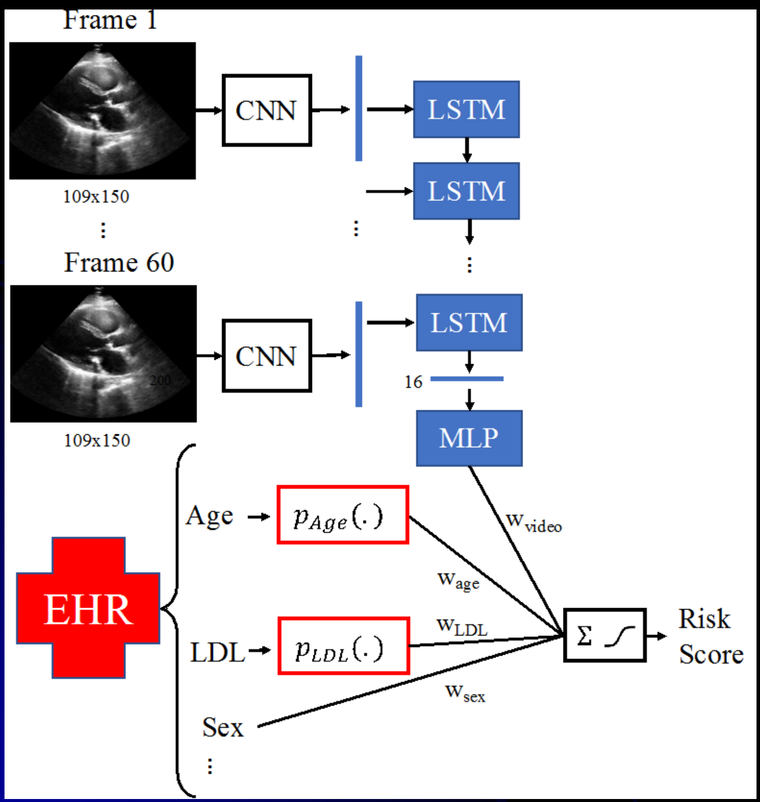

## Common Variation

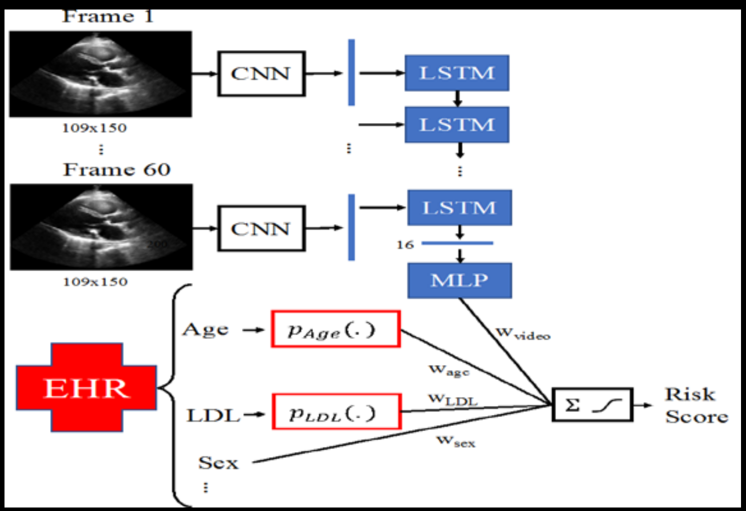


# Interpretable Neural Networks for Predicting Mortality Risk using Multi-modal Electronic Health Records

Alvaro E. Ulloa Cerna*†, Marios Pattichis†, *Senior Member, IEEE*, David P. vanMaanen*, Linyuan Jing*, Aalpen A. Patel*‡, Joshua V. Stough**, Christopher M. Haggerty*¶, and Brandon K. Fornwalt*‡¶
*Department of Imaging Science and Innovation, Geisinger, PA 17822 USA
†Department of Electrical and Computer Engineering, University of New Mexico, NM 87106, USA
‡Department of Radiology, Geisinger, PA 17822, USA ¶The Heart Institute, Geisinger, PA 17822, USA

