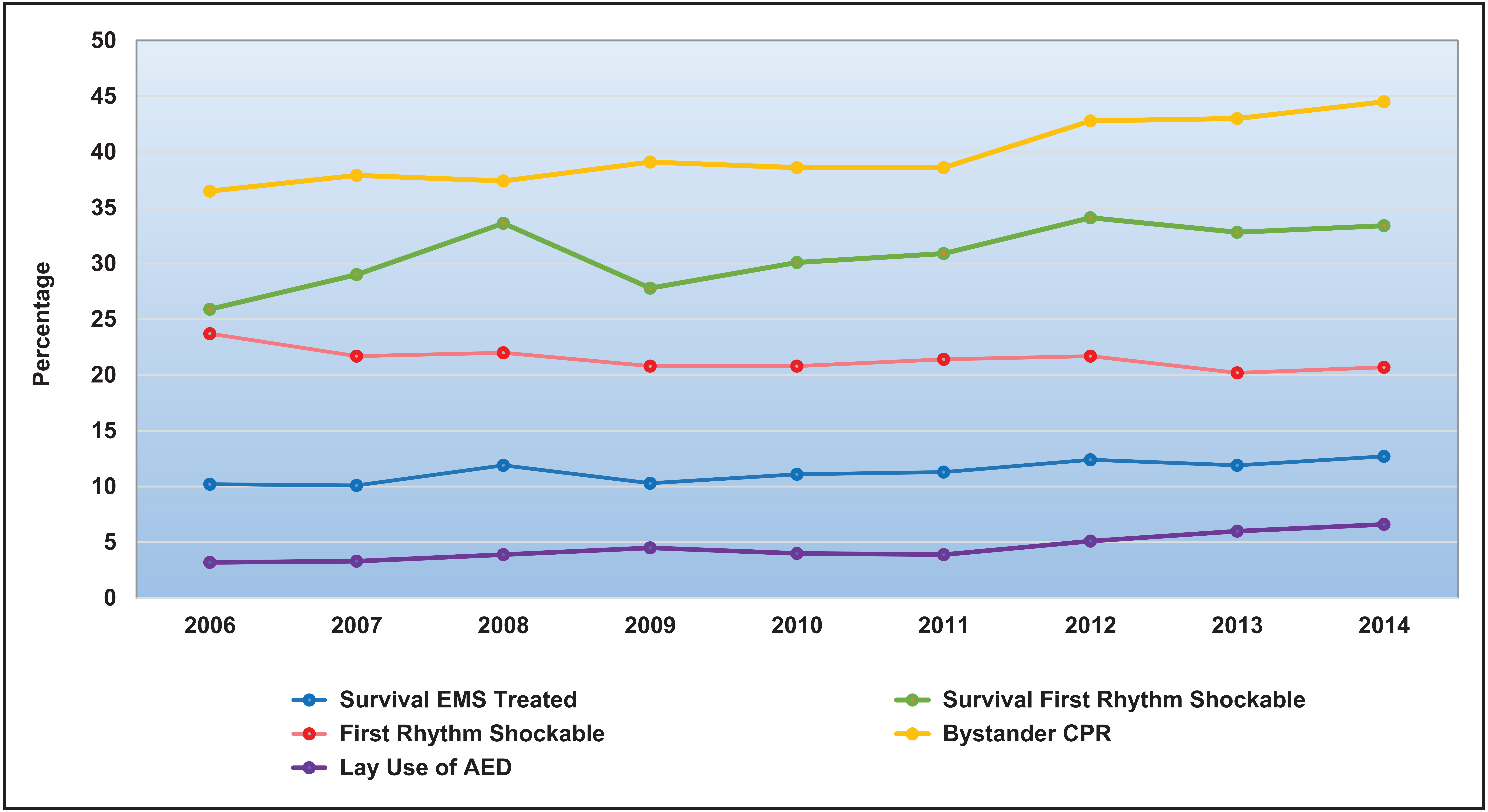
# Predicting and explaining in-hospital mortality following cardiac arrest.

## A machine learning approach

**Shane Nanayakkara** | Sam Fogarty, Michael Tremeer, Kelvin Ross, Brent Richards, Christoph Bergmeir, Sheng Xu, Dion Stub, Karen Smith, David Pilcher, David Kaye




Benjamin EJ et al. Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. 2018.

Clinical paper

# APACHE II scoring to predict outcome in post-cardiac arrest

Michael W. Donnino[a,b,*], Justin D. Salciccioli[a], Andre Dejam[c], Tyler Giberson[a], Brandon Giberson[a], Cristal Cristia[a], Shiva Gautam[d], Michael N. Cocchi[a,e]

*Conclusions:* APACHE II score is a poor predictor of outcome at time zero for out-of-hospital cardiac arrest (OHCA) post-arrest patients consistent with the original development of the score in the critically ill. For in-hospital cardiac arrest (IHCA) at time zero and for both IHCA and OHCA at 24 h and beyond, the APACHE II score was a modest indicator of illness severity and predictor of mortality/neurologic morbidity.

Resuscitation, 2013

Clinical paper

# Survival and outcome prediction using the Apache III and the out-of-hospital cardiac arrest (OHCA) score in patients treated in the intensive care unit (ICU) following out-of-hospital, in-hospital or ICU cardiac arrest☆

M.B. Skrifvars[a,b,*], B. Varghese[a], M.J. Parr[a,c]

*Conclusions:* Latency to ROSC seems to be the most important determinant of survival in patients following ICU care after a cardiac arrest in this single center trial. The OHCA score and the Apache III score offer moderate predictive accuracy in ICU cardiac arrest patients but correlated weakly with each other. Illness severity adjustment for cardiac arrest patients in ICU should include features of both these scoring systems.

Clinical Paper

# Effectiveness of SAPS III to predict hospital mortality for post-cardiac arrest patients☆

Magali Bisbal[a,b,*], Elisabeth Jouve[c], Laurent Papazian[d,e], Sophie de Bourmont[a,b], Gilles Perrin[b], Beatrice Eon[b], Marc Gainnier[a,b]

*Conclusions:* The SAPS III did not predict mortality in patients admitted to ICU after CA. The amount of time before specialized CPR, the low-flow interval and the absence of an initial ventricular arrhythmia appeared to be independently associated with mortality and these factors should be used to predict mortality for these patients.

# Key predictive variables

Summed from the literature.

rhythm on admission

time to ROSC

bystander CPR

time to adrenaline

duration of CPR

simplicity of prognostication

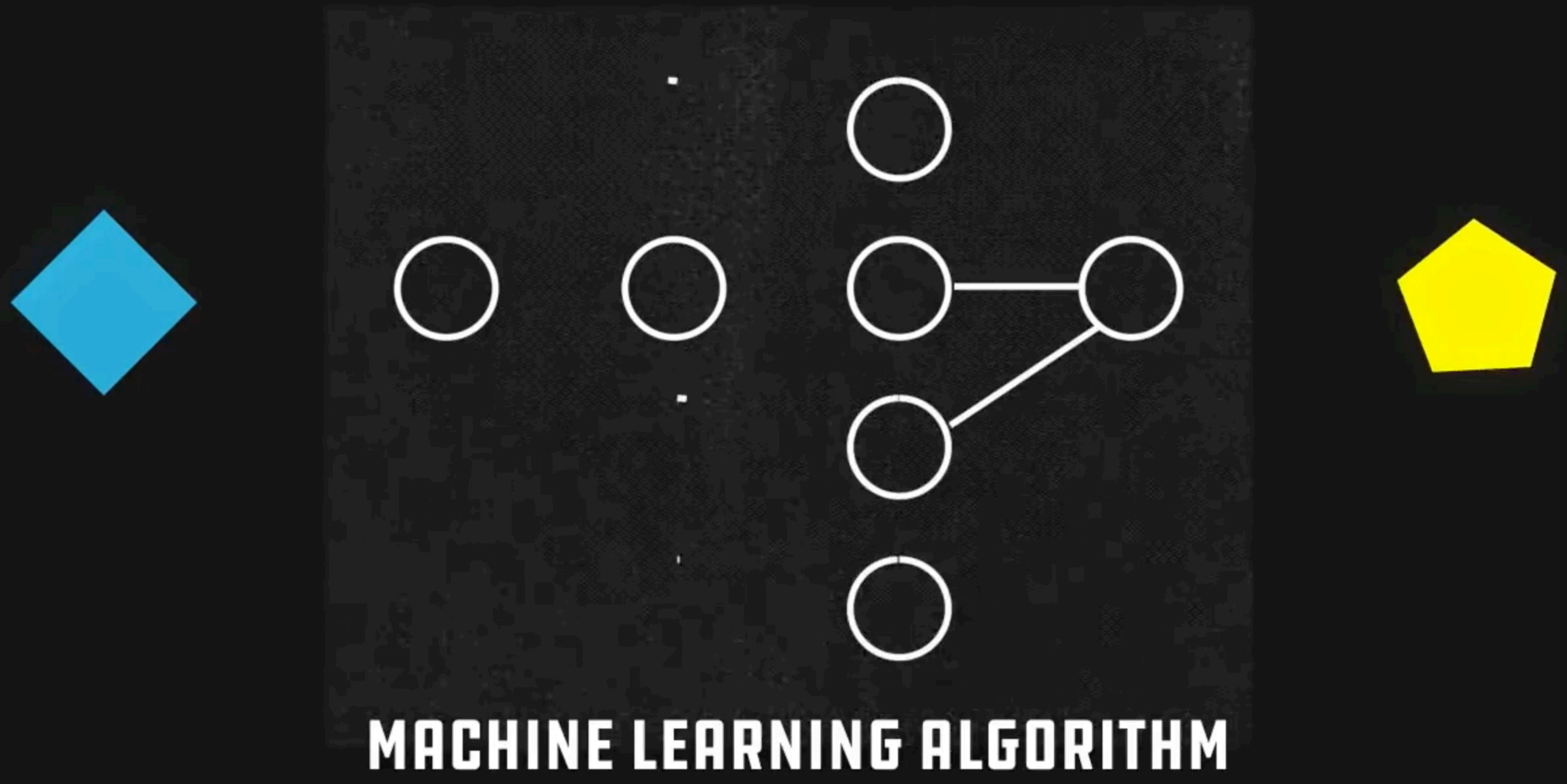
After hearing for several decades that computers will soon be able to assist with difficult diagnoses, the practicing physician may well wonder why the revolution has not occurred.

APRIL 1987

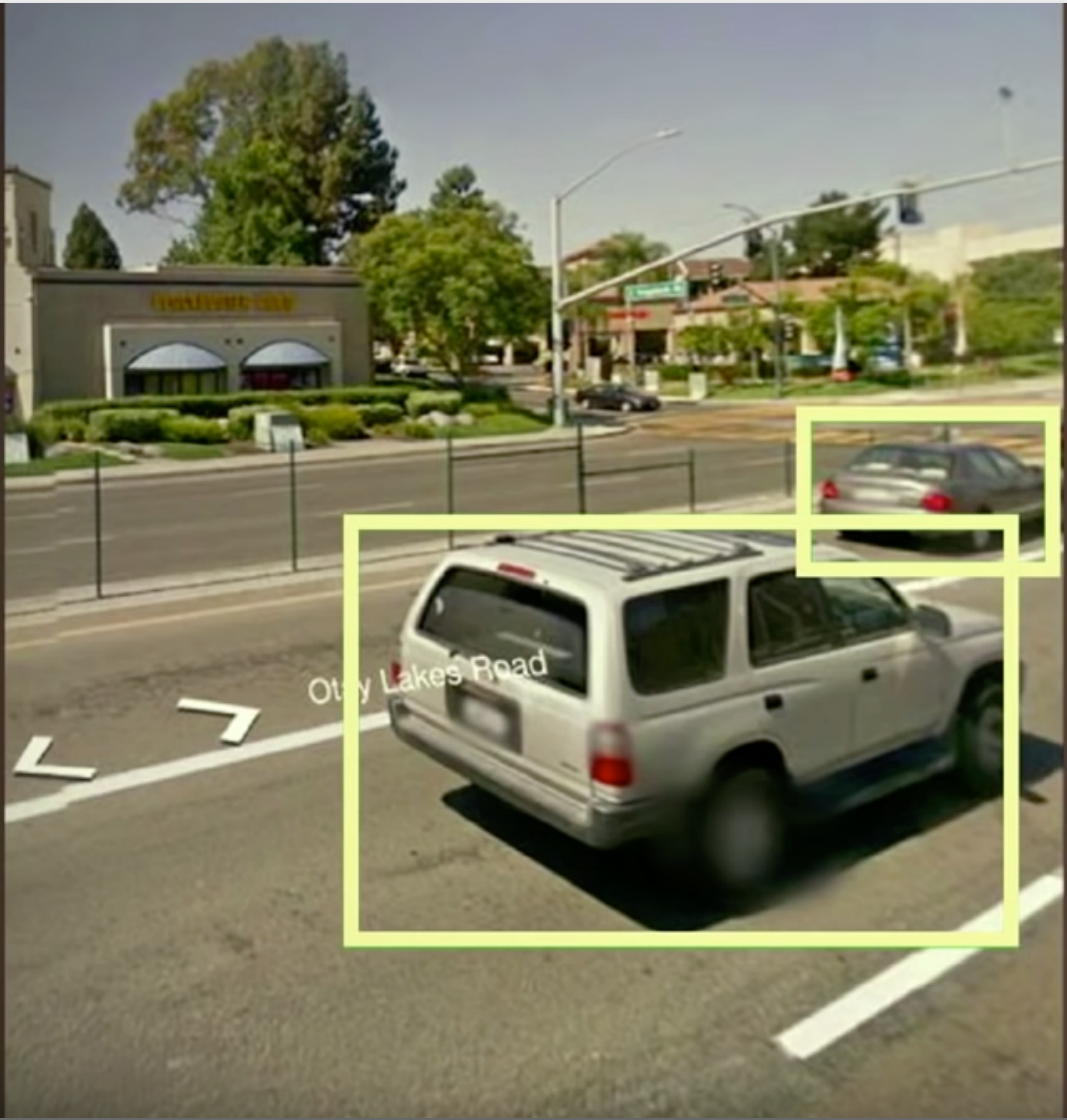
The NEW ENGLAND JOURNAL of MEDICINE

N Engl J Med 1987; 316:685-688

"Machine learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other."



