Many Good Models Leads To…

Cynthia Rudin

Gilbert, Louis, and Edward Lehrman
Distinguished Professor of Computer Science

Duke University

Joint work with students and former students Chudi Zhong, Jiachang Liu, Zhi Chen, Lesia Semenova, Hayden McTavish, Rui Xin, Jon Donnelly, Srikar Katta, Varun Babbar, Jiayun Dong, and faculty colleagues Margo Seltzer and Ron Parr

Amazing Things Come From Having Many Good Models

Cynthia Rudin [1*] Chudi Zhong [1] Lesia Semenova [1] Margo Seltzer [2] Ronald Parr [1] Jiachang Liu [1]
Srikar Katta [1] Jon Donnelly [1] Harry Chen [1] Zachery Boner [1]

ICML 2024 spotlight

Japanese movie, 1950

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

McCullagh and Nelder (1989) “Data will often point with almost equal emphasis on several possible models, and it is important that the statistician recognize and accept this.”

Breiman (2001): “What I call the Rashomon Effect is that there is often a multitude of different descriptions [equations $f(x)$] in a class of functions giving about the same minimum error rate.”

Amazing Things Come From Having Many Good Models

Cynthia Rudin [1*] Chudi Zhong [1] Lesia Semenova [1] Margo Seltzer [2] Ronald Parr [1] Jiachang Liu [1]
Srikar Katta [1] Jon Donnelly [1] Harry Chen [1] Zachery Boner [1]

ICML 2024 spotlight

We address how the Rashomon Effect
(1) impacts the existence of simple-yet-accurate models

fico.force.com/FICOCommunity/s/explainable-machine-learning-challen

1 of 2 matches Contains blac

FICO COMMUNITY

Search... SEARCH

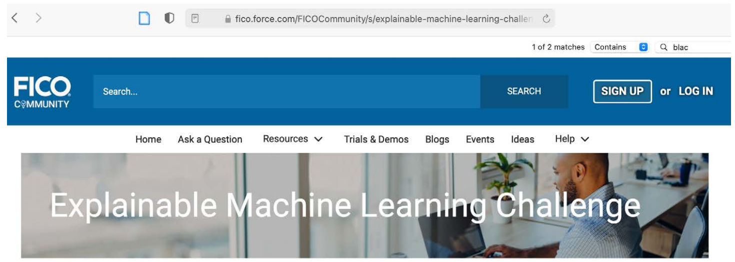
SIGN UP or LOG IN

Home | Ask a Question | Resources | Trials & Demos | Blogs | Events | Ideas | Help

Home Equity Line of Credit (HELOC) Dataset

This competition focuses on an anonymized dataset of Home Equity Line of Credit (HELOC) applications made by real homeowners. A HELOC is a line of credit typically offered by a bank as a percentage of home equity (the difference between the current market value of a home and its purchase price). The customers in this dataset have requested a credit line in the range of $5,000 - $150,000. The fundamental task is to use the information about the applicant in their credit report to predict whether they will repay their HELOC account within 2 years. This prediction is then used to decide whether the homeowner qualifies for a line of credit and, if so, how much credit should be extended.

The data



- ~10K loan applicants
- Factors:
  - External Risk Estimate
  - Months Since Oldest Trade Open
  - Months Since Most Recent Trade Open
  - Average Months In File
  - Number of Satisfactory Trades
  - Number Trades 60+ Ever
  - Number Trades 90+ Ever
  - Number of Total Trades
  - Number Trades Open In Last 12 Months
  - Percent Trades Never Delinquent
  - Months Since Most Recent Delinquency
  - Max Delinquency / Public Records Last 12 Months
  - Max Delinquency Ever
  - :

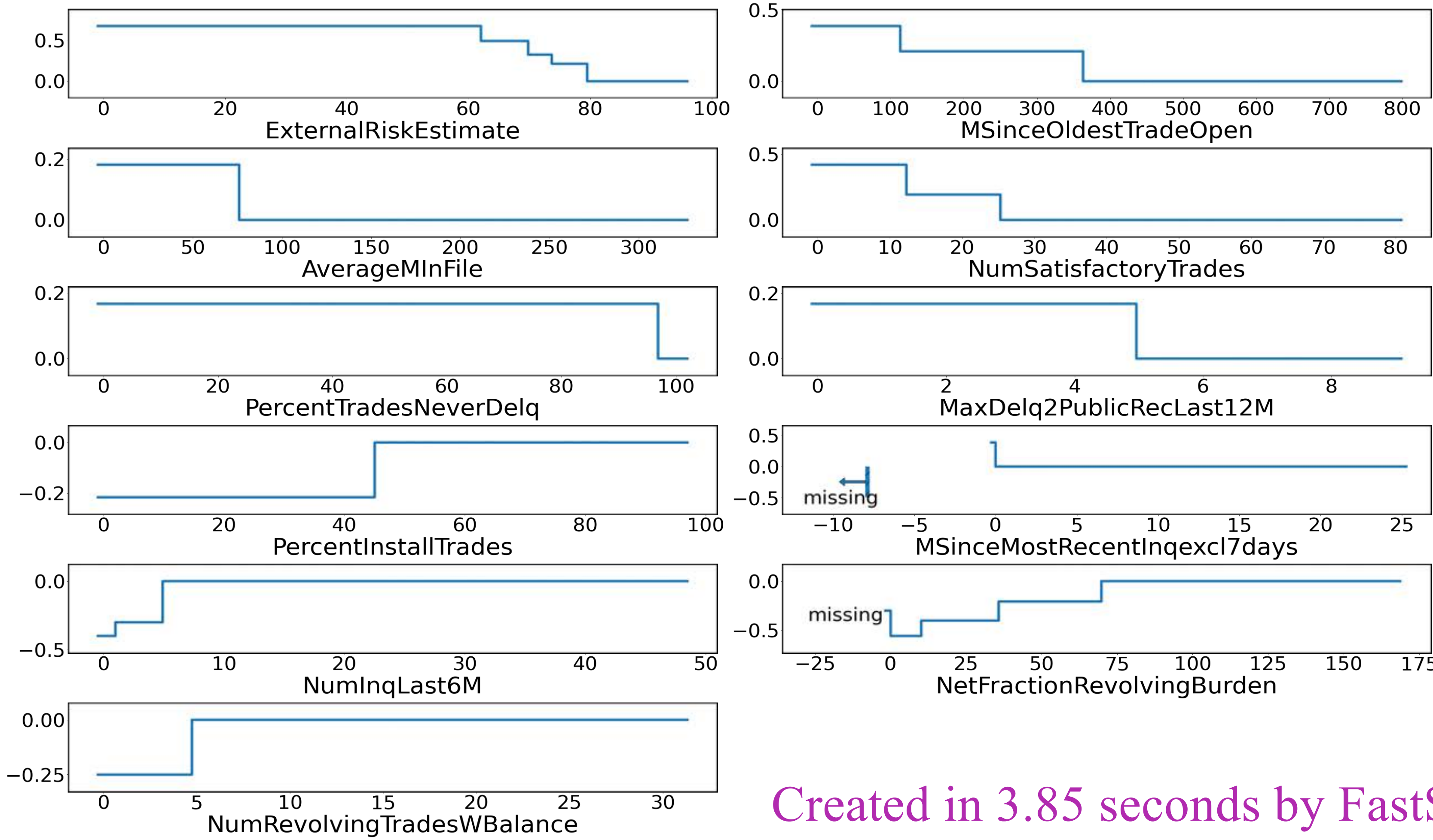
Best black box accuracy
(boosted decision trees) 73%

Best black box AUC
(2-layer neural network) .80

Performance of FastSparse (Liu et al., 2022)
Train/Test Accuracy: 73.05±0.28, 72.35±1.24
Train/Test AUC: .803±0.0025, .791±.0010

On the next slide…
The whole machine learning model

Generalized Additive Model on the FICO Dataset



Created in 3.85 seconds by FastSparse

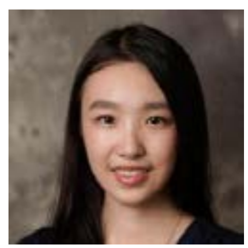
The FastSparse Algorithm

Fast Sparse Classification for Generalized Linear and Additive Models

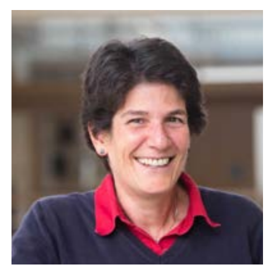
Jiachang Liu[1] **Chudi Zhong**[1] **Margo Seltzer**[2] **Cynthia Rudin**[1]
[1]Duke University [2] University of British Columbia
{jiachang.liu, chudi.zhong}@duke.edu, mseltzer@cs.ubc.ca, cynthia@cs.duke.edu

AISTATS, 2022

Jiachang Liu



Chudi Zhong



Margo Seltzer

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

Breiman (2001): “Accuracy generally requires more complex prediction methods. Simple and interpretable functions do not make the most accurate predictors.”

2001 Omnitech Custom Pentium 3 computer - Model OTS-8100SD02815 Transcendental Airwaves

https://novgblog.wordpress.com/2016/05/21/90s-games-computers-life-part-1/

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

Breiman (2001): "Accuracy generally requires more complex prediction methods. Simple and interpretable functions do not make the most accurate predictors."

"On interpretability, trees rate an A+."

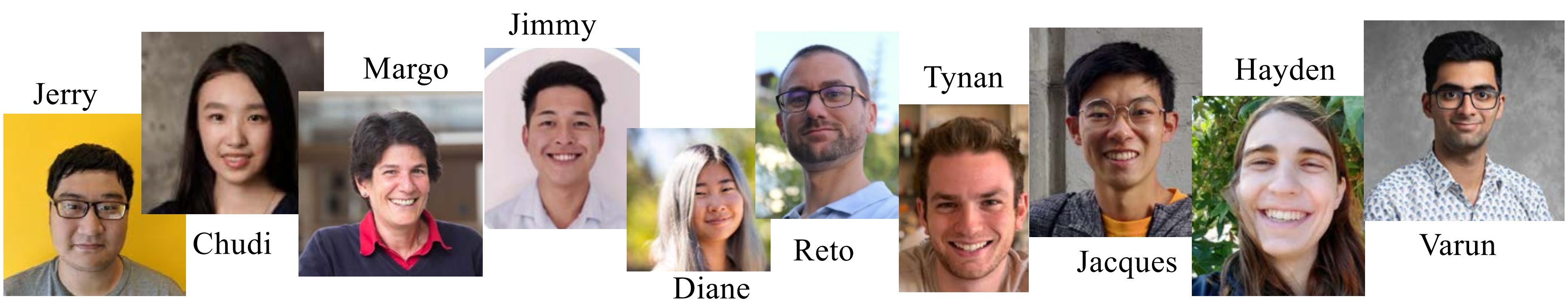
"While trees rate an A+ on interpretability, they are good, but not great, predictors. Give them, say, a B on prediction."