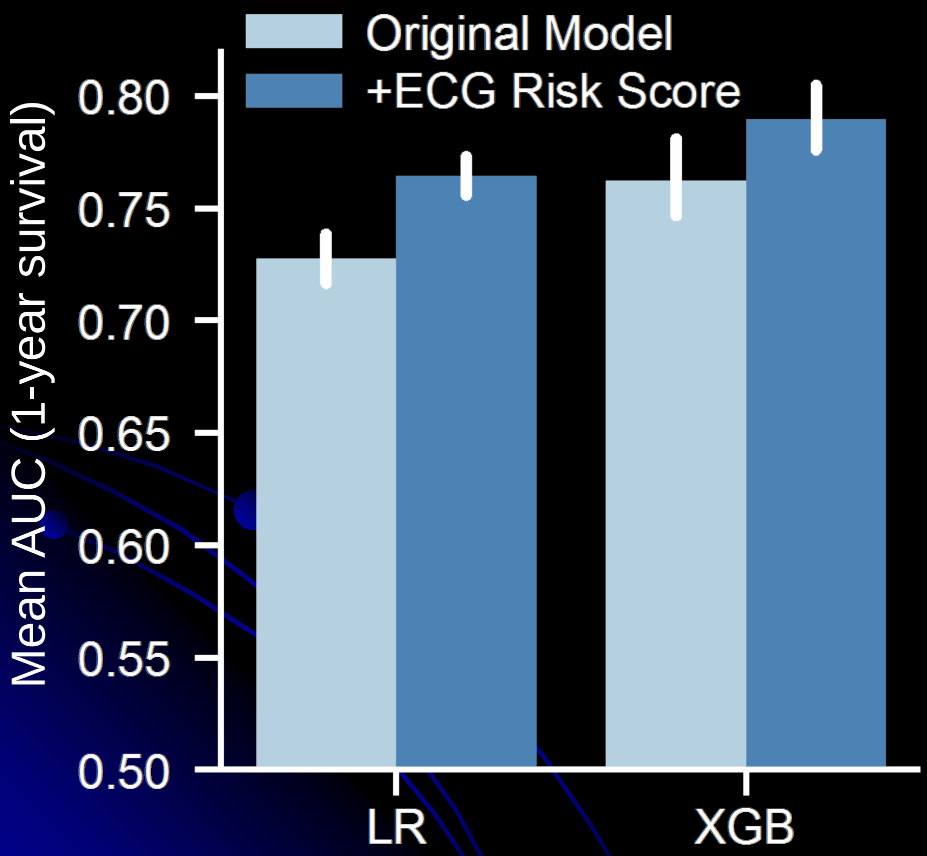
A Ulloa PhD

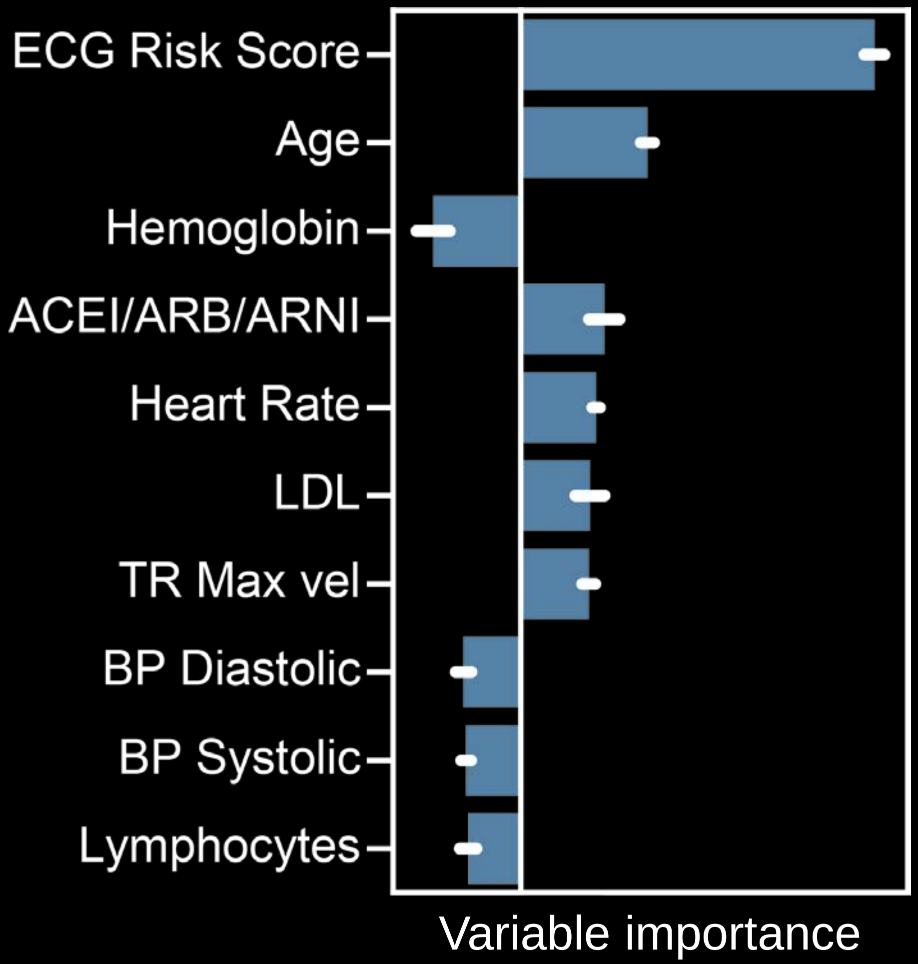



https://arxiv.org/abs/1901.08125

# Heart Failure Multi-Modal Input Example: ECG Highly Important




Trained classifier on 1.3 Million ECGs

# Summary

Large clinical datasets and machine learning will change medicine

- Radiology informatics
- Optimizing predictions from large datasets
- Managing disease populations
- Actionable predictions from complex multi-modal datasets

# This is Innovation

# But Innovation Often Fails

# Acknowledgements


Jeff Adams, MHA
Amro Alsaid, MBBCh
M Arbabshirani, PhD
Dominik Beer, DO
James Brazeal, JD
John Bulger, MD
Dave Carey, PhD
Brendan Carry, MD
Al Casale, MD
Joe Chronowski
John Cleland, MD
Brian Delisle, PhD
Sanjay Doddamani, MD
Mike Evans, RPh
Seth Gazes, MS

Patrick Gladding, MD
Chris Good, DO
Jerry Greskovic, RPh
Dustin Hartzel
Les Kirchner, PhD
Joe Leader
David Ledbetter, PhD
Neil Manus
Greg Moore, MD, PhD
Aalpen Patel, MD
Marios Pattichis, PhD
Nathan Sauers, PharmD