LEFT REARWARD VEHICLE CAMERA
MEDIUM RANGE VEHICLE CAMERA
RIGHT REARWARD VEHICLE CAMERA
MOTION FLOW
LANE LINES
LANE LINES
ROAD FLOW
IN-PATH OBJECTS
ROAD LIGHTS
OBJECTS
ROAD SIGNS

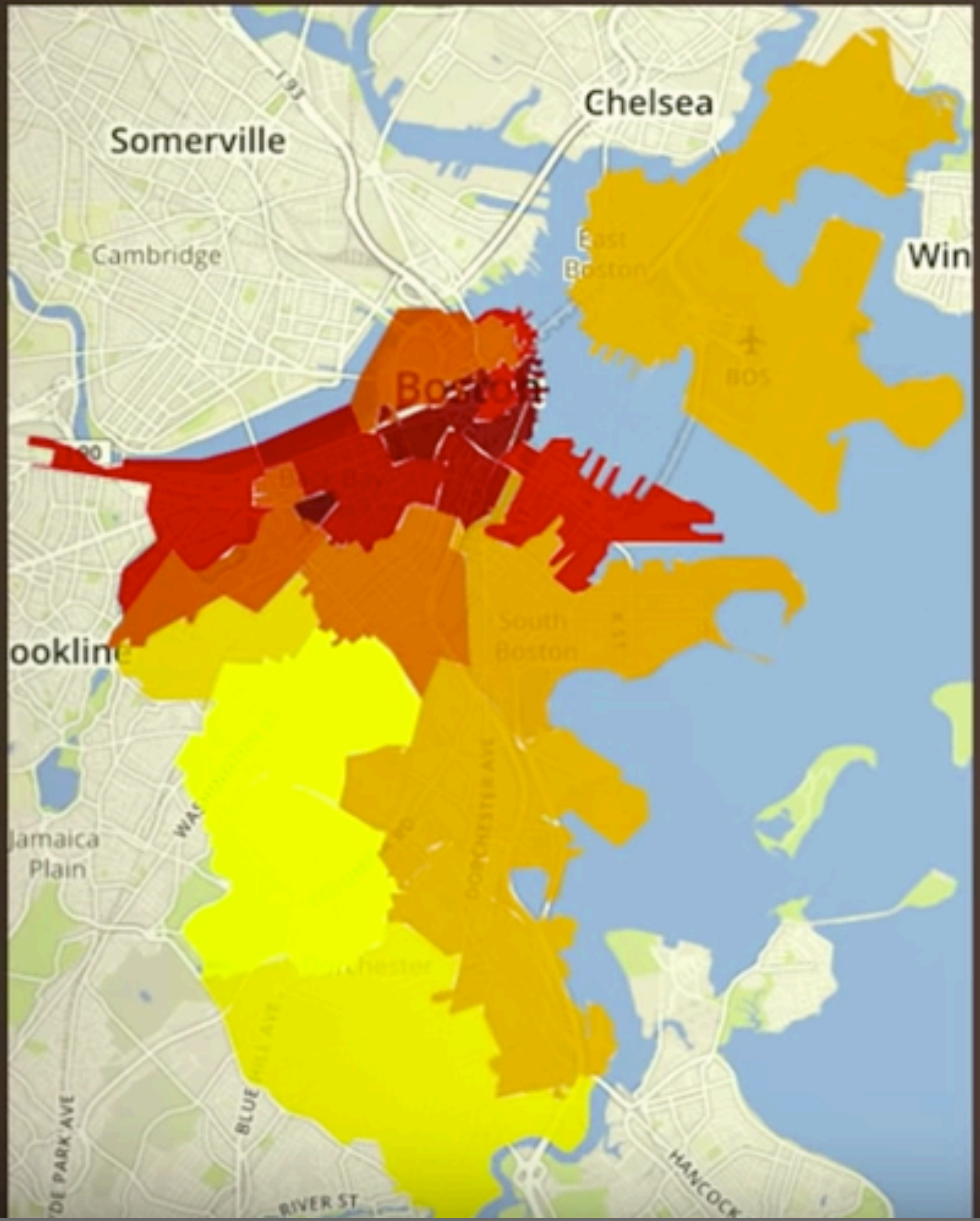


2005 Nissan
Sentra 1.8 S

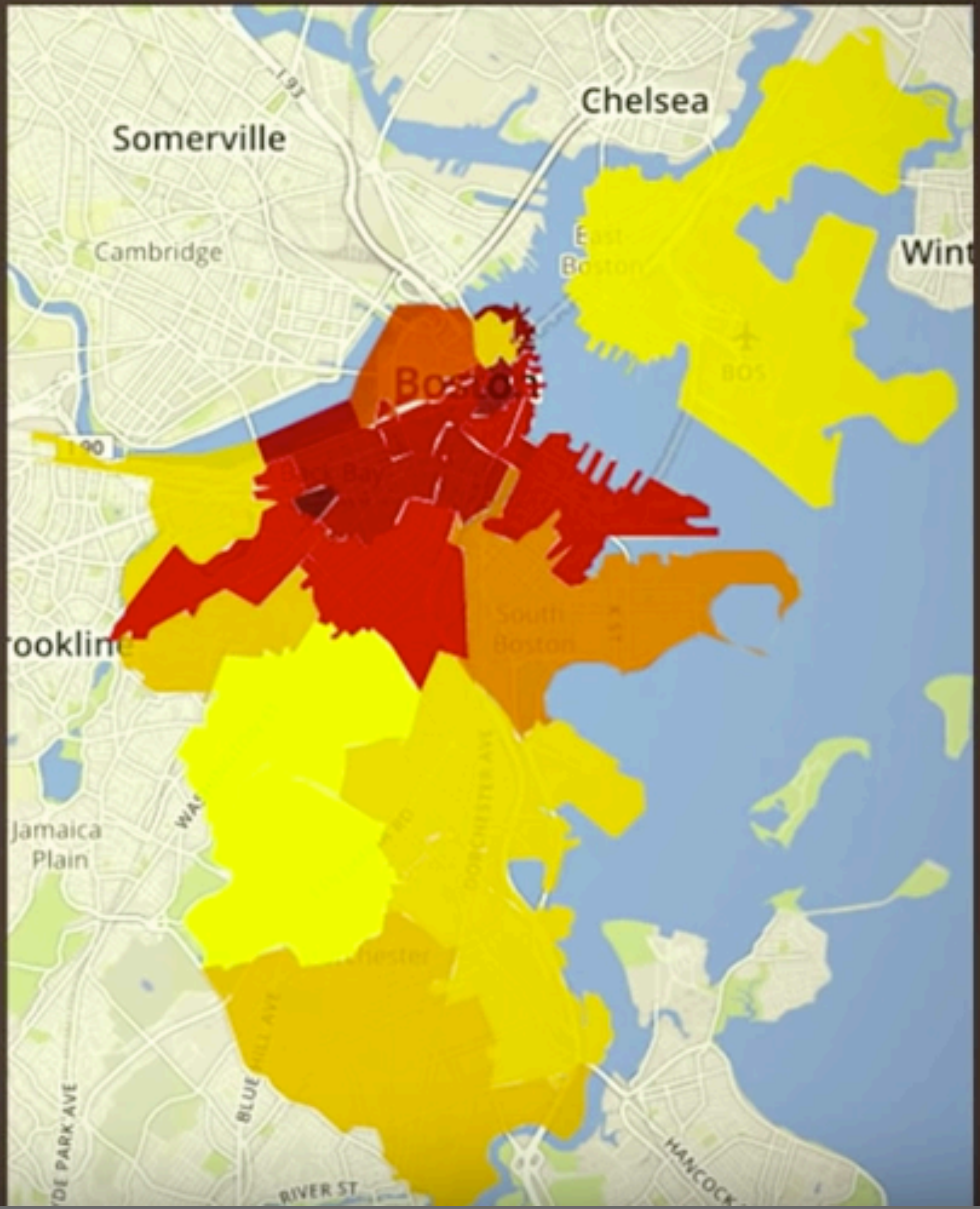
2000 Toyota
4Runner Limited

# Income in Boston, MA

## Actual

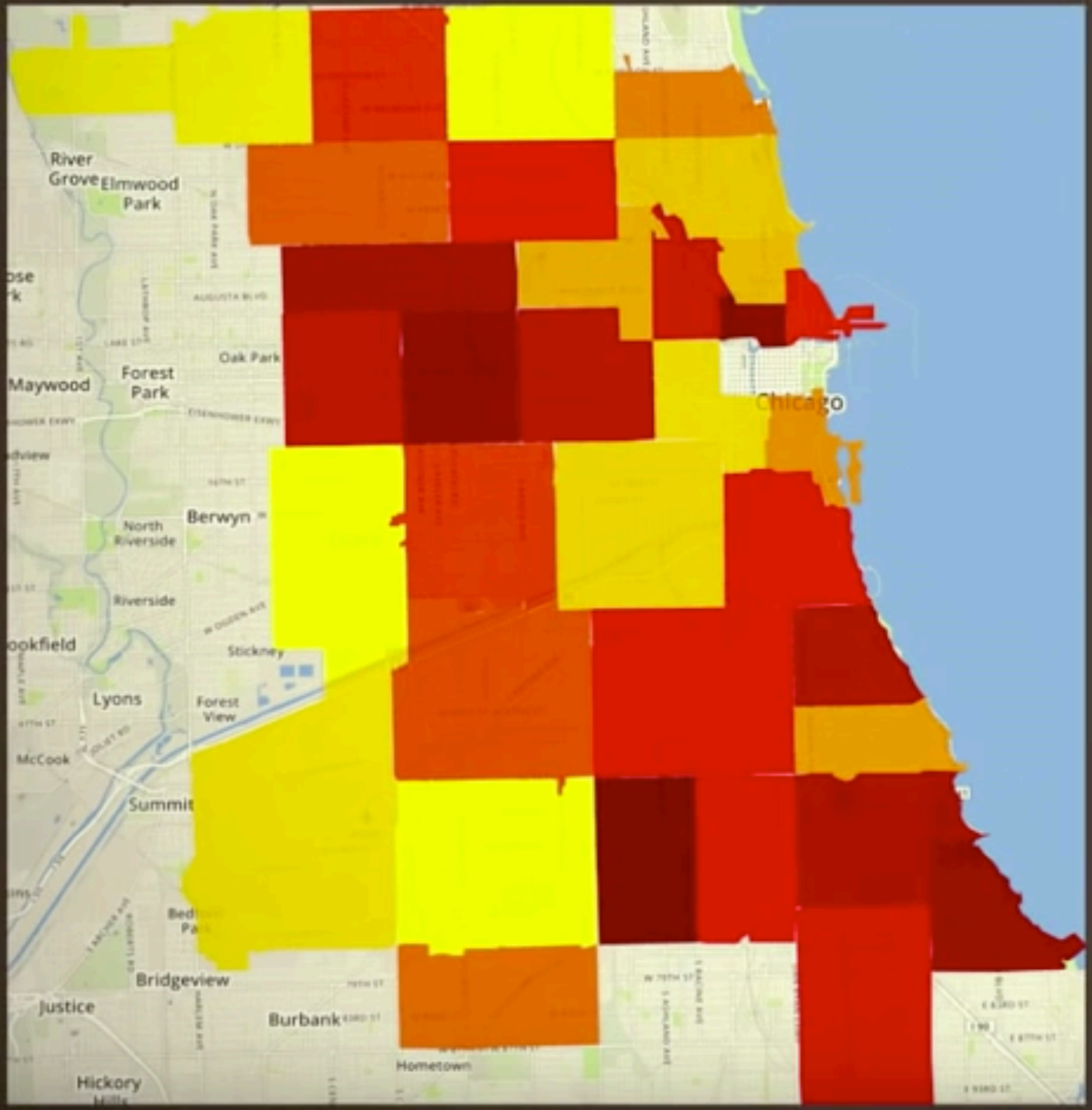



## Predicted by Car Price

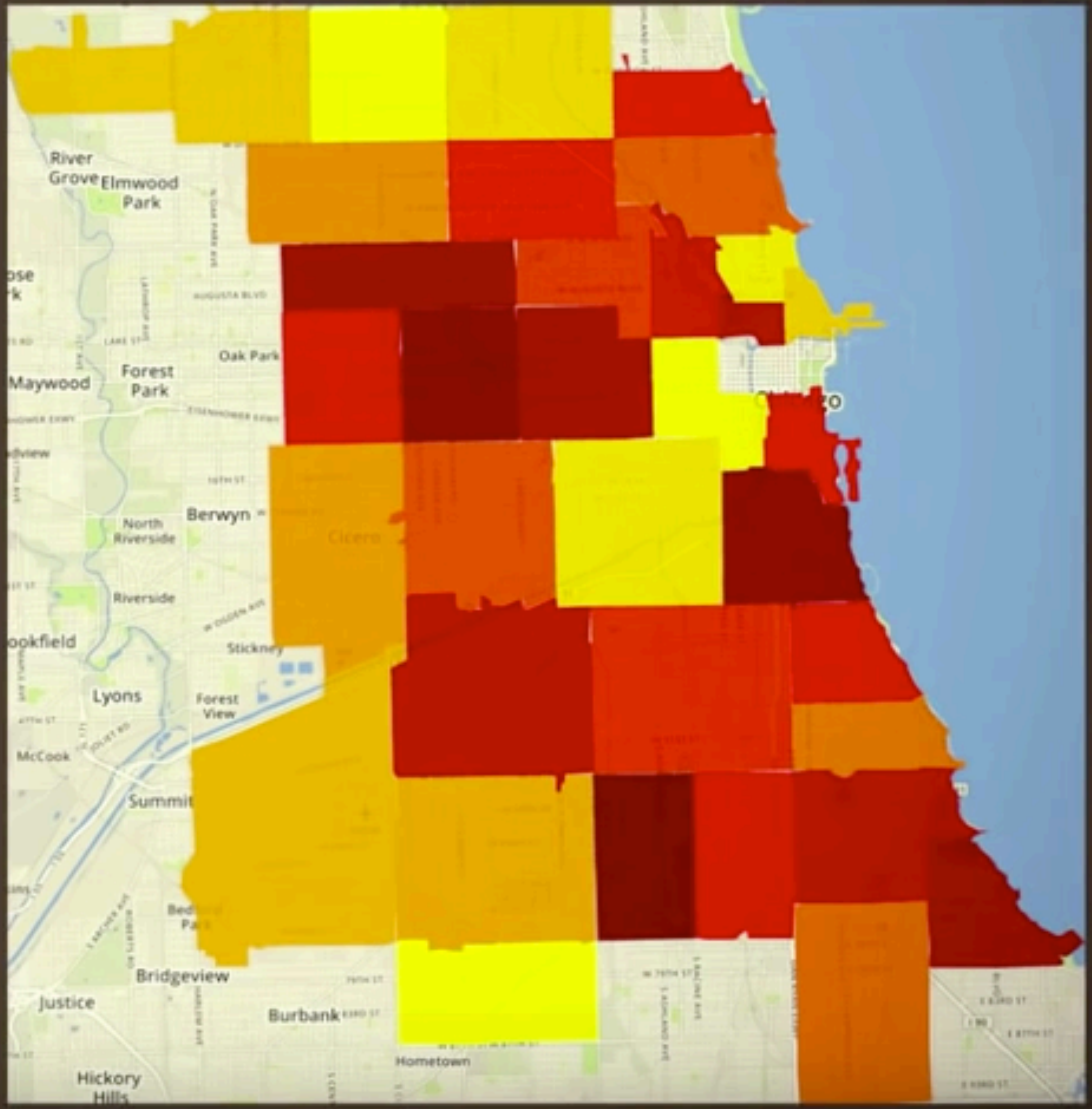


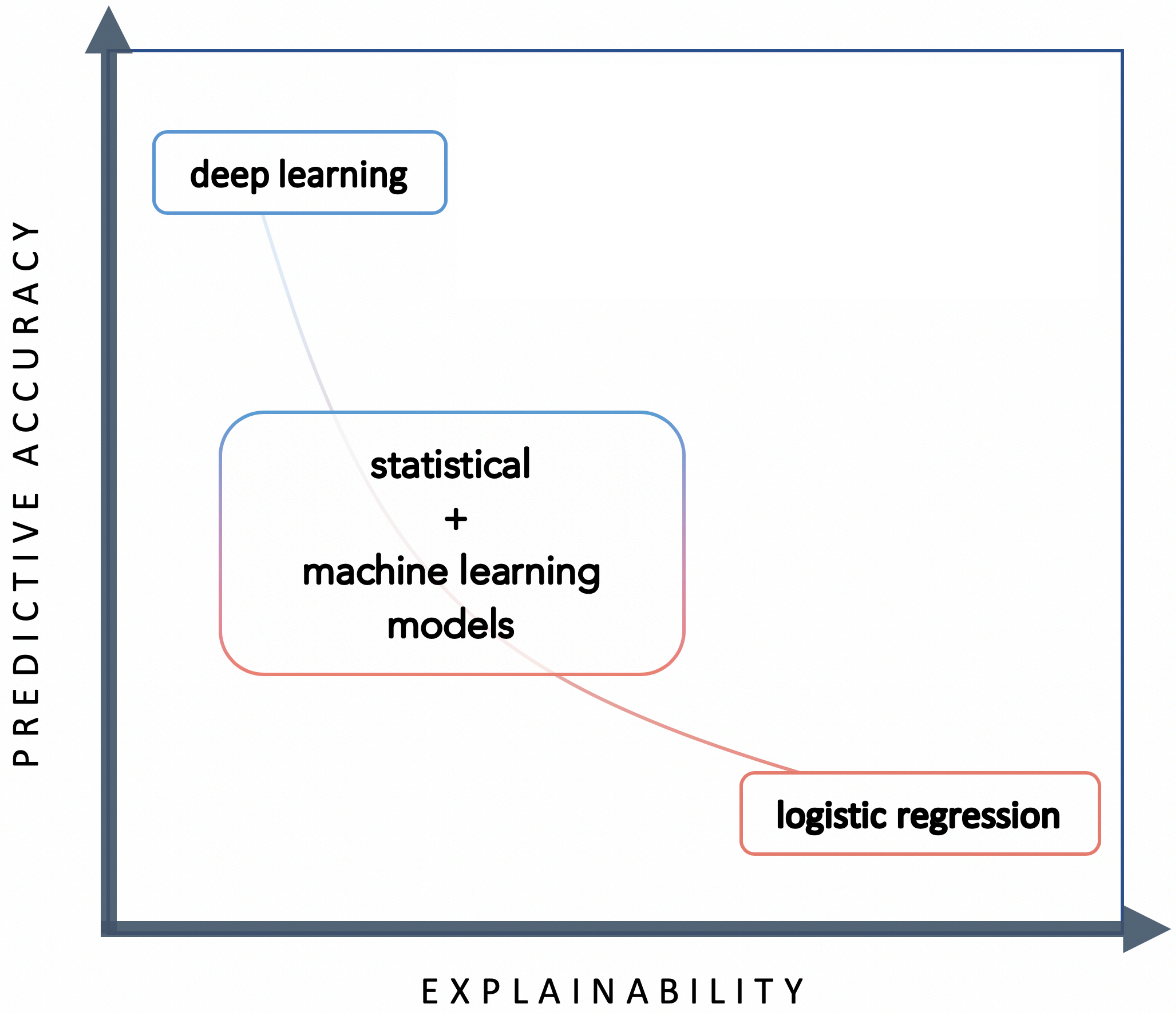
# Crime in Chicago, IL

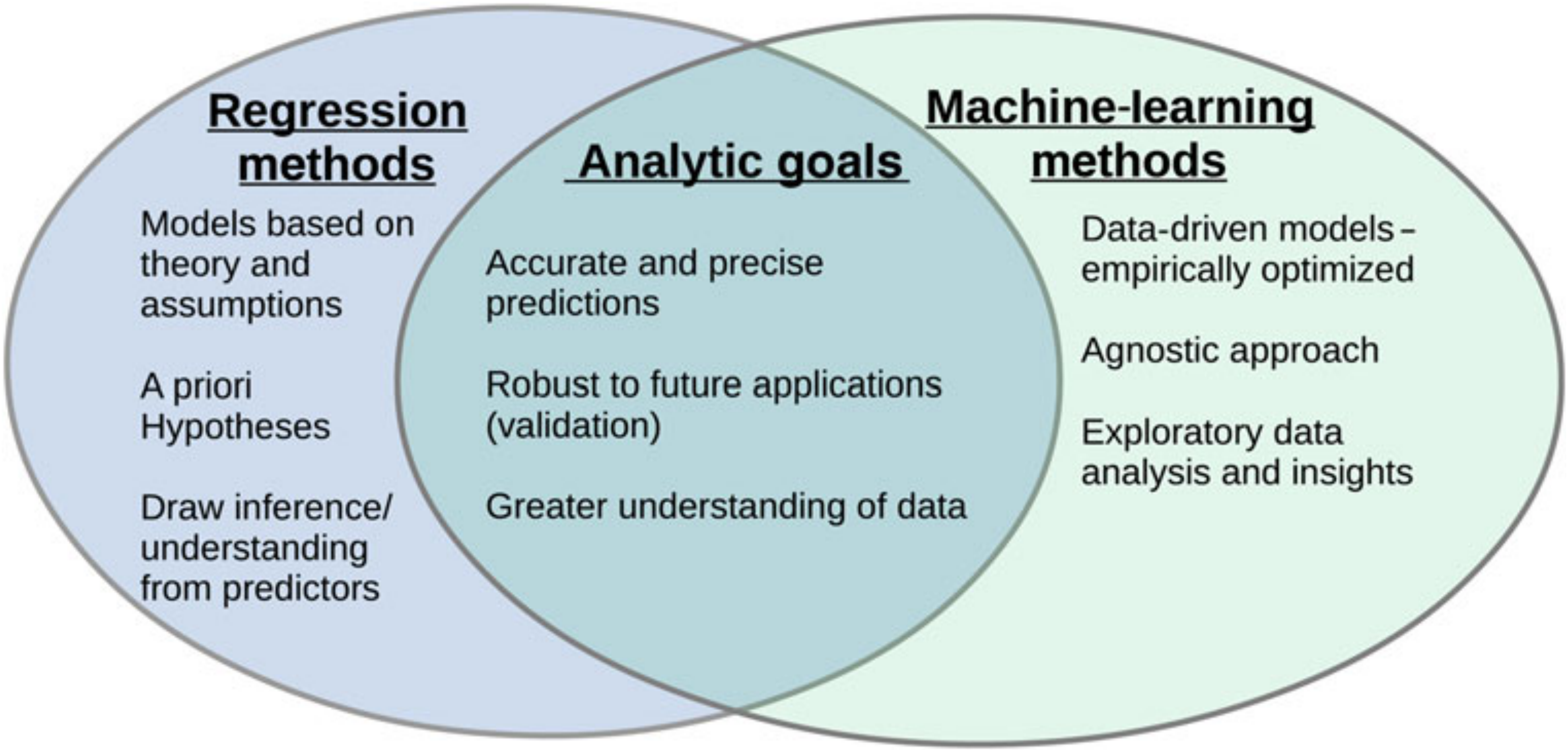
## Actual




## Predicted by Car Price

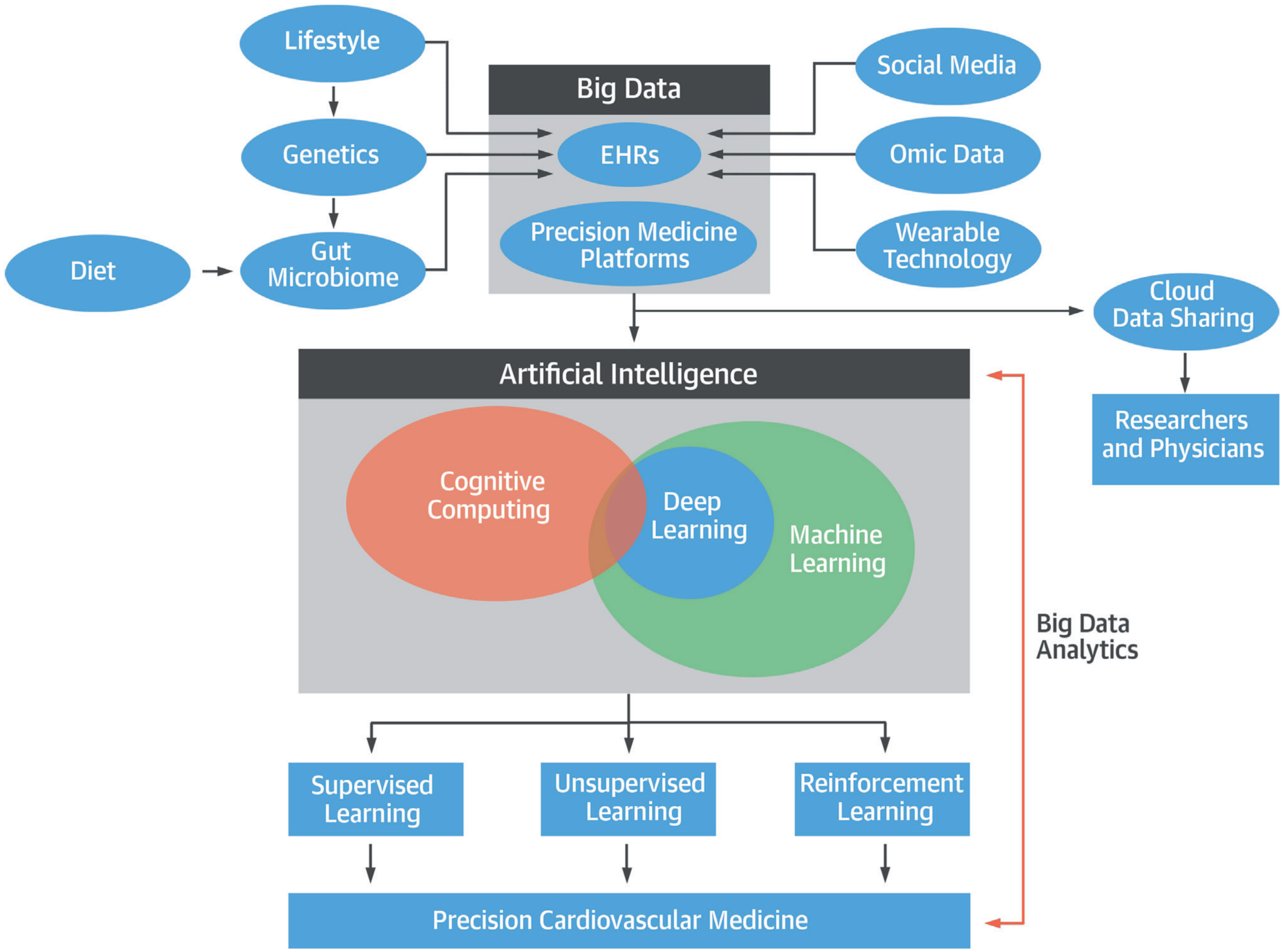



deep learning
statistical
+
machine learning
models
logistic regression
PREDICTIVE ACCURACY
EXPLAINABILITY




Goldstein, B. A., Navar, A. M. & Carter, R. E. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J ehw302 (2016).

**CENTRAL ILLUSTRATION** **Artificial Intelligence in Precision Cardiovascular Medicine**




**Krittanawong, C. et al. J Am Coll Cardiol. 2017;69(21):2657–64.**

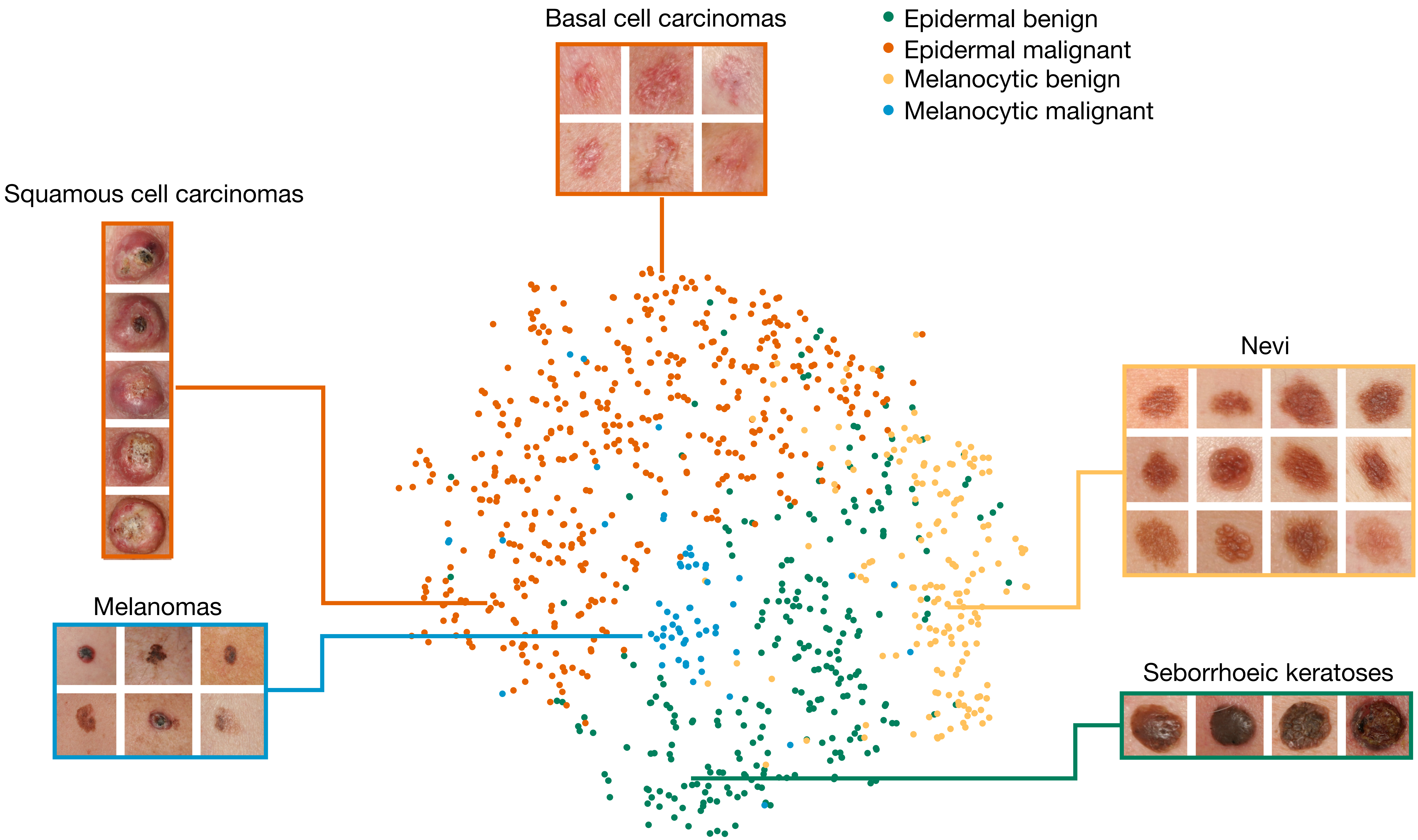