2001 Omnitech Custom Pentium 3 computer - Model OTS-8100SD02815 Transcendental Airwaves

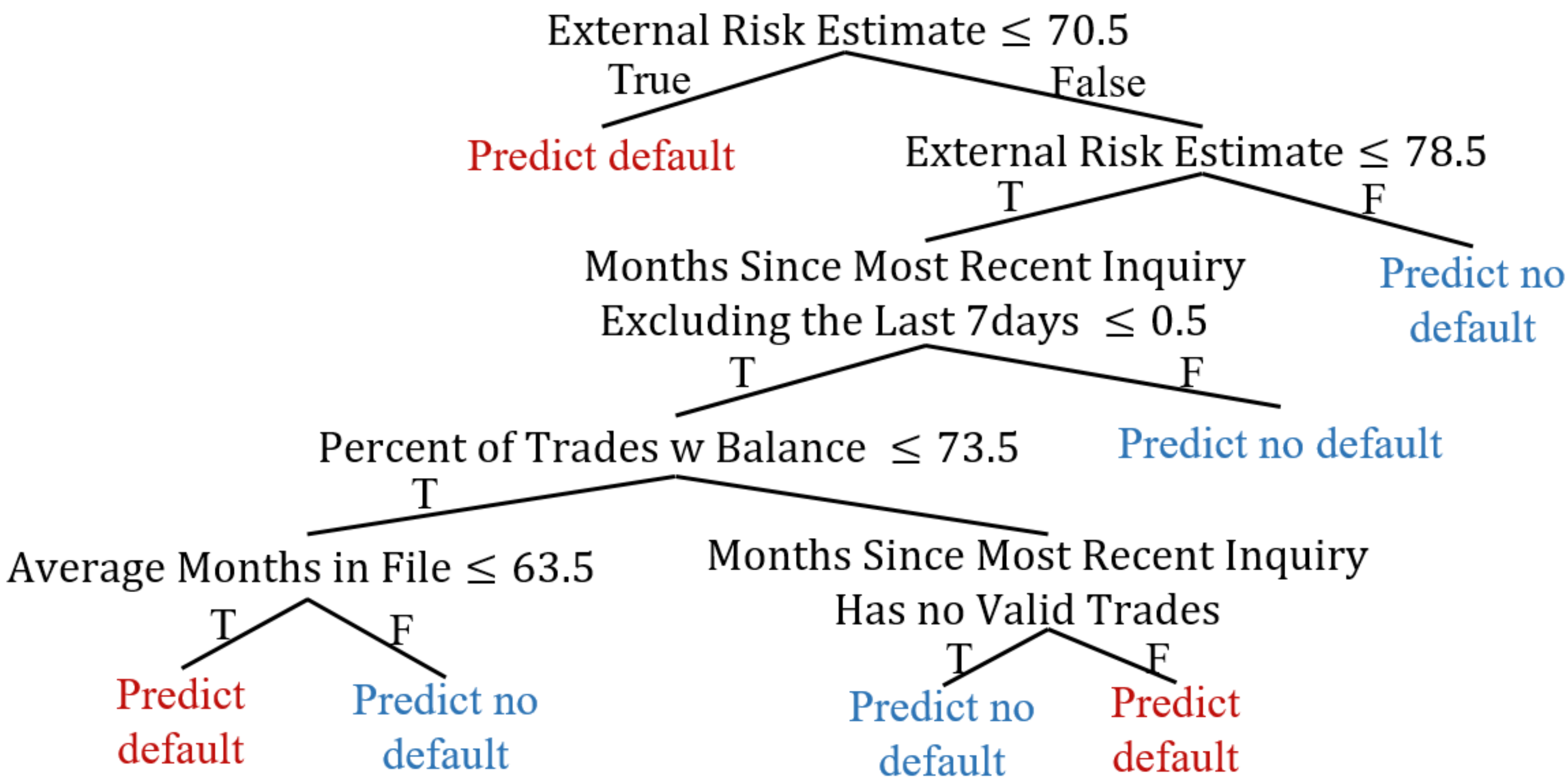
Generalized Optimal Sparse Decision Trees (GOSDT)

also GOSDT+Guesses, SPLIT

with Chudi Zhong, Hayden McTavish, Jimmy Lin, Reto Achermann, Ilias Karimalis, Jacques Chen, Diane Hu, Tynan Seltzer, Margo Seltzer, Bingyao (Jerry) Wang, Varun Babbar



https://github.com/Jimmy-Lin/GeneralizedOptimalSparseDecisionTrees



Created in 8.1 seconds by GOSDT
~72% accuracy

Why do simple-yet-accurate models exist?

$$0 < P(y=1|\mathbf{x}) < 1$$

On the Existence of Simpler Machine Learning Models

Lesia Semenova, Cynthia Rudin, and Ronald Parr

{*lesia*, *cynthia*, *parr*}@cs.duke.edu

ACM Conference on Fairness, Accountability, and Transparency, 2022

A Path to Simpler Models Starts With Noise

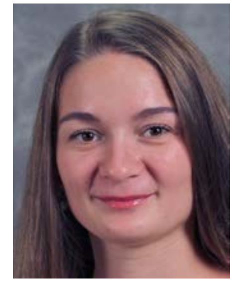
Lesia Semenova **Harry Chen** **Ronald Parr** **Cynthia Rudin**
Department of Computer Science, Duke University
{lesia.semenova,harry.chen084,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2023

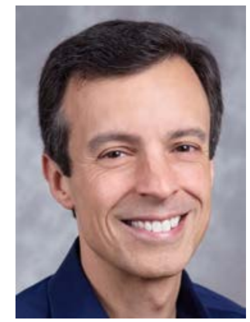
Using Noise to Infer Aspects of Simplicity Without Learning

Zachery Boner* **Harry Chen*** **Lesia Semenova*** **Ronald Parr** **Cynthia Rudin**
Department of Computer Science, Duke University
{zachery.boner,harry.chen084,lesia.semenova,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2024



Lesia Semenova



Ron Parr

Harry Chen

Zack Boner

On the Existence of Simpler Machine Learning Models

Lesia Semenova, Cynthia Rudin, and Ronald Parr

{*lesia, cynthia, parr*}@cs.duke.edu

ACM Conference on Fairness, Accountability, and Transparency, 2022

A Path to Simpler Models Starts With Noise

Lesia Semenova **Harry Chen** **Ronald Parr** **Cynthia Rudin**
Department of Computer Science, Duke University
{lesia.semenova,harry.chen084,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2023

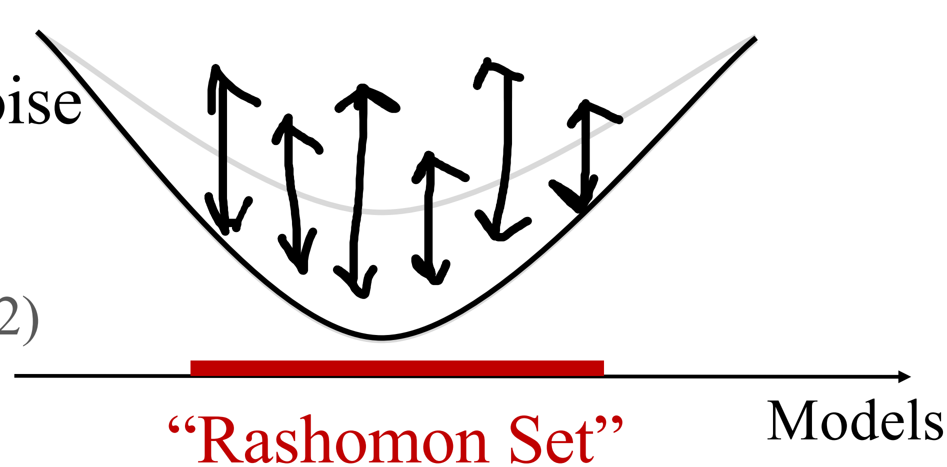
"Rashomon set"

=

set of optimal and almost optimal models

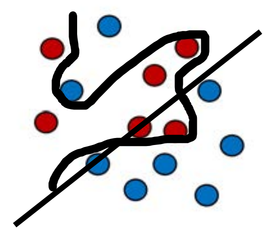
1) Noise in the world leads to increased variance of the outcomes and loss terms.
2) Higher error variance leads to worse generalization. (Bernstein's inequality)
3) Cross-validation reveals this. Analyst compensates by simplifying the class.

Prediction loss
with noise

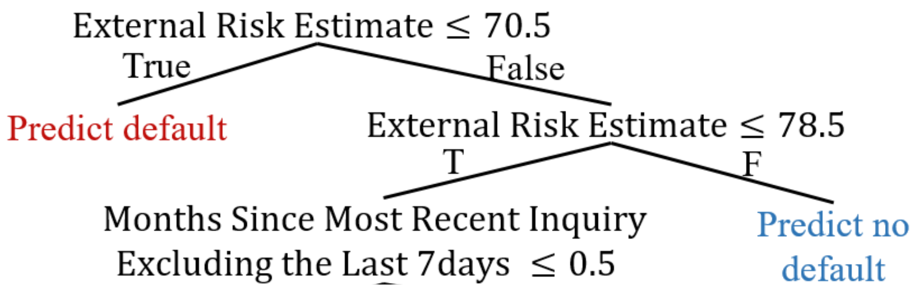


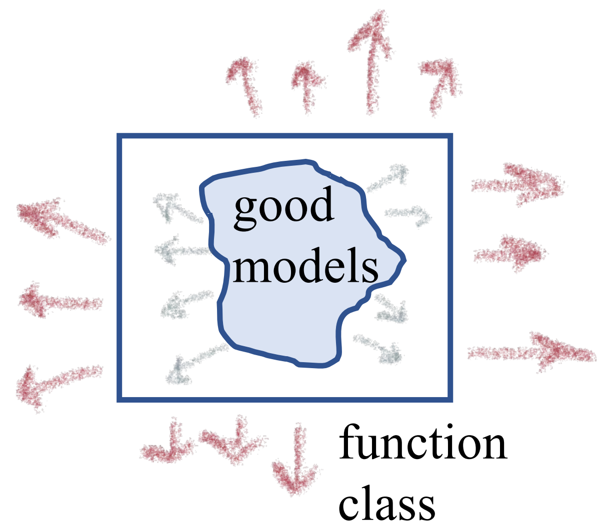
Uniform label flipping noise (Theorem 2)
Labels flip with probability p(x) (Theorem 12)
"Margin noise" (Theorem 15)

1) Noise in the world leads to increased variance of the outcomes and loss terms.
2) Higher error variance leads to worse generalization. (Bernstein's inequality)
3) Cross-validation reveals this. Analyst compensates by simplifying the class.



1) Noise in the world leads to increased variance of the outcomes and loss terms.
2) Higher error variance leads to worse generalization. (Bernstein's inequality)
3) Cross-validation reveals this. Analyst compensates by simplifying the class.
4) New smaller function class has a larger *Rashomon ratio* - large fraction of good models.