Steve Steinhubl, MD
Joshua Stough, PhD
Eric Topol, MD

CARDIAC IMAGING TECHNOLOGY LAB

Eric Carruth, PhD
Sam Fielden, PhD
Ally Haggerty, MBA
Chris Haggerty, PhD
Linyuan Jing, PhD
Bern McCarty
Chris Nevius
S Raghunath, PhD

Jeff Ruhl
Gargi Schneider, MD
Nathan Stoudt
Jonathan Suever, PhD
Alvaro Ulloa, MS
D vanMaanen, MS
Greg Wehner, PhD

Geisinger

UK UNIVERSITY OF KENTUCKY

UNM

Imperial College London

REGENERON GENETICS CENTER

$$ NIH DP5 OD-012132 to BKF
NIH R01 HL-141901 to CMH
NIH UL1 TR-002550 to E Topol
American Heart Assoc Competitive Catalyst Award to B Delisle
**Geisinger Health Plan** and Clinic
PA Dept of Health

NVIDIA
PURESTORAGE

Colin Compas, PhD
Rory Kelleher
Raghav Mani
Chris Harvey

# We Are Hiring At All Levels!!

bkfornwalt@geisinger.edu

Linyuan Jing, PhD
Math & Computational Scientist

*"Our team is encouraged to innovate and research topics that will actually impact the way we practice medicine. The data that is available to Geisinger researchers is incredible."*

www.Geisinger.org/DISI