Big data (genetics, social media, environmental, and lifestyle-related factors, or "omic" data) can be stored through EHRs or precision medicine platforms, and can be shared for data analysis with other physicians or researchers through secure cloud systems. Big data analytics using artificial intelligence (machine learning, deep learning, or cognitive computing) and 3 main types of learning algorithms (supervised, unsupervised, and reinforcement learning) will enable precision cardiovascular medicine. EHR = electronic health record.

We can automate anything a person can do with less than one second of thinking.

Andrew Ng

# Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]



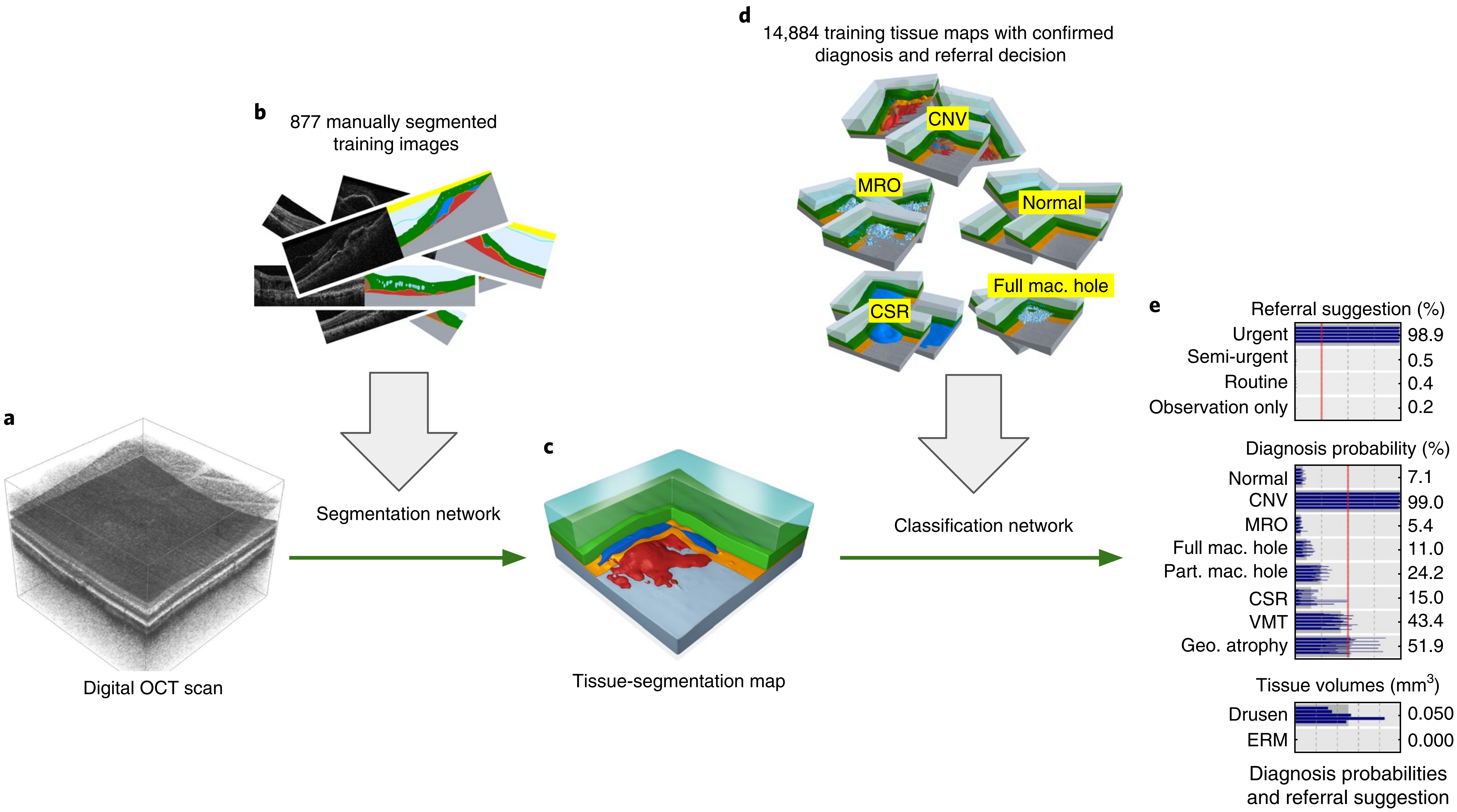

Nature **542**, 115–118 (02 February 2017) nature medicine

ARTICLES

https://doi.org/10.1038/s41591-018-0107-6

# Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw[1], Joseph R. Ledsam[1], Bernardino Romera-Paredes[1], Stanislav Nikolov[1], Nenad Tomasev[1], Sam Blackwell[1], Harry Askham[1], Xavier Glorot[1], Brendan O'Donoghue[1], Daniel Visentin[1], George van den Driessche[1], Balaji Lakshminarayanan[1], Clemens Meyer[1], Faith Mackinder[1], Simon Bouton[1], Kareem Ayoub[1], Reena Chopra[2], Dominic King[1], Alan Karthikesalingam[1], Cían O. Hughes[1,3], Rosalind Raine[3], Julian Hughes[2], Dawn A. Sim[2], Catherine Egan[2], Adnan Tufail[2], Hugh Montgomery[3], Demis Hassabis[1], Geraint Rees[3], Trevor Back[1], Peng T. Khaw[2], Mustafa Suleyman[1], Julien Cornebise[1,3,4], Pearse A. Keane[2,4*] and Olaf Ronneberger[1,4*]




Nature Medicine, 2018

Intelligent Machines

# New Research Aims to Solve the Problem of AI Bias in "Black Box" Algorithms

As we automate more and more decisions, being able to understand how an AI thinks is increasingly important.

by Jackie Snow November 7, 2017

# ANZICS

Adult Patient Database.