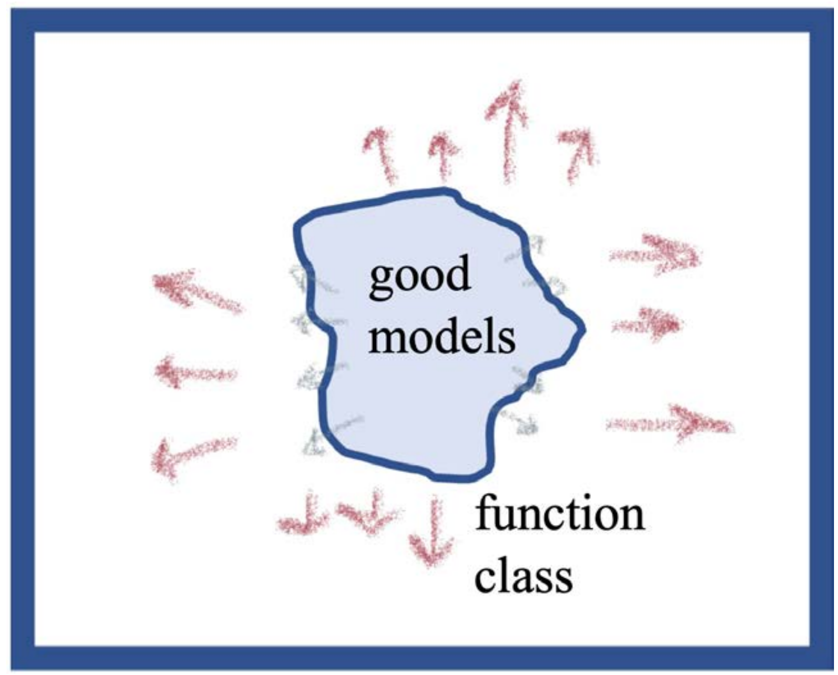




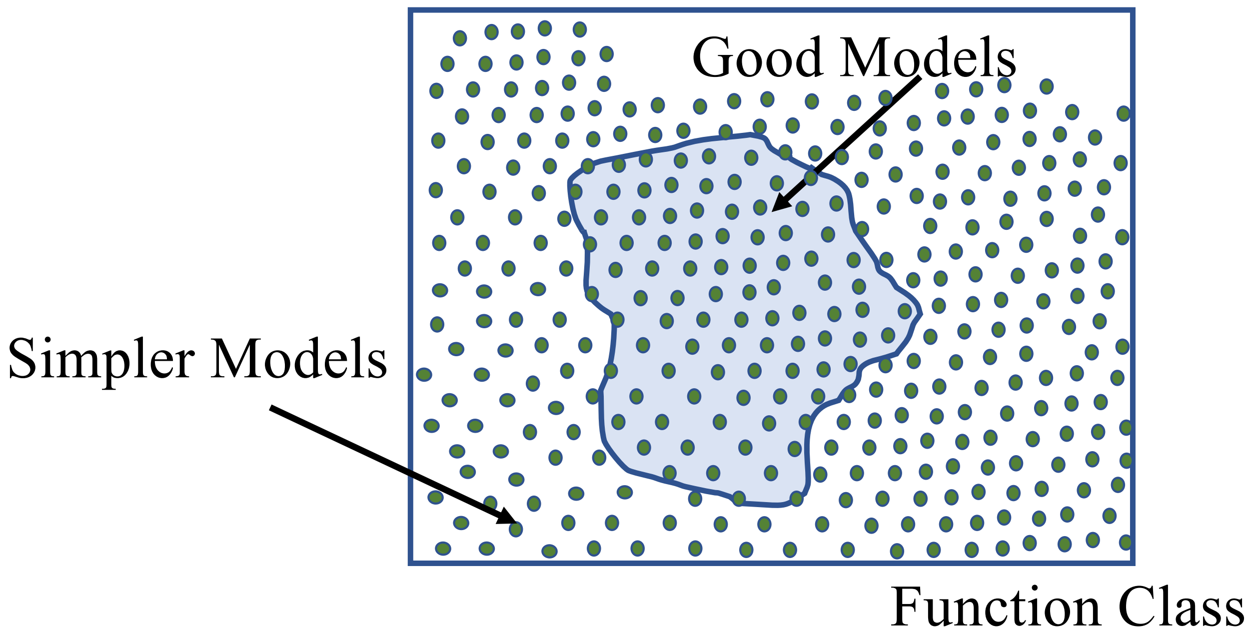
1) Noise in the world leads to increased variance of the outcomes and errors.
2) Higher error variance leads to worse generalization. (Bernstein's inequality)
3) Cross-validation reveals this. Analyst compensates by simplifying the class.
4) New smaller function class has a larger *Rashomon ratio* - large fraction of good models.

Proposition 6 (Rashomon ratio is larger for decision trees of smaller depth)

Theorem 7 (Rashomon ratio increases with noise for ridge regression)



1) Noise in the world leads to increased variance of the outcomes and errors.
2) Higher error variance leads to worse generalization. (Bernstein's inequality)
3) Cross-validation reveals this. Analyst compensates by simplifying the class.
4) New smaller function class has a larger *Rashomon ratio* - large fraction of good models.



On the Existence of Simpler Machine Learning Models

Lesia Semenova, Cynthia Rudin, and Ronald Parr

{*lesia, cynthia, parr*}@cs.duke.edu

ACM Conference on Fairness, Accountability, and Transparency, 2022

A Path to Simpler Models Starts With Noise

Lesia Semenova Harry Chen Ronald Parr Cynthia Rudin
Department of Computer Science, Duke University
{lesia.semenova,harry.chen084,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2023

Using Noise to Infer Aspects of Simplicity Without Learning

Zachery Boner* Harry Chen* Lesia Semenova* Ronald Parr Cynthia Rudin
Department of Computer Science, Duke University
{zachery.boner,harry.chen084,lesia.semenova,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2024

Noise in the world is equivalent to adding implicit regularization.

Simpler models are more likely to exist.

Using Noise to Infer Aspects of Simplicity Without Learning

Zachery Boner* **Harry Chen*** **Lesia Semenova*** **Ronald Parr** **Cynthia Rudin**
Department of Computer Science, Duke University
{zachery.boner,harry.chen084,lesia.semenova,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2024

On the Existence of Simpler Machine Learning Models

Lesia Semenova, Cynthia Rudin, and Ronald Parr

{*lesia*, *cynthia*, *parr*}@cs.duke.edu

ACM Conference on Fairness, Accountability, and Transparency, 2022

A Path to Simpler Models Starts With Noise

Lesia Semenova Harry Chen Ronald Parr Cynthia Rudin
Department of Computer Science, Duke University
{lesia.semenova,harry.chen084,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2023

Using Noise to Infer Aspects of Simplicity Without Learning

Zachery Boner* Harry Chen* Lesia Semenova* Ronald Parr Cynthia Rudin
Department of Computer Science, Duke University
{zachery.boner,harry.chen084,lesia.semenova,ronald.parr,cynthia.rudin}@duke.edu

NeurIPS, 2024

Practitioners don’t need to understand any of this math!

The "Rashomon Set" Theory

Implication:

Optimizing for simplicity won't sacrifice accuracy for a vast set of problems.

Amazing Things Come From Having Many Good Models

Cynthia Rudin[1*] Chudi Zhong[1] Lesia Semenova[1] Margo Seltzer[2] Ronald Parr[1] Jiachang Liu[1]
Srikar Katta[1] Jon Donnelly[1] Harry Chen[1] Zachery Boner[1]

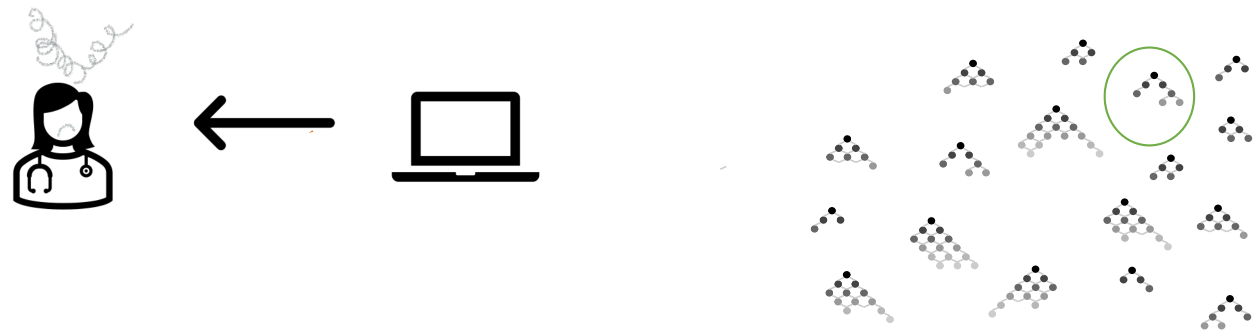
ICML 2024 spotlight

We address how the Rashomon Effect impacts:

(1) the existence of simple-yet-accurate models

(2) flexibility to address user preferences, such as fairness and monotonicity, without losing performance

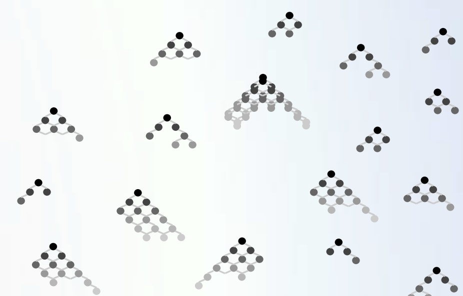
The “Interaction Bottleneck”



ML algorithms only return one model. No human interaction.

A New Paradigm of Machine Learning

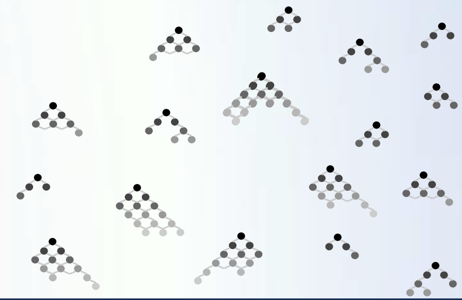
Training Set → ML Algorithm → All Good Predictive Models!



Rashomon Set

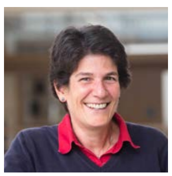
optimization + enumeration + visualization

Training Set → ML Algorithm → All Good Predictive Models!

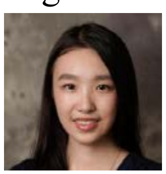


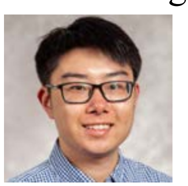
optimization + enumeration + visualization


Rui Xin

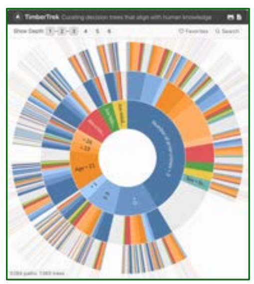

Margo Seltzer


Jay Wang


Chudi Zhong


Zhi Chen

TimberTrek



(Jay Wang et al., IEEE VIS, 2022)

TreeFARMS

Exploring the Whole Rashomon Set of Sparse Decision Trees

(Rui Xin*, Chudi Zhong*, Zhi Chen*, Takuya Takagi, Margo Seltzer, Cynthia Rudin, NeurIPS oral, 2022)

No More Interaction Bottleneck

Constraints are Now Easy

Amazing Things Come From Having Many Good Models

Cynthia Rudin [1*] Chudi Zhong [1] Lesia Semenova [1] Margo Seltzer [2] Ronald Parr [1] Jiachang Liu [1]
Srikar Katta [1] Jon Donnelly [1] Harry Chen [1] Zachery Boner [1]

ICML 2024 spotlight

We address how the Rashomon Effect impacts:

(1) the existence of simple-yet-accurate models

(2) flexibility to address user preferences, such as fairness and monotonicity, without losing performance

(3) uncertainty in predictions, fairness, and explanations

(4) reliable variable importance

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

So here are three possible pictures with RSS or test set error within 1.0% of each other:

Picture 1

$$y = 2.1 + 3.8x_3 - 0.6x_8 + 83.2x_{12} - 2.1x_{17} + 3.2x_{27},$$

Picture 2

$$y = -8.9 + 4.6x_5 + 0.01x_6 + 12.0x_{15} + 17.5x_{21} + 0.2x_{22},$$

Picture 3

$$y = -76.7 + 9.3x_2 + 22.0x_7 - 13.2x_8 + 3.4x_{11} + 7.2x_{28}.$$

Which one is better? The problem is that each one tells a different story about which variables are important.

The Rashomon Effect also occurs with decision trees and neural nets. In my experiments with trees,

(3) uncertainty in predictions, fairness,

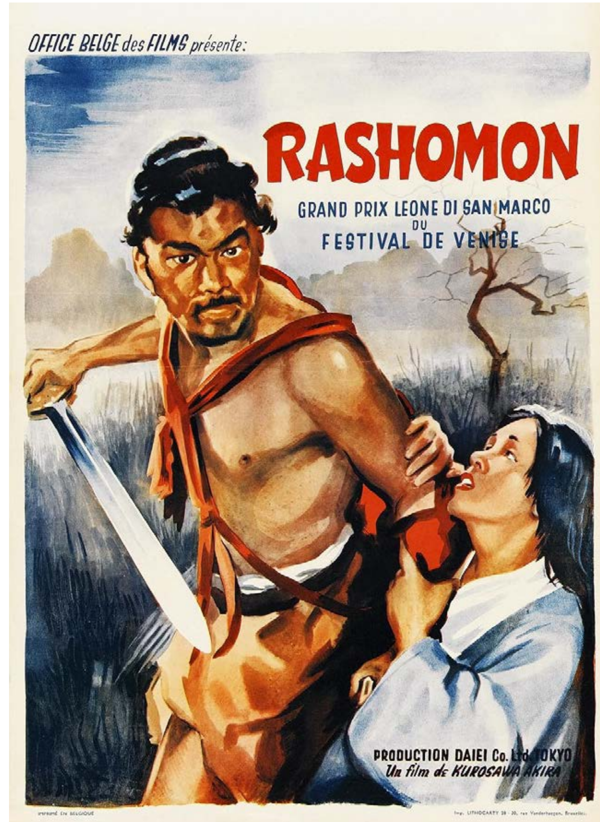
(4) reliable variable importance

Want importance for the data generation process, not any one model.

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman



Which one is better? The problem is that each one tells a different story about which variables are important.

XAI often asks the wrong question.

Understanding XAI: SHAP, LIME, and Other Key Techniques

By RoX818 / November 5, 2024

https://aicompetence.org/understanding-xai-shap-lime-and-beyond/

The Core Goals of XAI

In general, XAI frameworks aim to:

- Increase **trust** in model decisions by making them understandable.
- Provide insights into model **biases** or **weaknesses**.
- Enhance model **debugging** and **improvement**.
- Facilitate **compliance** with ethical and legal standards (e.g., GDPR).

Understanding XAI: SHAP, LIME, and Other Key Techniques

By RoX818 / November 5, 2024

https://aicompetence.org/understanding-xai-shap-lime-and-beyond/

The Core Goals of XAI

In general, XAI frameworks aim to:

able.

PR).

What Is SHAP?

SHAP (SHapley Additive exPlanations) is a popular XAI technique based on *Shapley values*, a concept from cooperative game theory. The Shapley values evaluate each feature's contribution to the model's prediction, giving a sense of "feature importance." SHAP aims to distribute the "payoff" (or prediction) fairly among features, indicating how much each feature contributed to the output.

Understanding XAI: SHAP, LIME, and Other Key Techniques

By RoX818 / November 5

https://aicompete

Key Strengths of SHAP

- **Theoretically Sound**: SHAP explanations are grounded in Shapley values, ensuring each feature's contribution is additive and fair.
- **Global and Local Interpretability**: SHAP can explain individual predictions (local) and broader model behavior (global).
- **Visualizations**: SHAP provides rich visuals, like beeswarm plots and summary plots, to make interpretability more accessible.

What Is SI

SHAP (SHapl

values, a concept from cooperative game theory. The Shapley values evaluate each feature's contribution to the model's prediction, giving a sense of "feature importance." SHAP aims to distribute the "payoff" (or prediction) fairly among features, indicating how much each feature contributed to the output.

Understanding XAI: SHAP, LIME, and Other Key Techniques

By RoX818 / November 5

https://aicompete

Key Strengths of SHAP

- **Theoretically Sound**: SHAP explanations are grounded in Shapley values, ensuring each feature's contribution is additive and fair.

Limitations of SHAP

While SHAP provides powerful insights, it can be computationally expensive, especially with complex models. Calculating Shapley values for every feature and instance may be prohibitive for large datasets or deep neural networks.

W

SH

val

fea

SHAP aims to distribute the "payoff" (or prediction) fairly among features, indicating how much each feature contributed to the output.

Need model independent variable importance

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

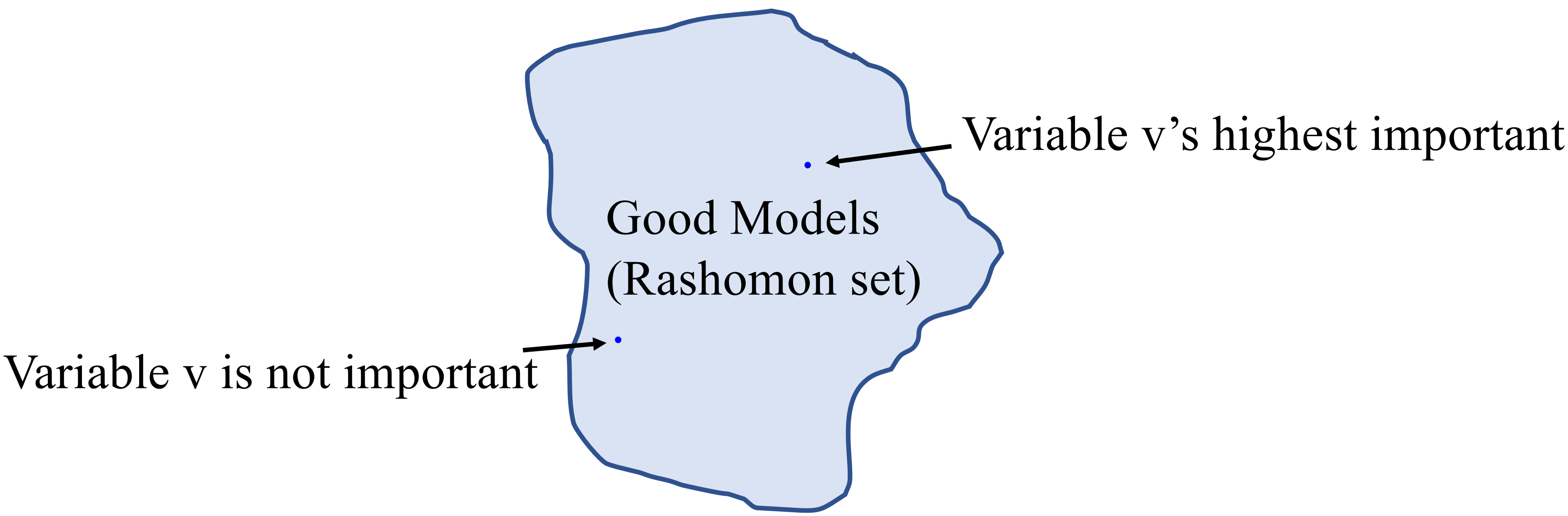
Leo Breiman

The goals in statistics are to use data to predict and to get information about the underlying data mechanism.

All Models are Wrong, but *Many* are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously

JMLR, 2019

Aaron Fisher, Cynthia Rudin, Francesca Dominici



Model Class Reliance range of variable importance within Rashomon set

Exploring the cloud of variable importance for the set of all good models

Jiayun Dong[1] and Cynthia Rudin[2]

Nature Machine Intelligence, 2020

Variable Importance Clouds
&
Variable Importance Diagrams

plot variable importance for all models within Rashomon set

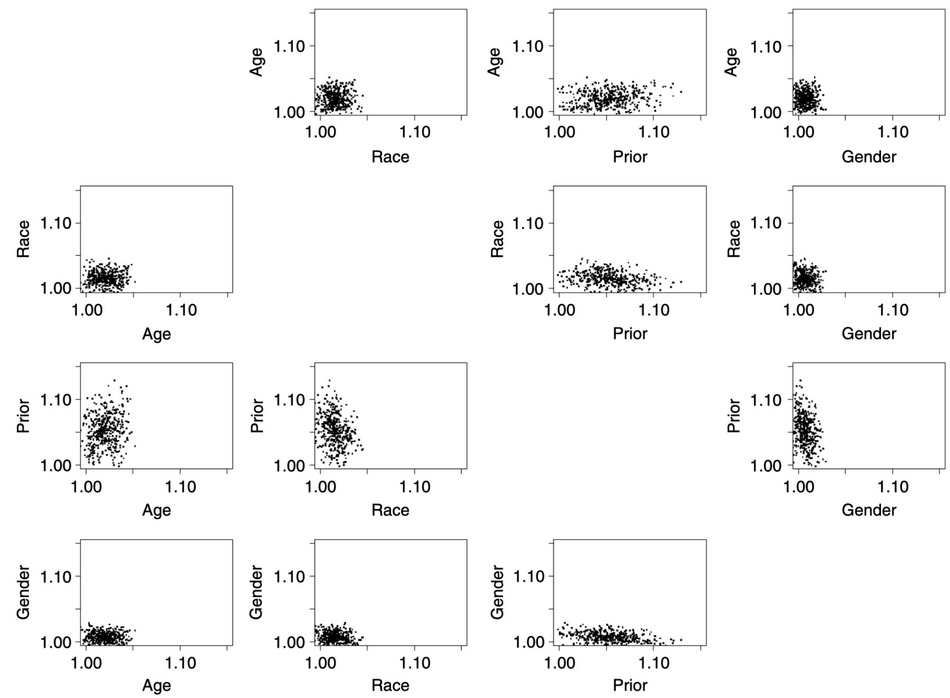
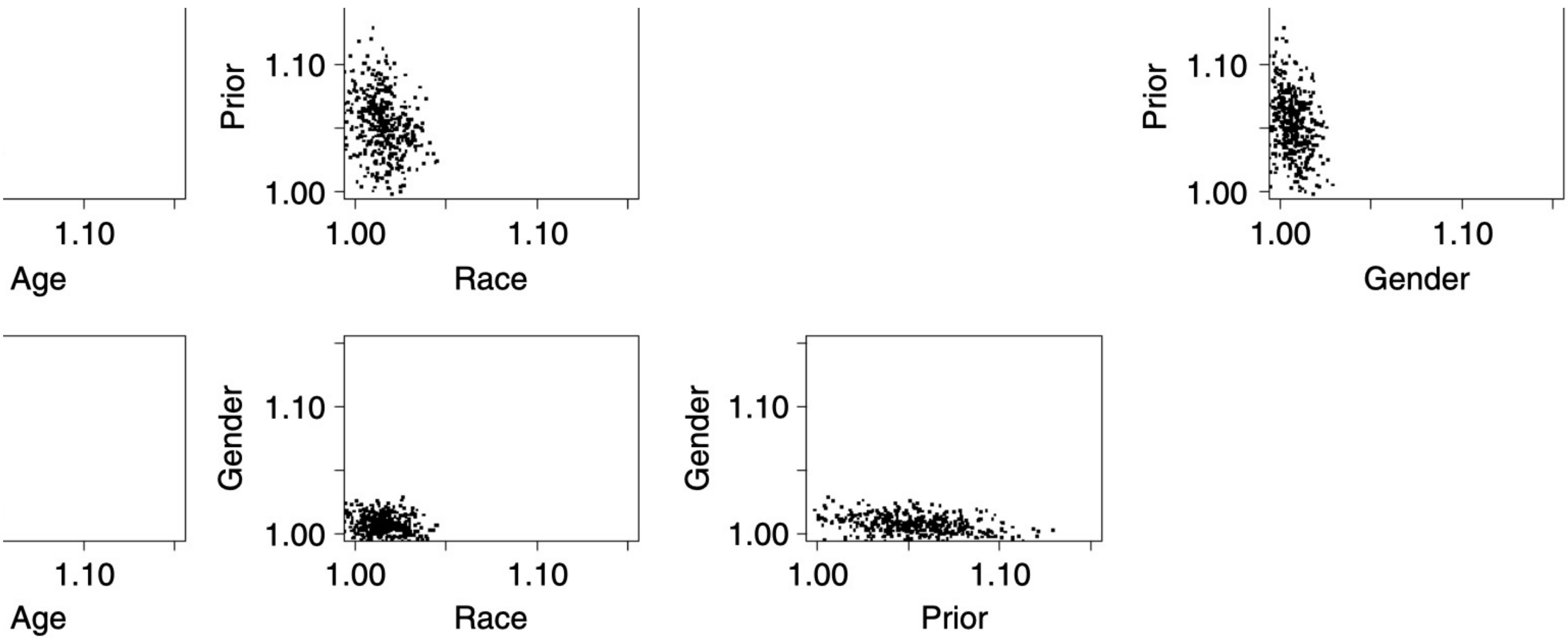