1,484,536 patients → cardiac arrest; elective admissions; inter-hospital transfer; treatment limitation; unknown mortality outcome → 39,566 patients

cardiac arrest

elective admissions
inter-hospital transfer
treatment limitation
unknown mortality outcome

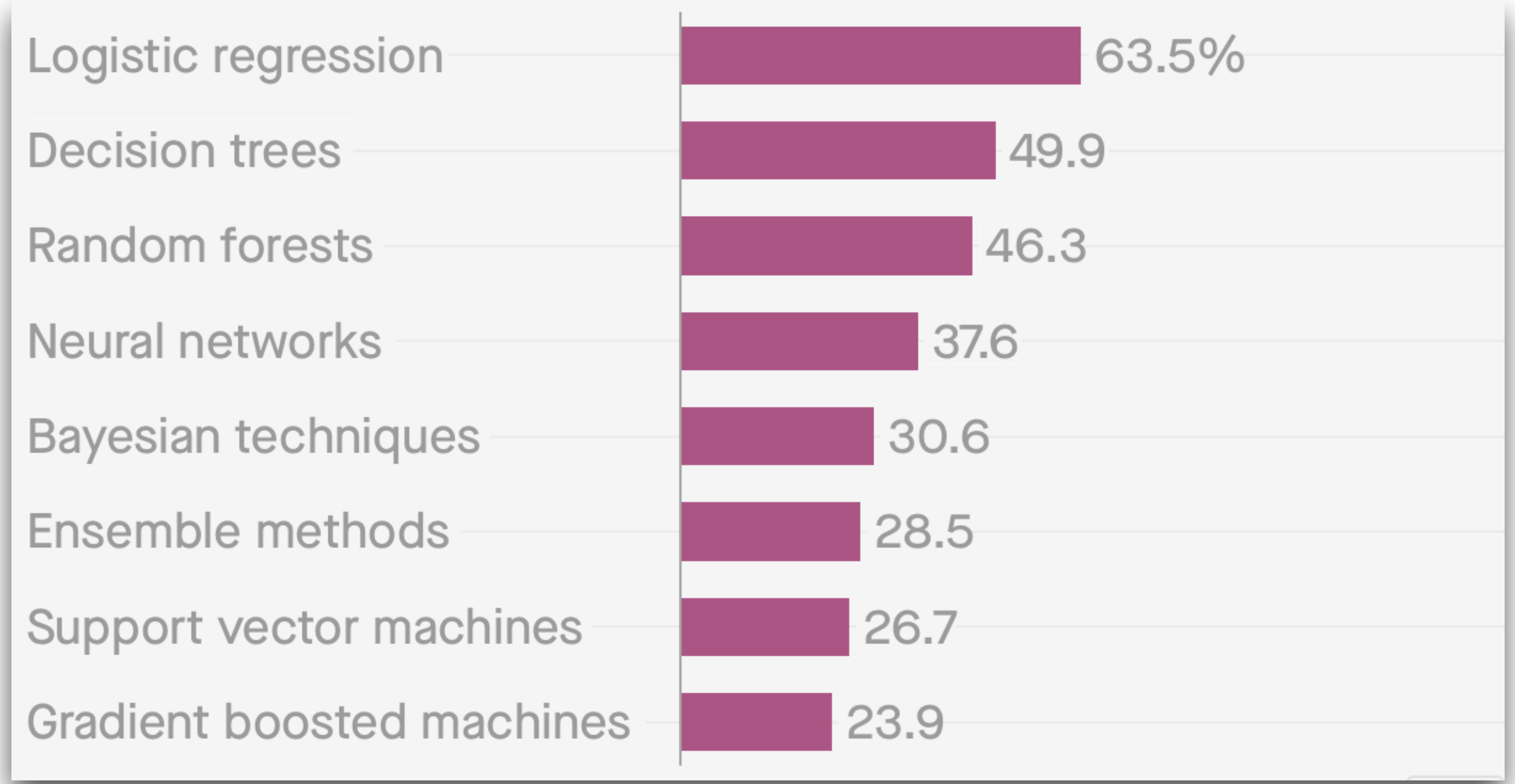
# Baseline characteristics

186 ICUs, 39,566 patients.

| n | 21,547<br>survivors | 18,019<br>non-survivors | p value |
|---|---|---|---|
| **age** | 63 [50-73] | 66 [52-77] | <0.001 |
| **sex (% male)** | 14,255 (66.2%) | 11,629 (64.5%) | 0.001 |
| **peak temperature** | 37.1±1.0 | 36.8±1.5 | <0.001 |
| **peak creatinine** | 101 [76-151] | 146 [102-198] | <0.001 |

# Most popular prediction methods used by data scientists

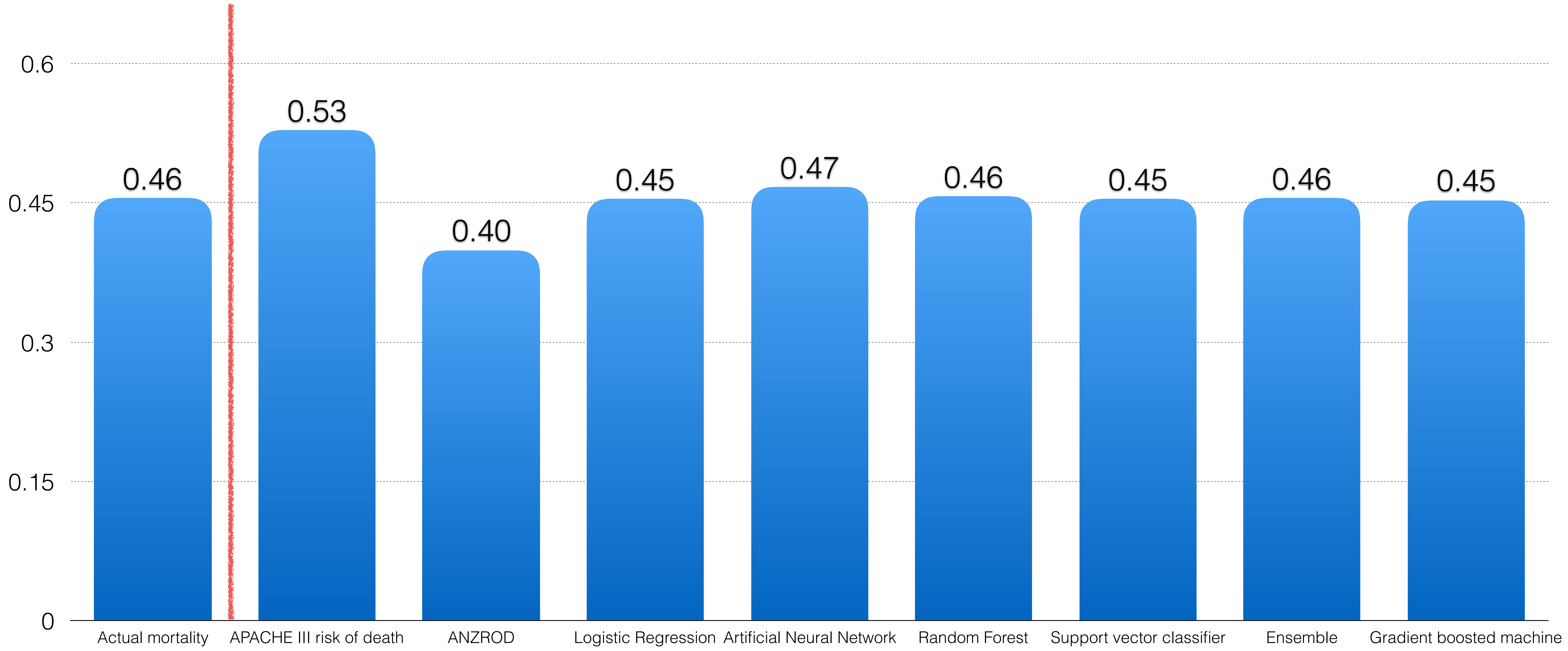
(Based on a survey of Kaggle users)




Kaggle

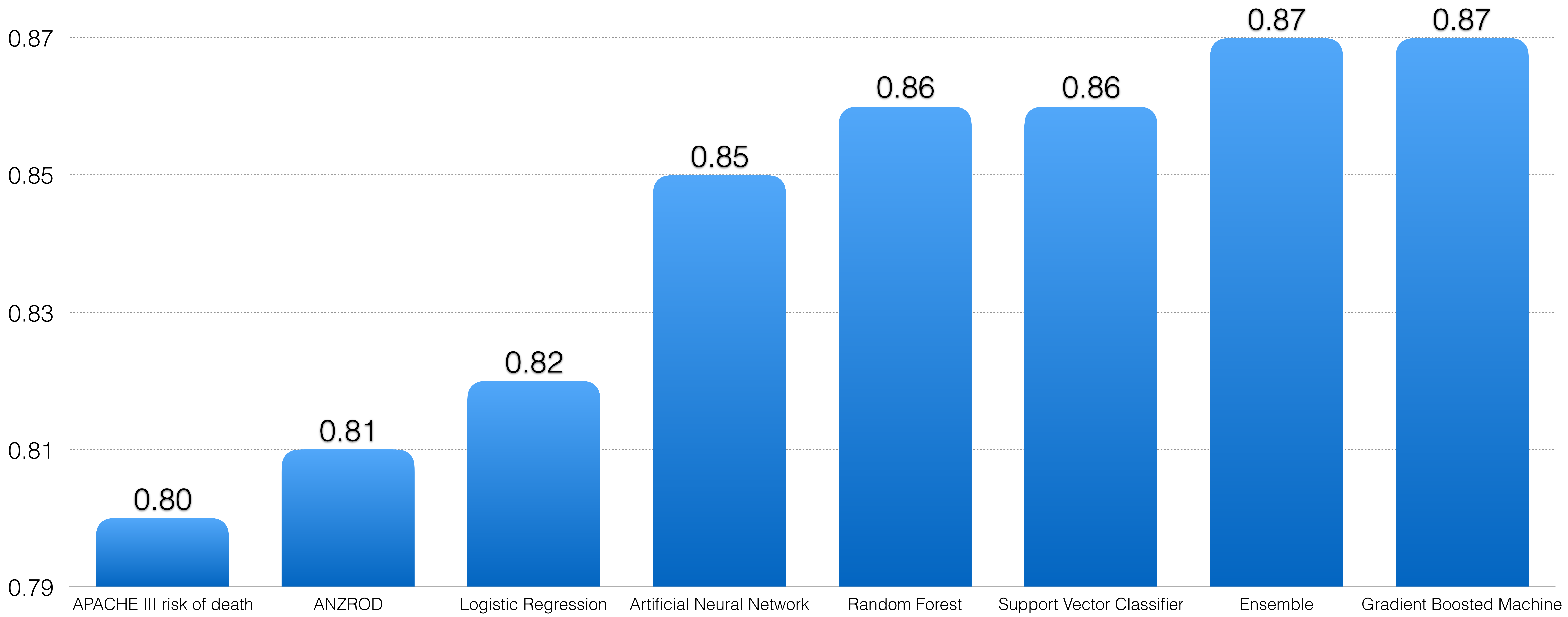
# Comparison of model accuracy

Predicted mortality | existing scores, logistic regression, and machine learning models.