Fig. 4 | VID for recidivism using logistic regression. This is the projection of the VIC onto the space spanned by the four variables of interest: age, race, prior criminal history and gender. The point, say (1.02, 1.03), in the first diagram in the first row suggests that there is a model in the Rashomon set with reliance 1.02 on race and 1.03 on age.



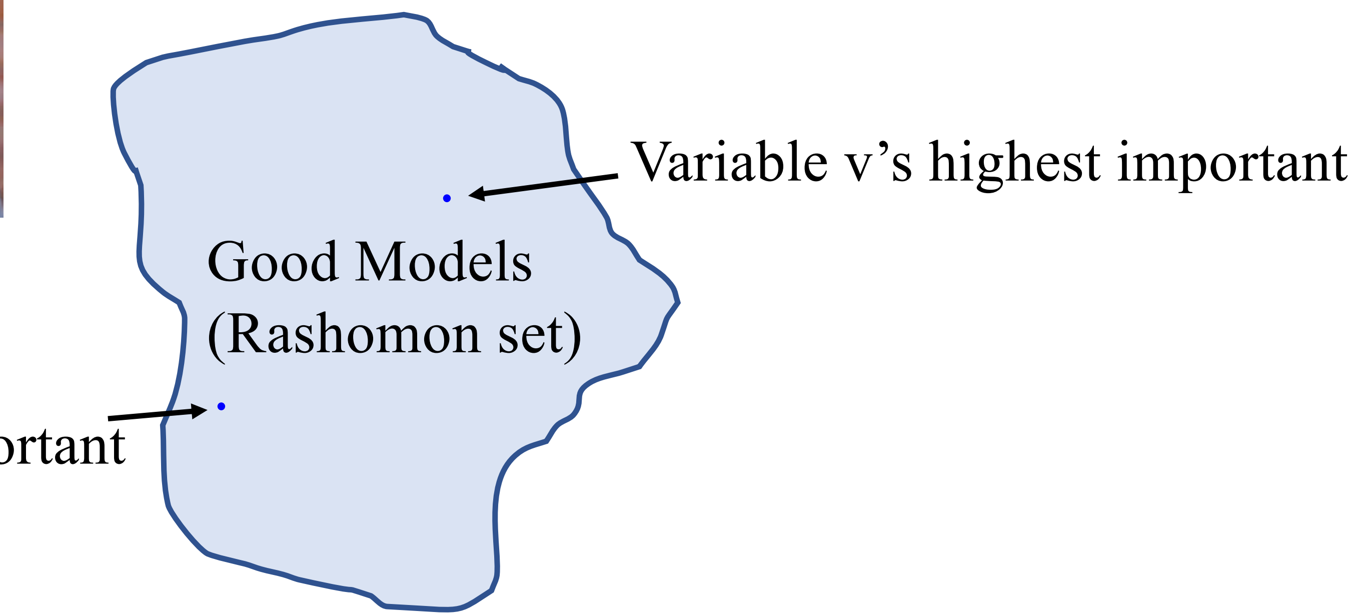
stic regression. This is the projection of the VIC onto the space spanned by the four variables of int
e point, say (1.02, 1.03), in the first diagram in the first row suggests that there is a model in the Rasl
.

But we want more information, not just a range or cloud.

Perhaps a probability distribution?

The Rashomon Importance Distribution: Getting RID of Unstable, Single Model-based Variable Importance

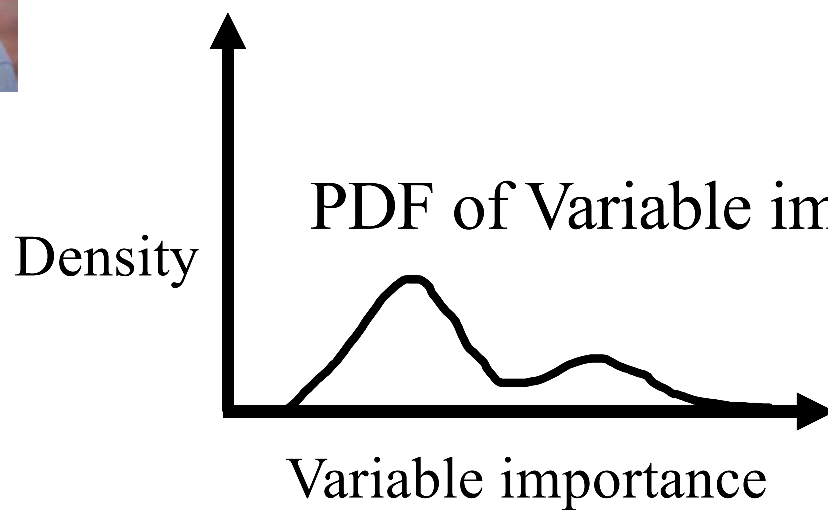
Jon Donnelly*, Srikar Katta*, Cynthia Rudin, and Edward Browne. NeurIPS 2023 (spotlight)



The Rashomon Importance Distribution: Getting RID of Unstable, Single Model-based Variable Importance

Jon Donnelly*, Srikar Katta*, Cynthia Rudin, and Edward Browne. NeurIPS 2023 (spotlight)

- Gives a PDF for each variable’s importance
- Stable (across bootstraps)
- Model independent (averaged across Rashomon sets)



Amazing Things Come From Having Many Good Models

Cynthia Rudin [1*] Chudi Zhong [1] Lesia Semenova [1] Margo Seltzer [2] Ronald Parr [1] Jiachang Liu [1]
Srikar Katta [1] Jon Donnelly [1] Harry Chen [1] Zachery Boner [1]

ICML 2024 spotlight

We address how the Rashomon Effect impacts:

(1) the existence of simple-yet-accurate models

(2) flexibility to address user preferences, such as fairness and monotonicity, without losing performance

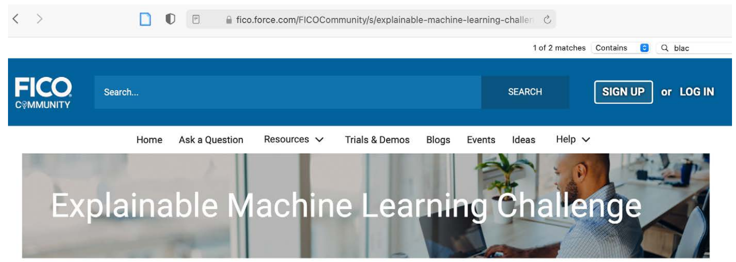
(3) uncertainty in predictions, fairness, and explanations

(4) reliable variable importance

(5) algorithm choice, specifically, providing advanced knowledge of which algorithms might be suitable for a given problem

Which algorithm should I use?

The data



- ~10K loan applicants
- Factors:
  - External Risk Estimate
  - Months Since Oldest Trade Open
  - Months Since Most Recent Trade Open
  - Average Months In File
  - Number of Satisfactory Trades
  - Number Trades 60+ Ever
  - Number Trades 90+ Ever
  - Number of Total Trades
  - Number Trades Open In Last 12 Months
  - Percent Trades Never Delinquent
  - Months Since Most Recent Delinquency
  - Max Delinquency / Public Records Last 12 Months
  - Max Delinquency Ever
  - :

Best black box accuracy
(boosted decision trees) 73%

Best black box AUC
(2-layer neural network) .80

Performance of FastSparse (Liu et al., 2022)
Train/Test Accuracy: 73.05±0.28, 72.35±1.24
Train/Test AUC: .803±0.0025, .791±.0010

On the next slide…
The whole machine learning model

Which algorithm should I use?

For noisy tabular data (loans, recidivism, etc.):

- Don’t use CART.
- Most other algorithms will perform similarly for many problems.
- Interpretable models are much easier to work with in practice.
- In the Rashomon Set paradigm: TreeFARMS/TimberTrek, FastSparse or OKRidge with GAMChanger, or FasterRisk with Riskomon.

For NLP, I don’t know the answer.

For images, time series (ECG, PPG, etc.), or other types of signals:

- Try interpretable neural networks (ProtoPNets, or other ideas)

Challenge: Make a discovery with interpretable AI that wasn’t possible with a black box.

Predict breast cancer up to 5 years in advance?

Prof. Regina Barzilay and Cynthia
MIT Campus, July 2023

Predict breast cancer 1-5 years in advance from mammograms

Regina's paper:

RSNA Journals CME Contact Us Subscribe Email Alerts

Radiology

A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction

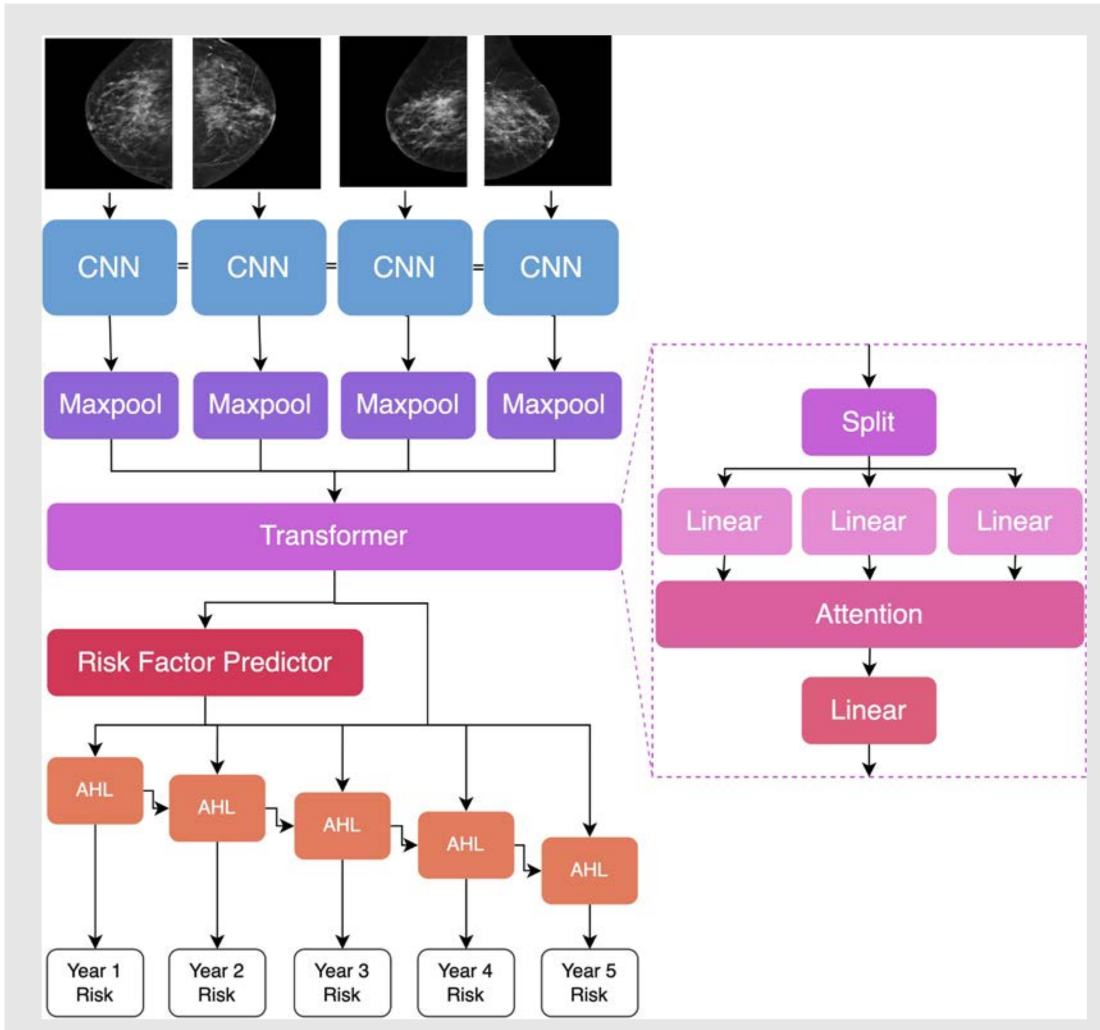
Adam Yala, Constance Lehman, Tal Schuster, Tally Portnoi, Regina Barzilay

"But Cynthia, even the doctors don't know what Mirai is doing…"

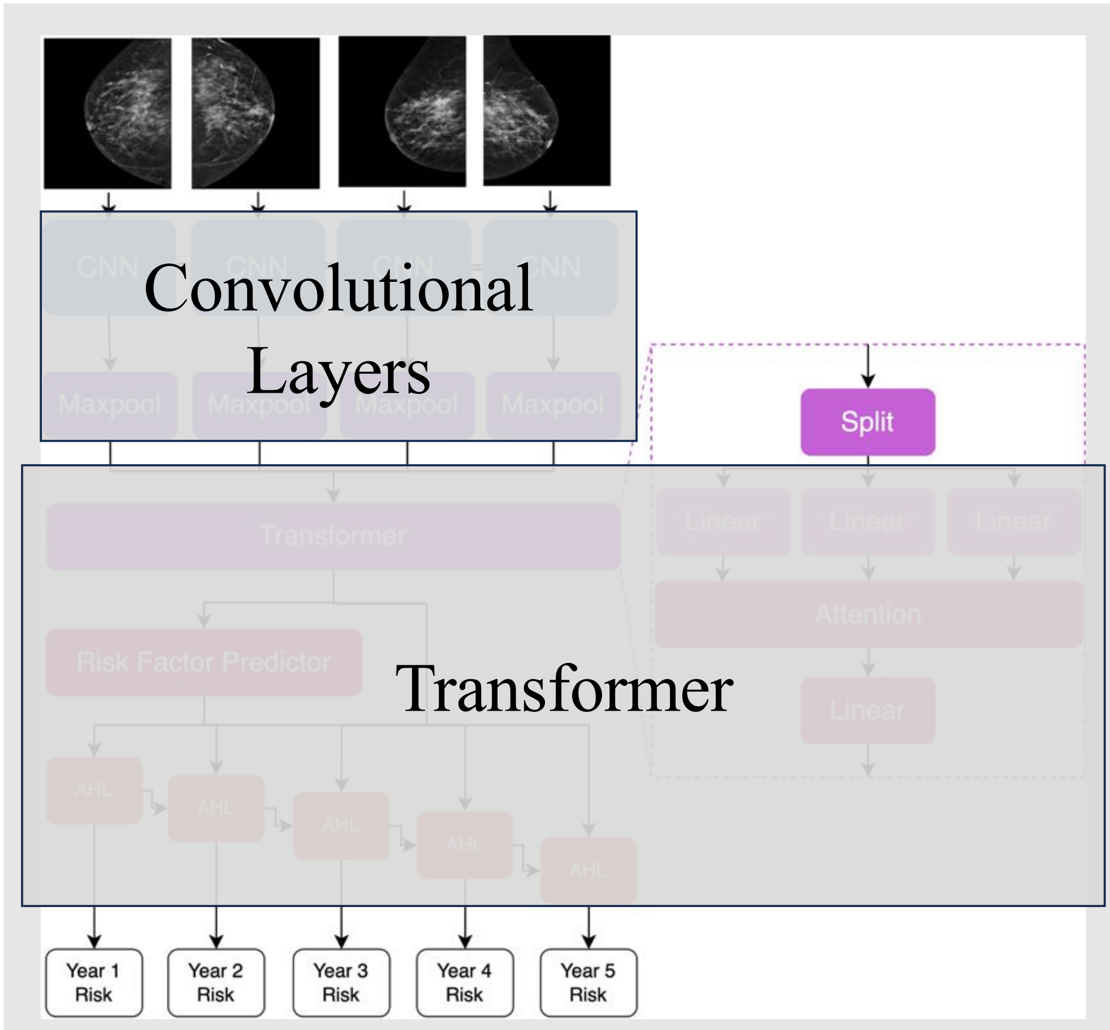
- Breast cancer is the most commonly diagnosed cancer worldwide.
- Screening frequency at least once per year for women over 40.

- 20 FDA approved AI tools, but nothing approved for 5-year risk prediction from mammograms.

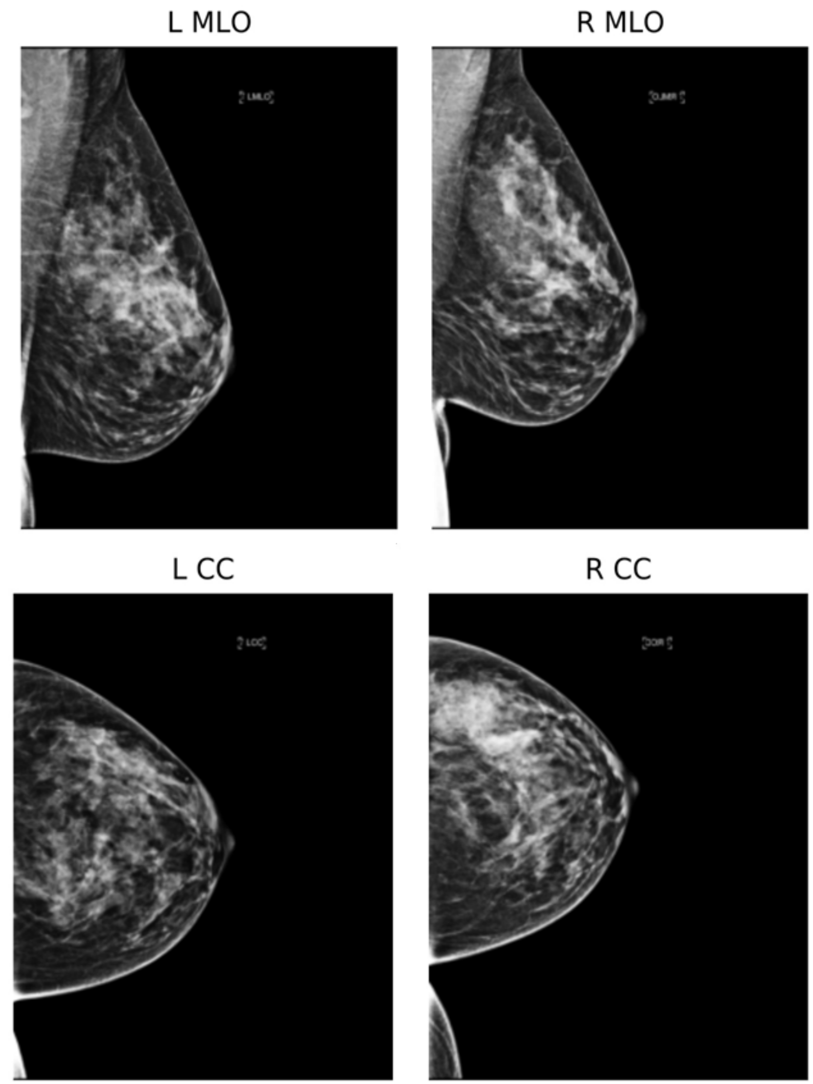
Mirai Architecture (from Regina's paper, Yala et al 2021)



Mirai Architecture (from Regina's paper, Yala et al 2021)



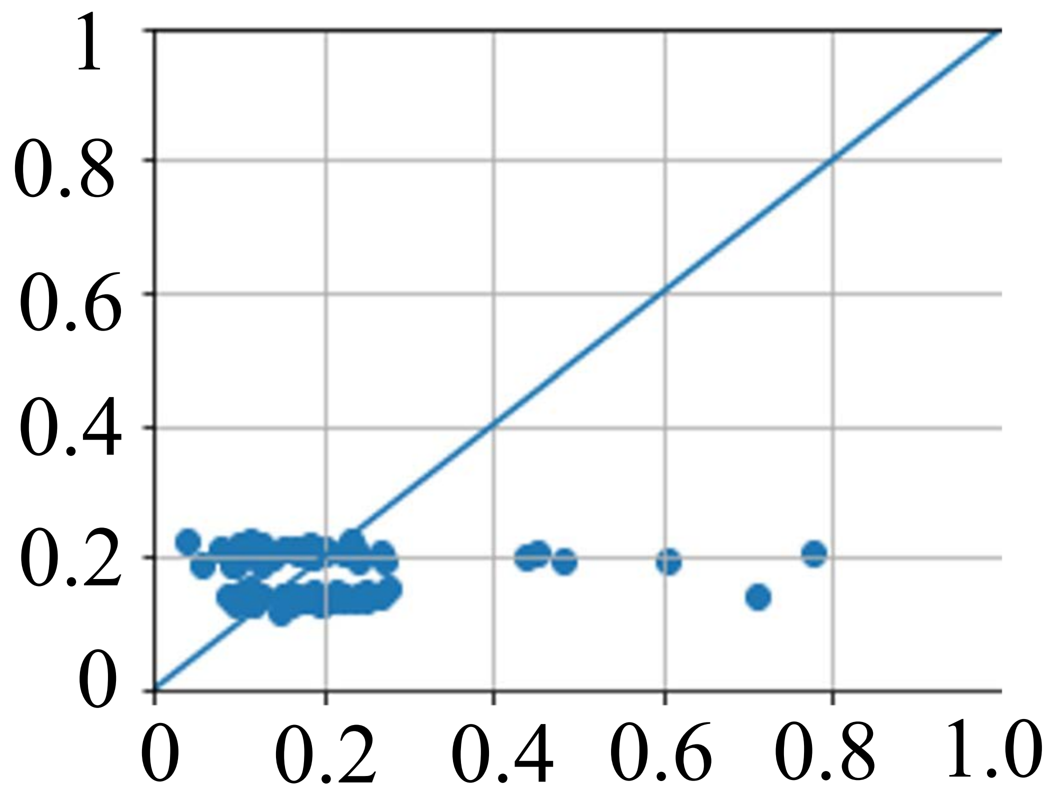
4 views, 2 of each side.