# Comparison of model accuracy (AUROC)

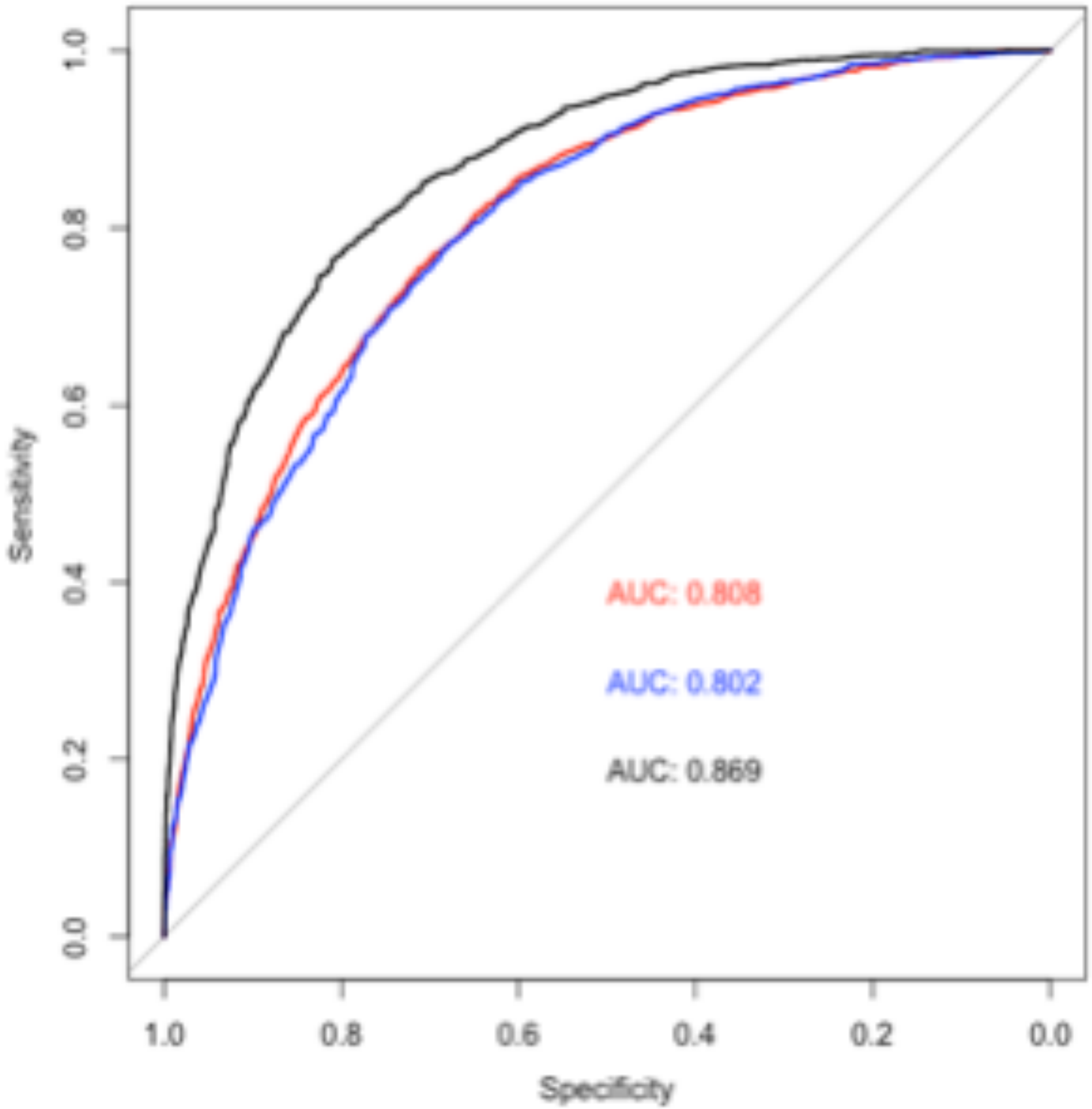
Area under the receiver operating curve | Existing scores, logistic regression, and machine learning models.



# Comparison of model accuracy

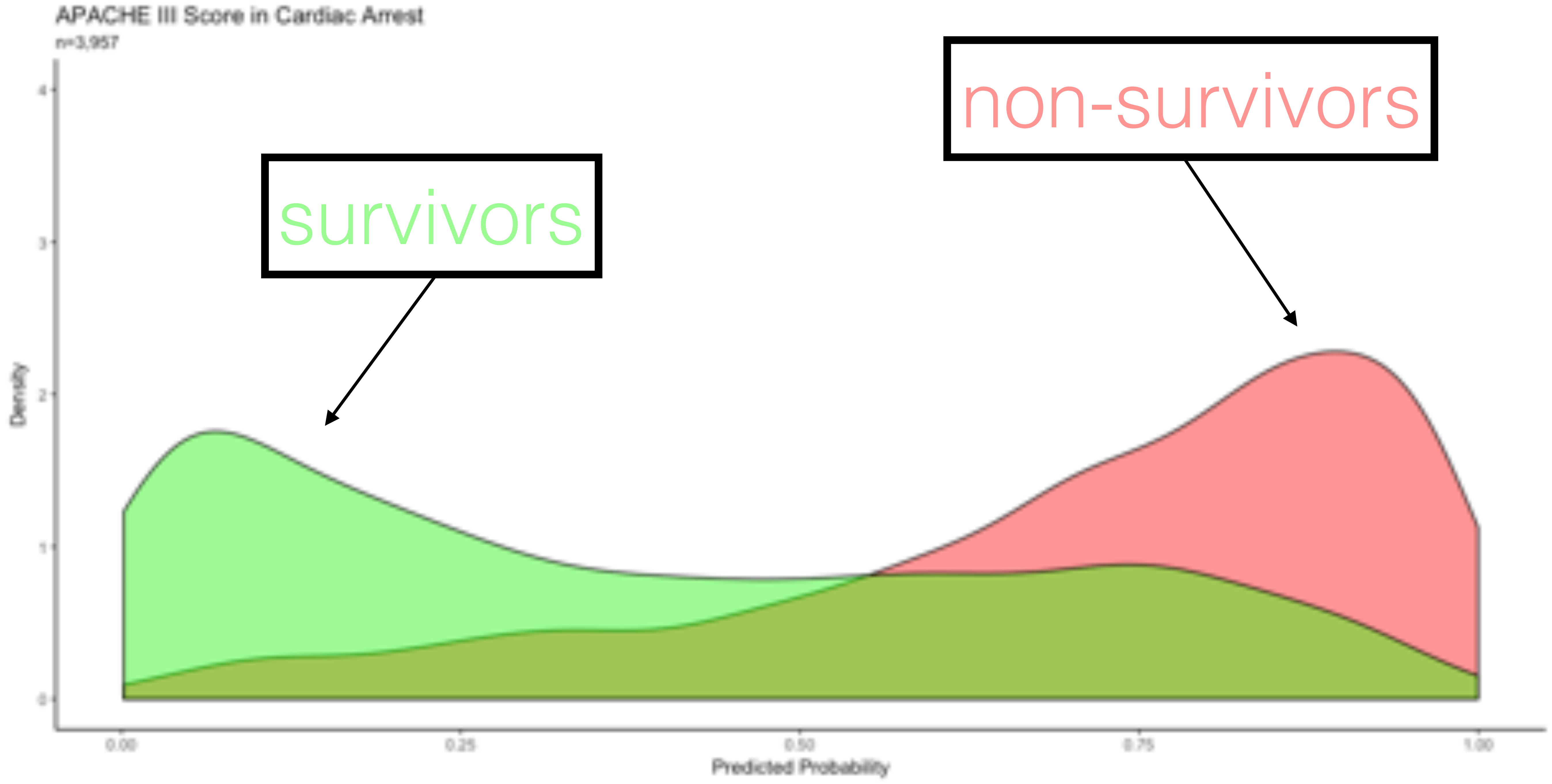
Existing scores, logistic regression, and machine learning models.

| Model | Prediction | AUC (95% CI) | Brier Score | Log Loss |
|---|---|---|---|---|
| Actual mortality | 45.5% | | | |
| APACHE III risk of death | 52.8% | 0.80 (0.79-0.82) | 0.190 | 0.57 |
| ANZROD | 39.9% | 0.81 (0.80-0.82) | 0.182 | 0.55 |
| Logistic Regression | 45.4% | 0.82 (0.81-0.83) | 0.170 | 0.51 |
| Artificial Neural Network | 46.7% | 0.85 (0.84-0.86) | 0.158 | 0.48 |
| Random Forest | 45.7% | 0.86 (0.84-0.87) | 0.156 | 0.47 |
| Support vector classifier | 45.4% | 0.86 (0.85-0.87) | 0.153 | 0.47 |
| Ensemble | 45.5% | 0.87 (0.86-0.88) | 0.148 | 0.45 |
| Gradient boosted machine | 45.3% | 0.87 (0.86-0.88) | 0.147 | 0.45 |

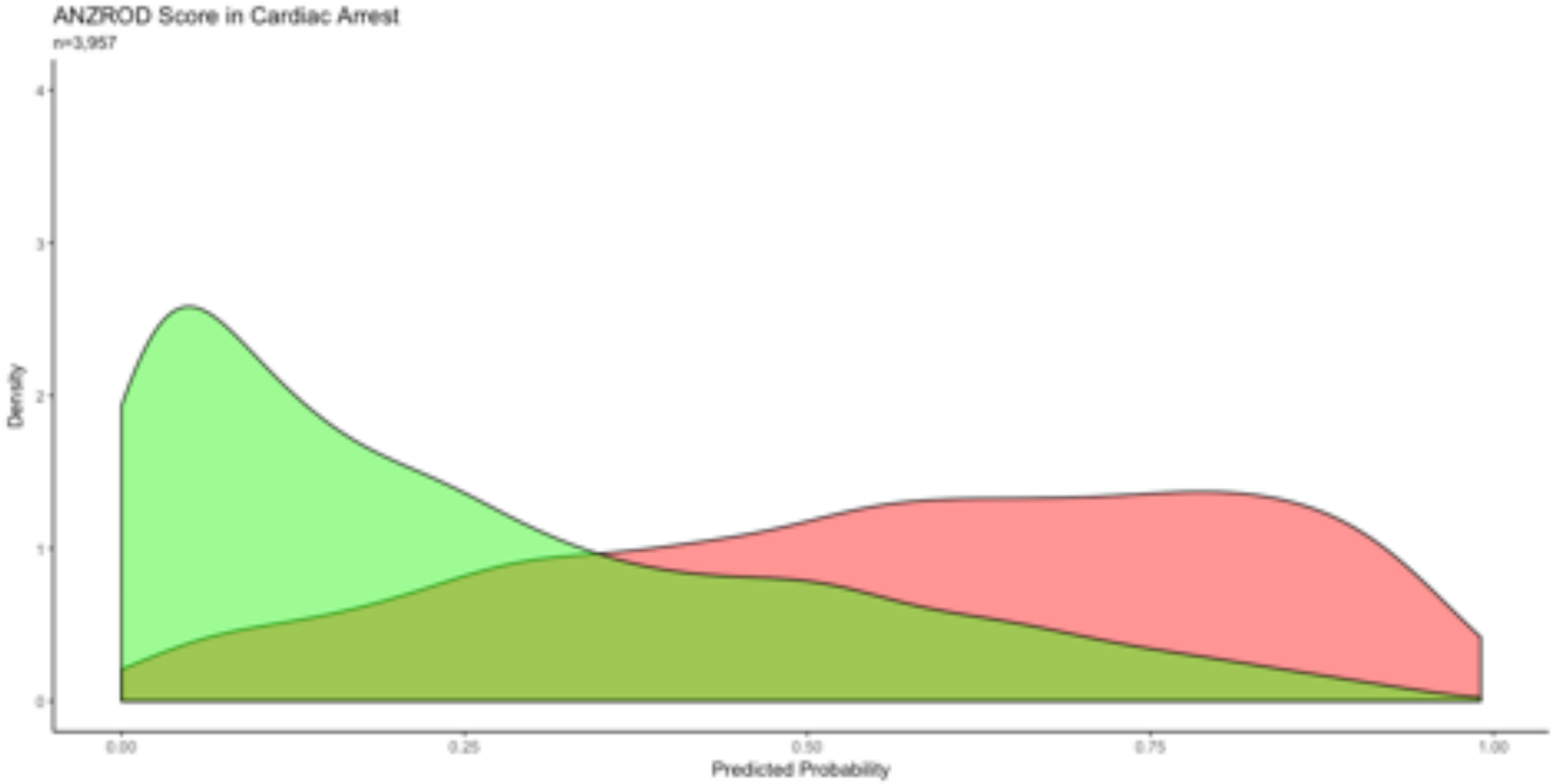


1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
AUC: 0.808
AUC: 0.802
AUC: 0.869
1.0
0.8
0.6
0.4
0.2
0.0
Specificity