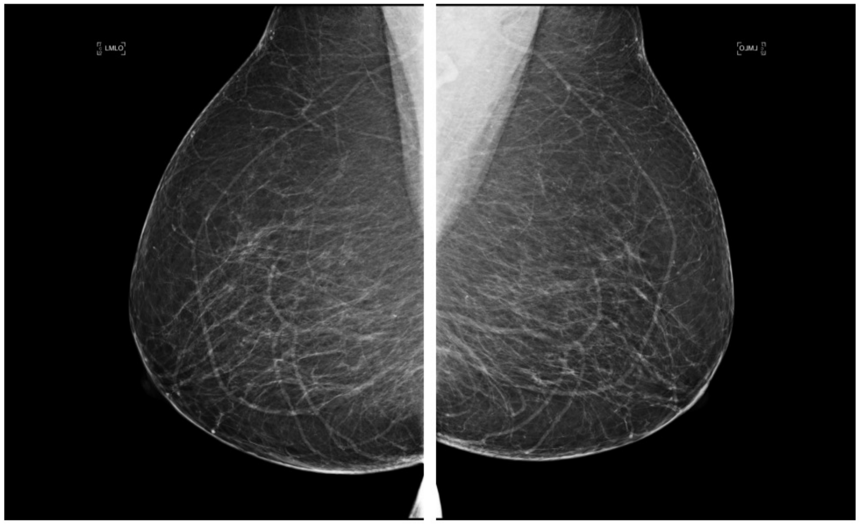
Mirai Predicts Low Risk on Mirrored Mammograms

(INBreast dataset, n=120)

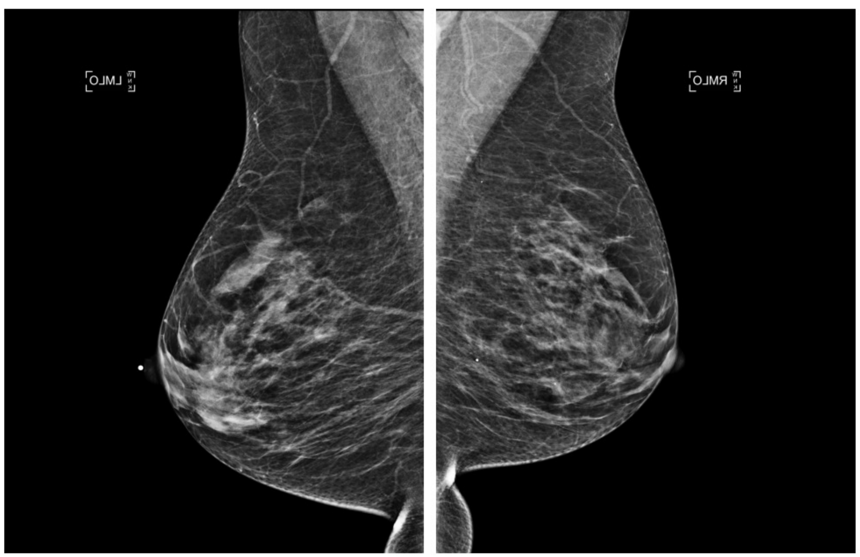
Mirai risk score from mirrored mammograms, left mirrored to both sides