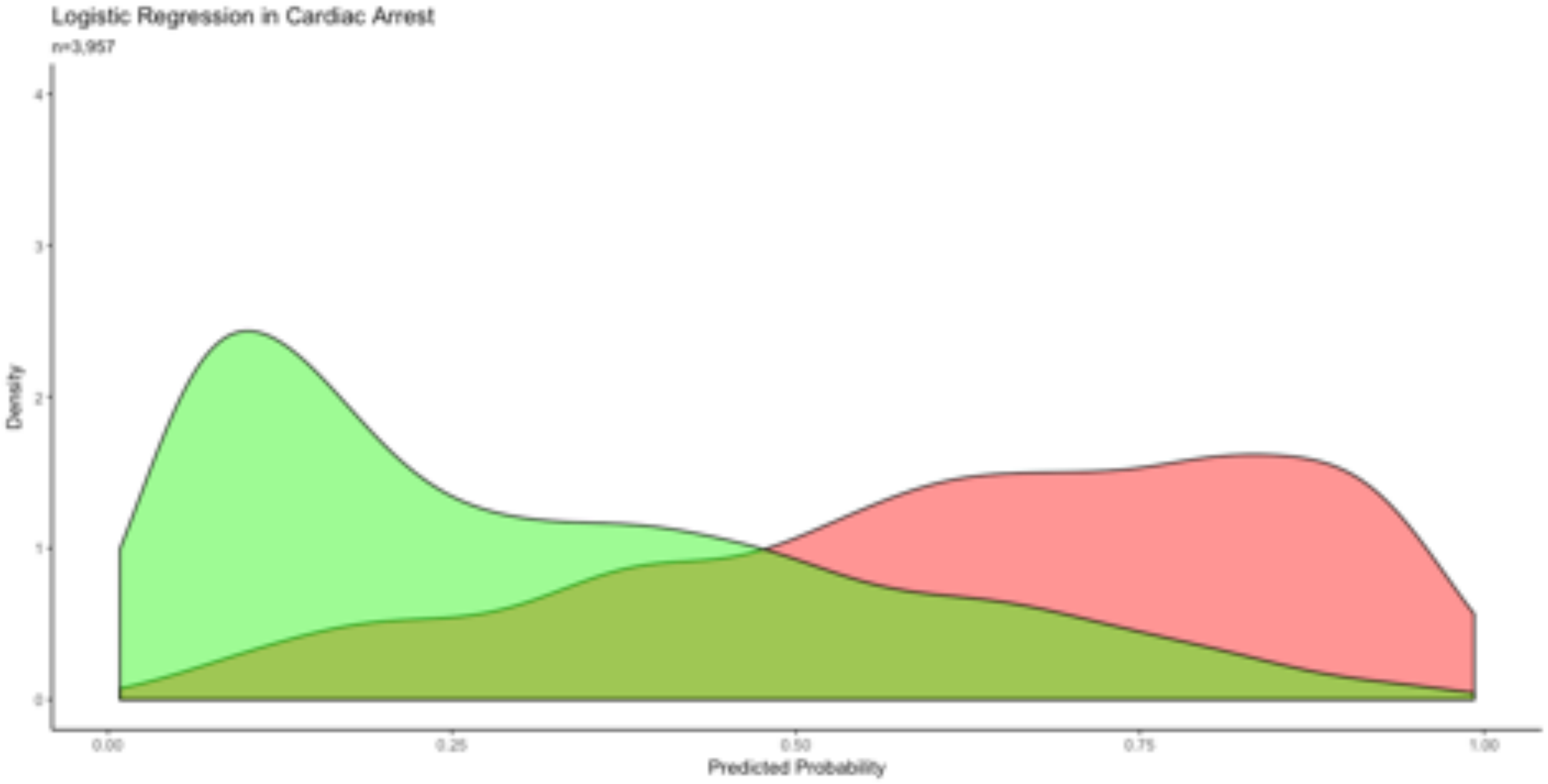
# APACHE III



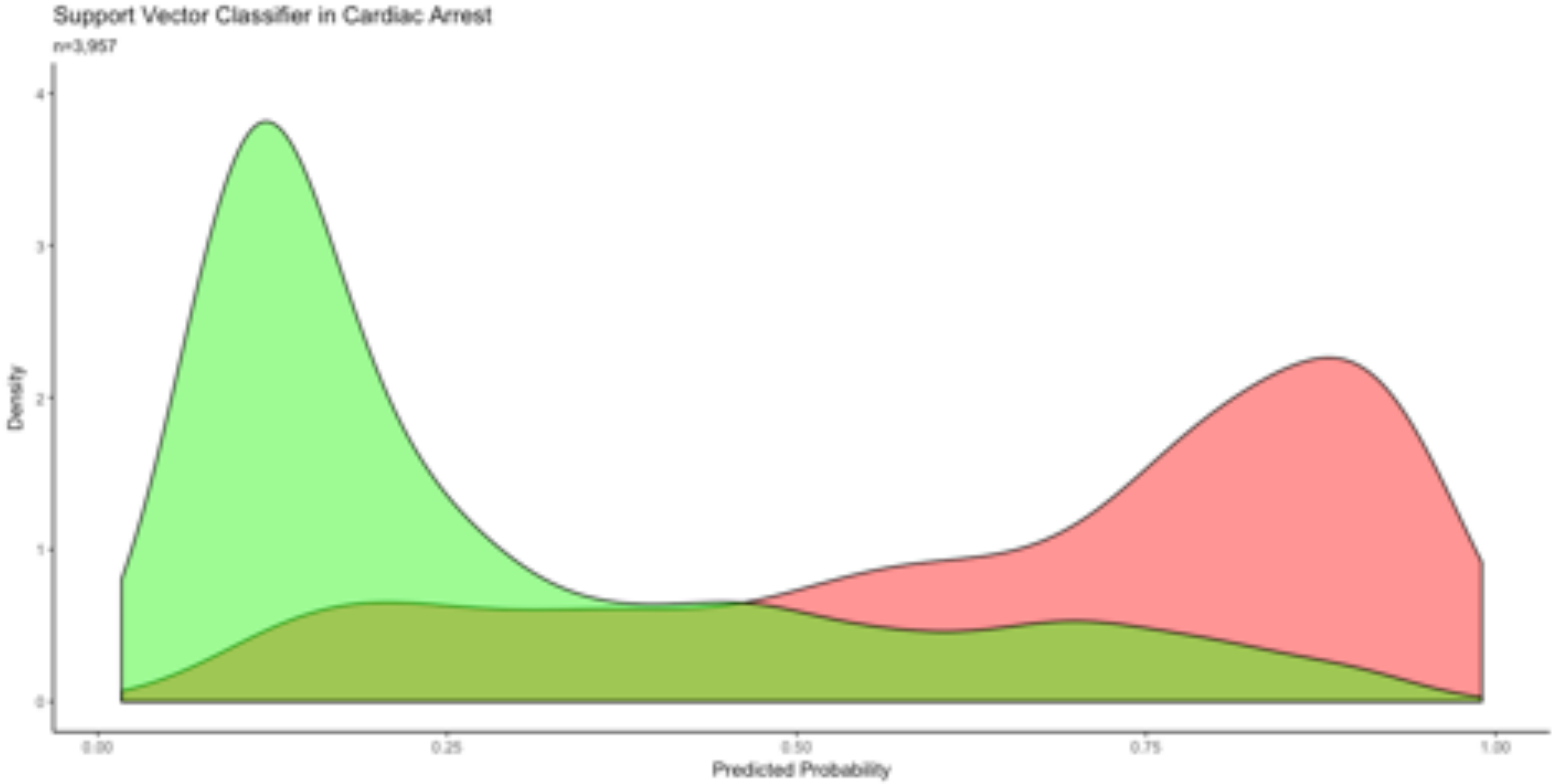

# ANZROD



# Logistic Regression

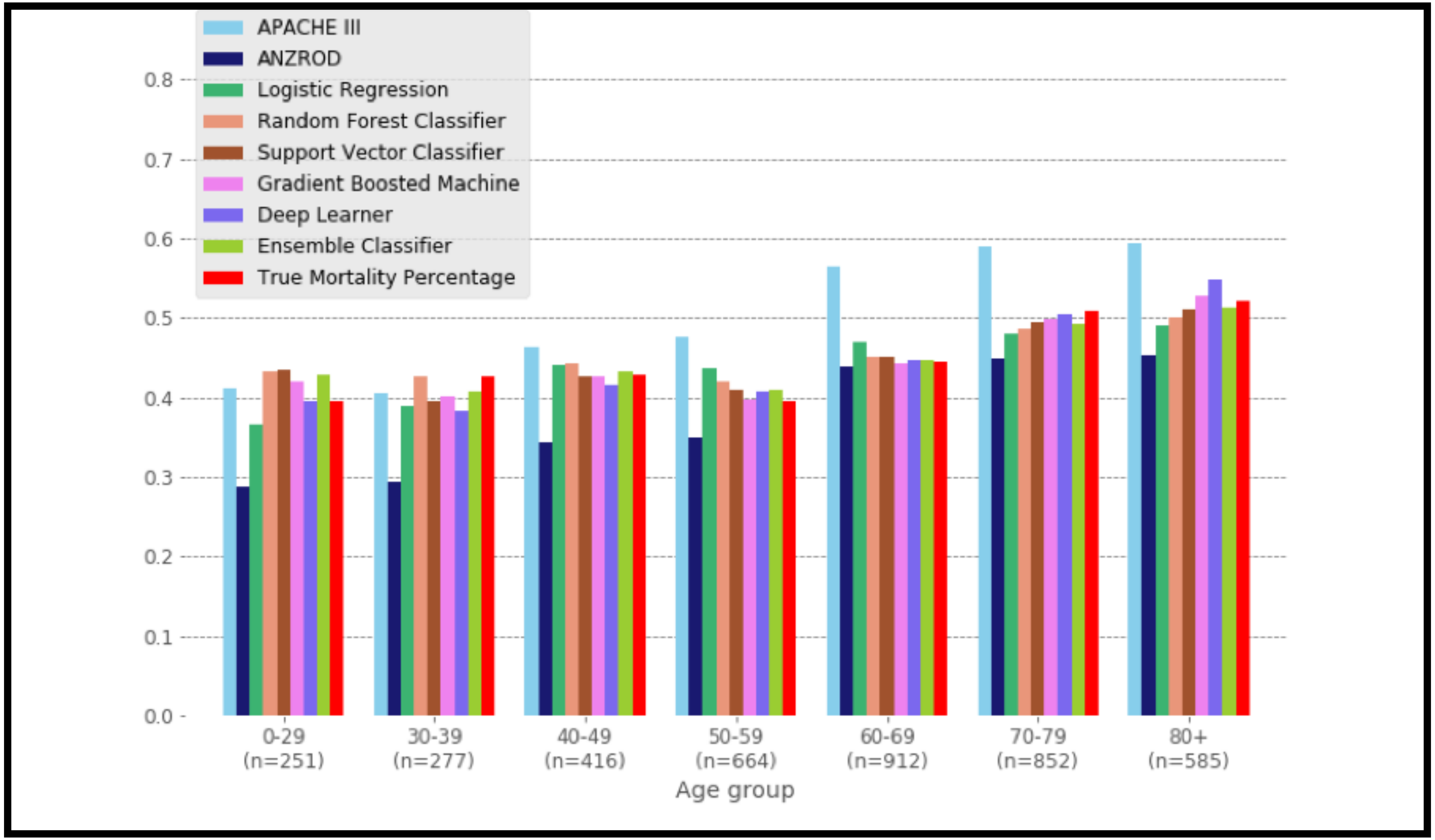



# Support Vector Machine



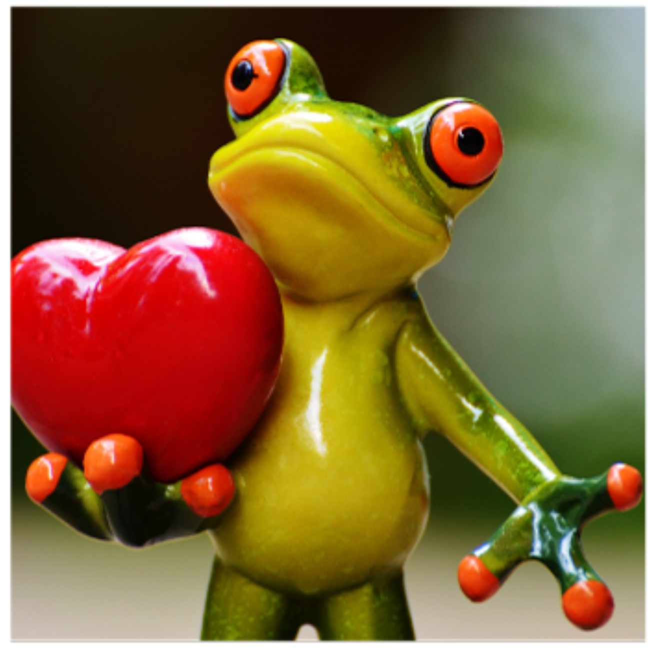
# Predicted mortality

Comparing age groups.



# LIME[1]

is a sensitivity analysis that reveals which parts of an input matter most to the eventual output.



**Turning off** all but a few interpretable components of this image reveals the probability that the model will identify …

... a tree frog
54%

... billiard balls
7%

... a balloon
5%

# Attention

shines a spotlight on where the model is looking when it makes a particular decision.

Words relevant to food quality ...

The fajita we tried was tasteless and burned and the mole sauce was way too sweet.

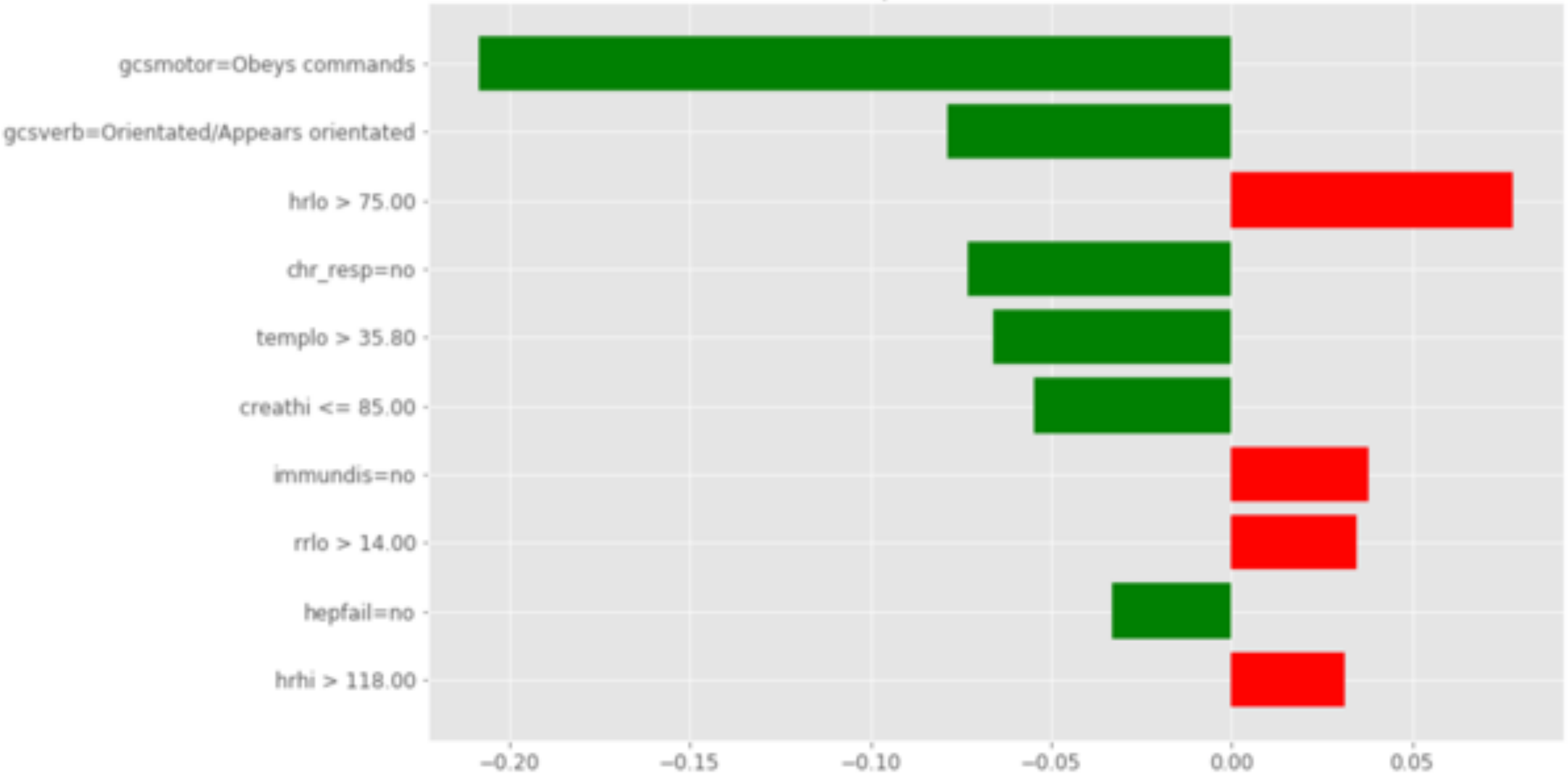
... or to service

They have one of the fastest delivery times in the city.

[1]LIME = local-interpretable-model-agnostic explanations.
**Source:** Carlos Guestrin, Marco Tulio Ribeiro, and Sameer Singh, "Introduction to local interpretable model-agnostic explanations (LIME)," August 12, 2016, O'Reilly, oreilly.com; Minlie Huang, Yequan Wang, Li Zhao, and Xiaoyan Zhu, *Attention-based LSTM for aspect-level sentiment classification*, Tsinghua University; Pixabay

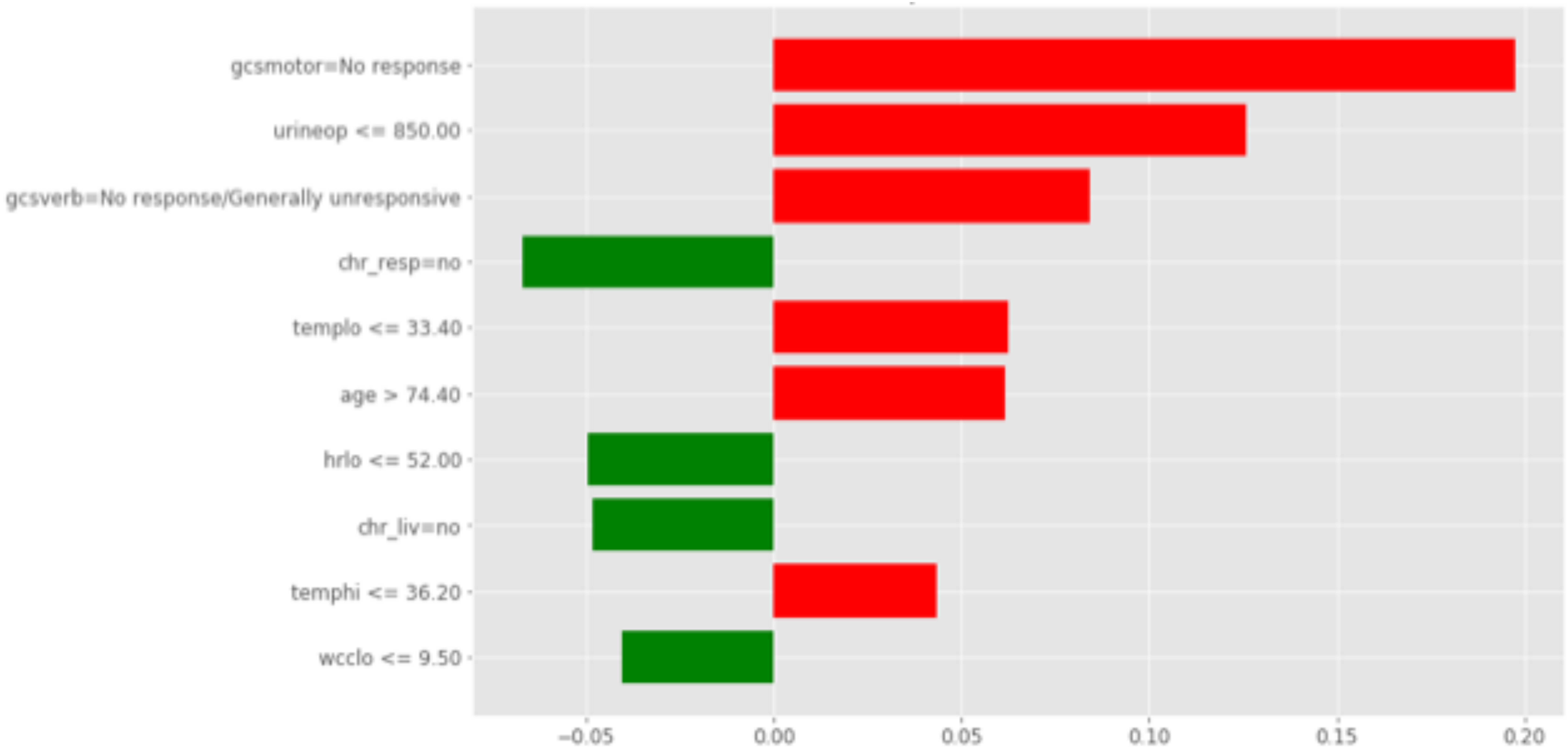
# Explaining the model

High probability of survival (83%).



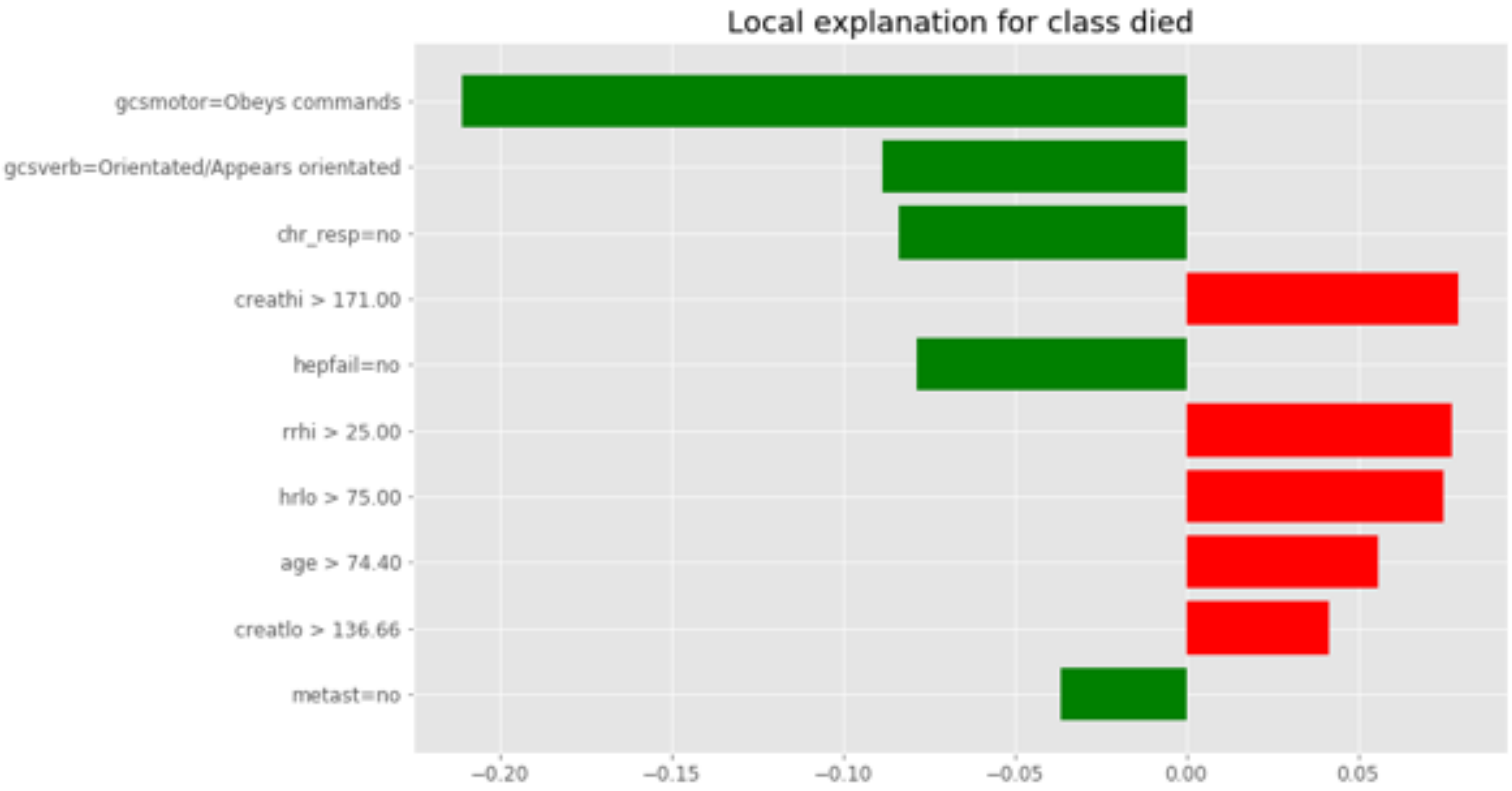
# Explaining the model

Low probability of survival (27%).



# Explaining the model

Incorrect prediction; 78% probability of survival, however the patient did not survive.



# Conclusions.

Predicting, explaining, and future directions.

able to advance on existing critical care models and statistical methods

able to add explainability