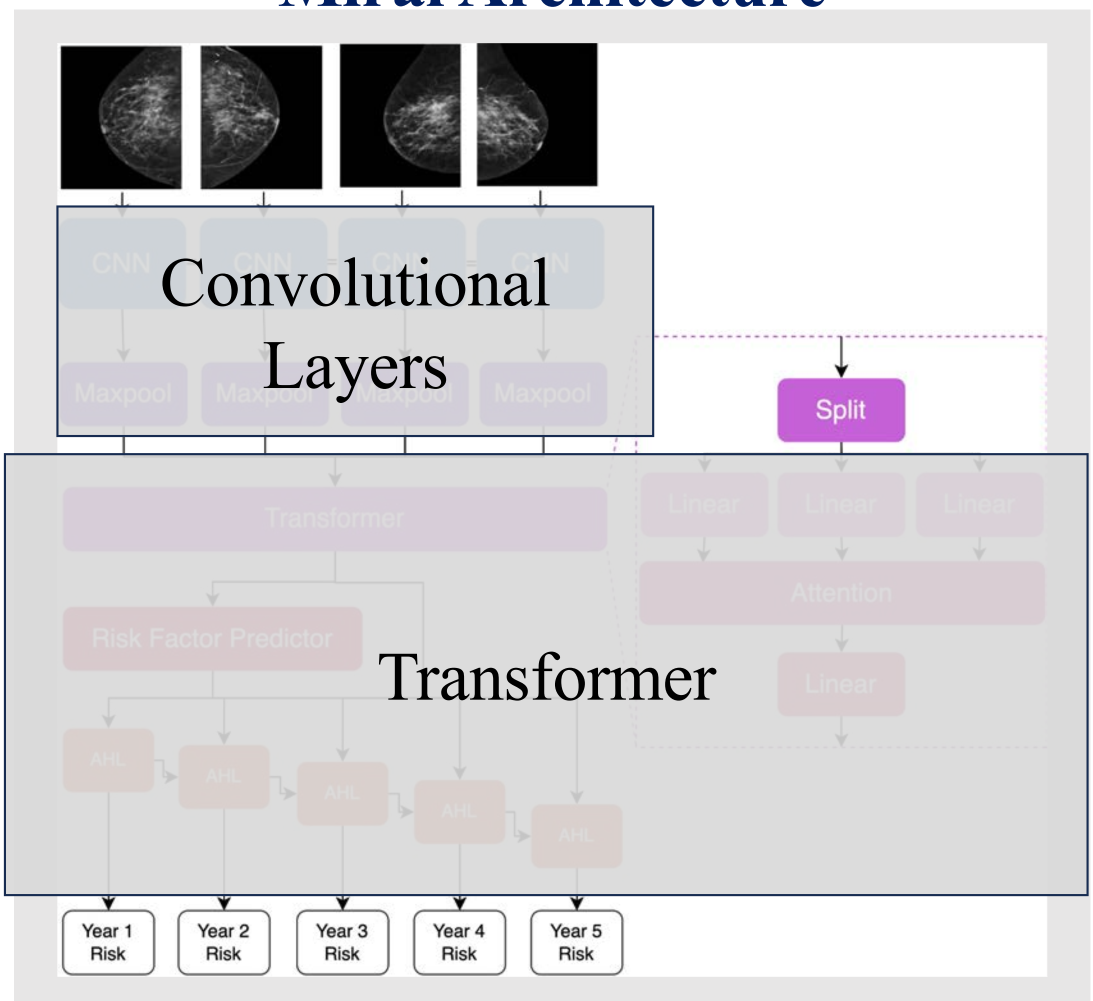
Mirai's original risk score



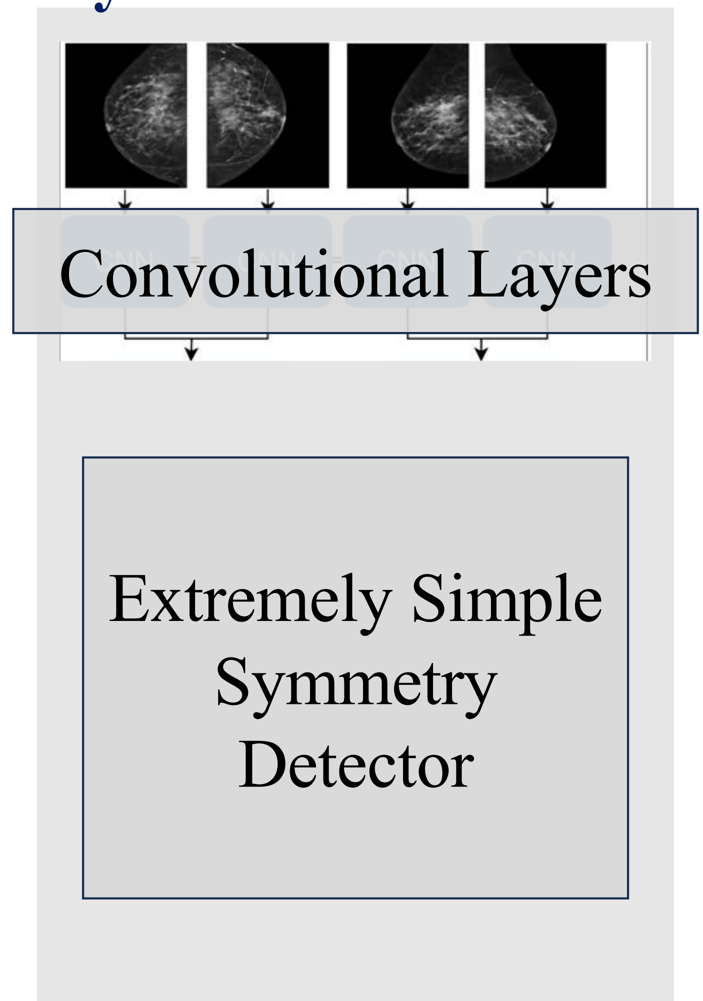

RMLO



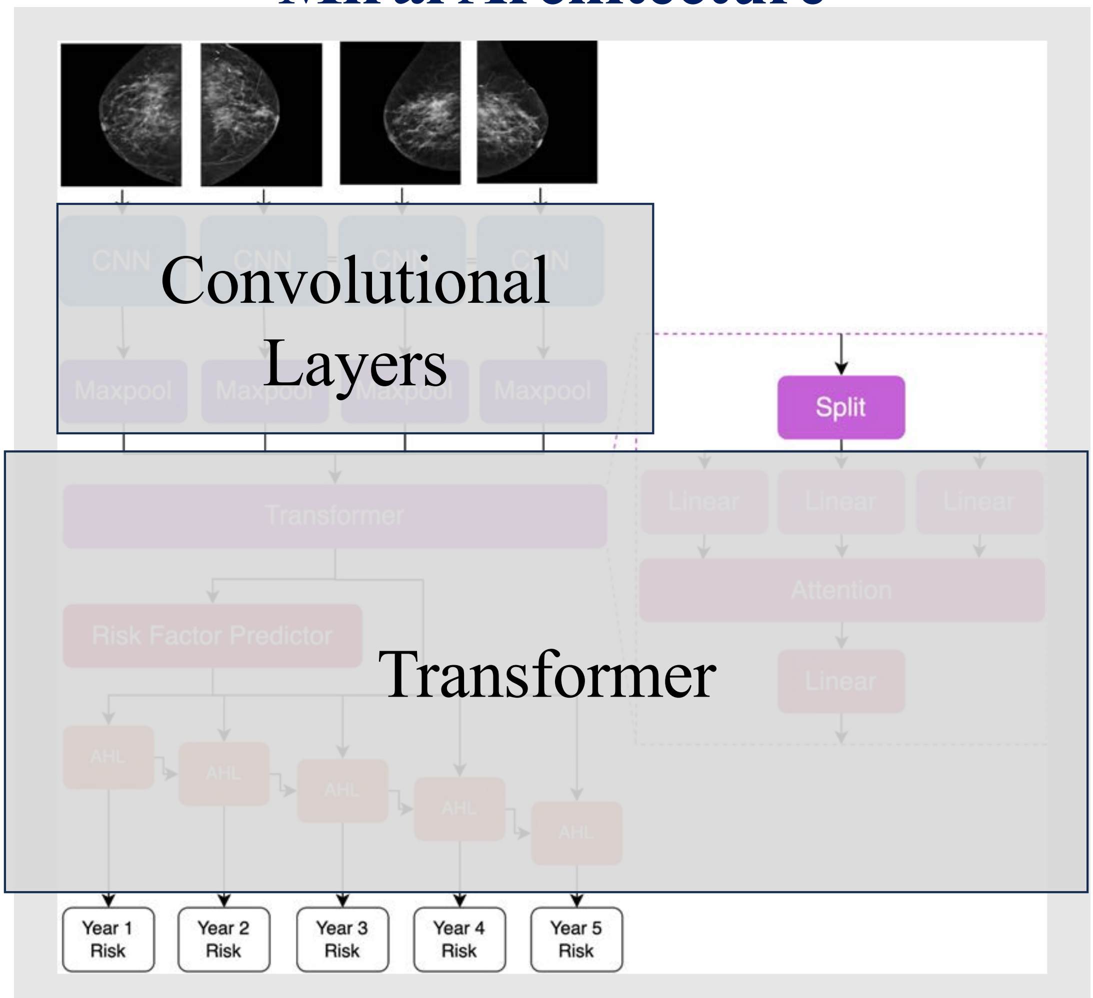
Mirai Architecture



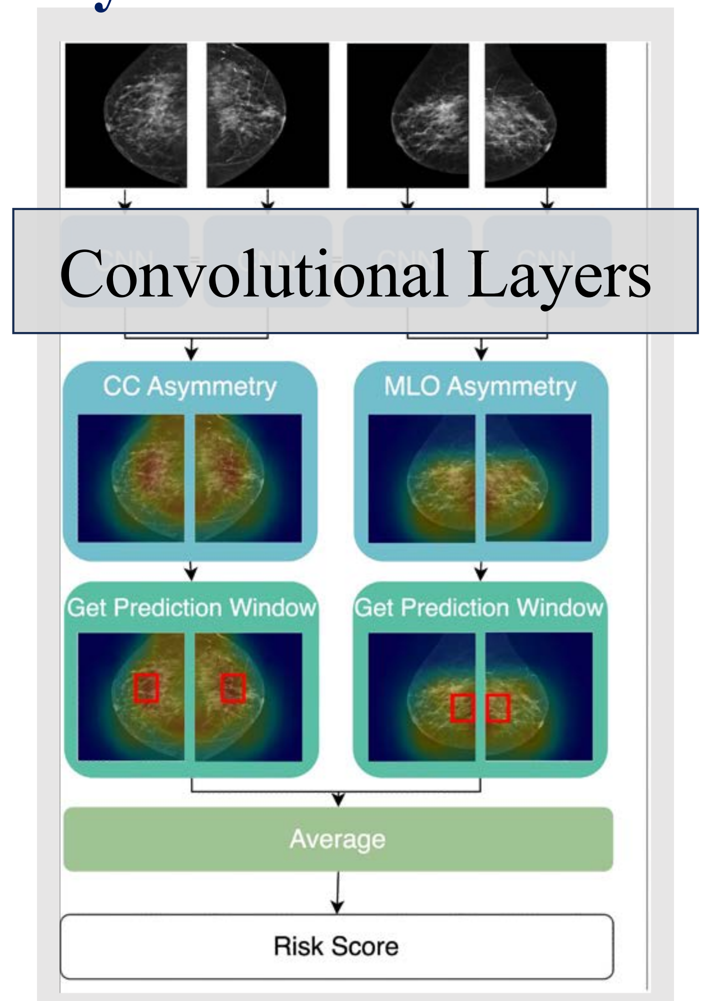
AsymMirai Architecture

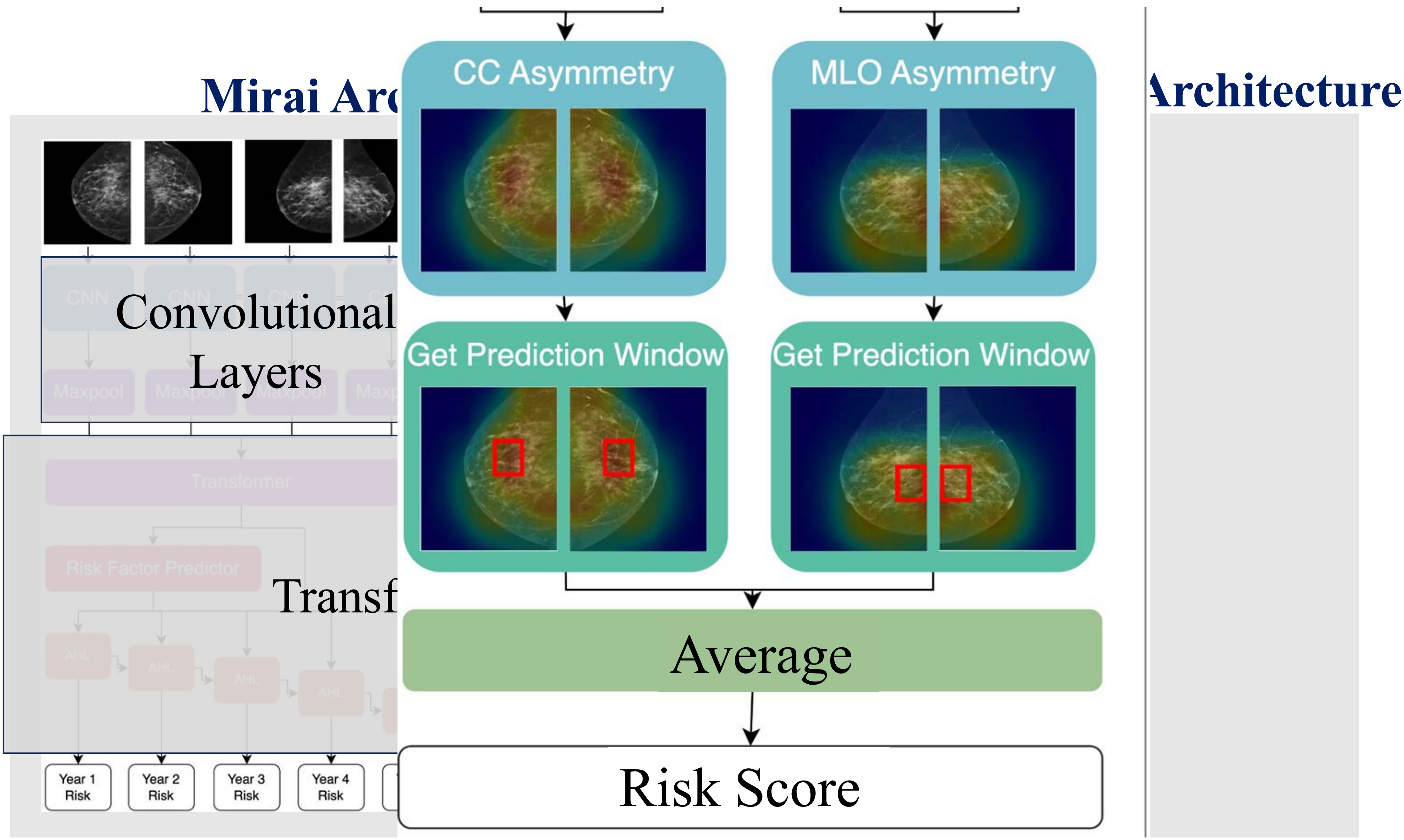


Mirai Architecture



AsymMirai Architecture

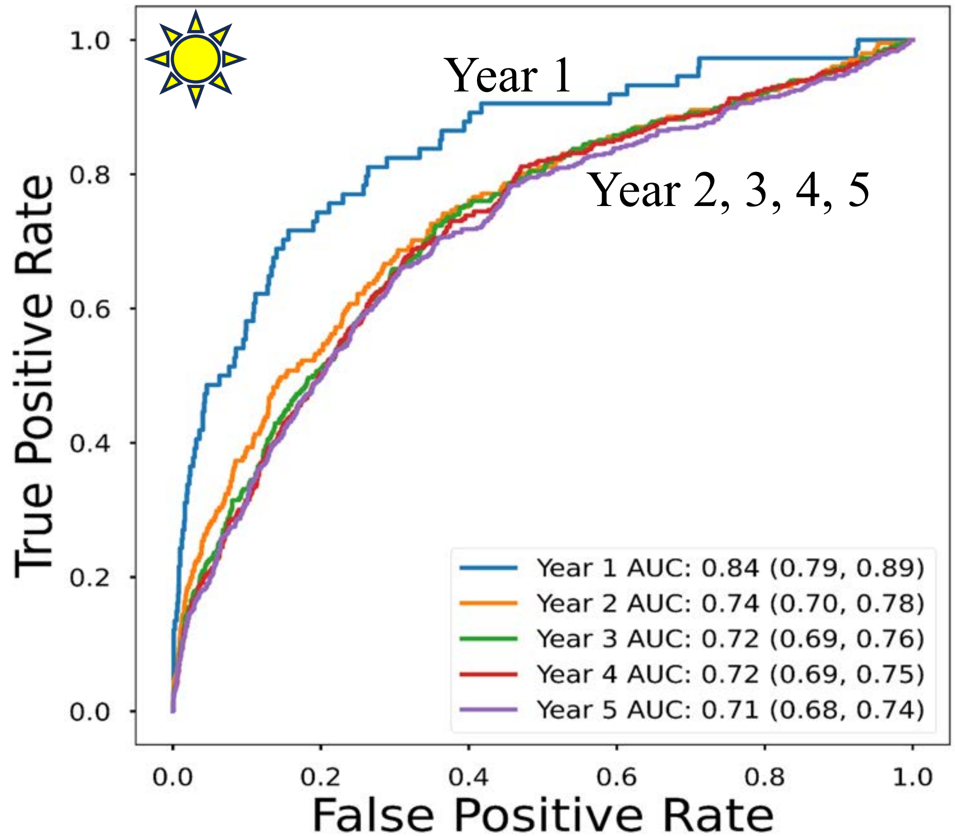




Mirai Arc
Architecture
CC Asymmetry
MLO Asymmetry
Get Prediction Window
Get Prediction Window
Average
Risk Score
CNN
CNN
Convolutional
Layers
Maxpool
Maxpool
Maxpool
Transformer
Risk Factor Predictor
Transf
Year 1
Risk
Year 2
Risk
Year 3
Risk
Year 4
Risk

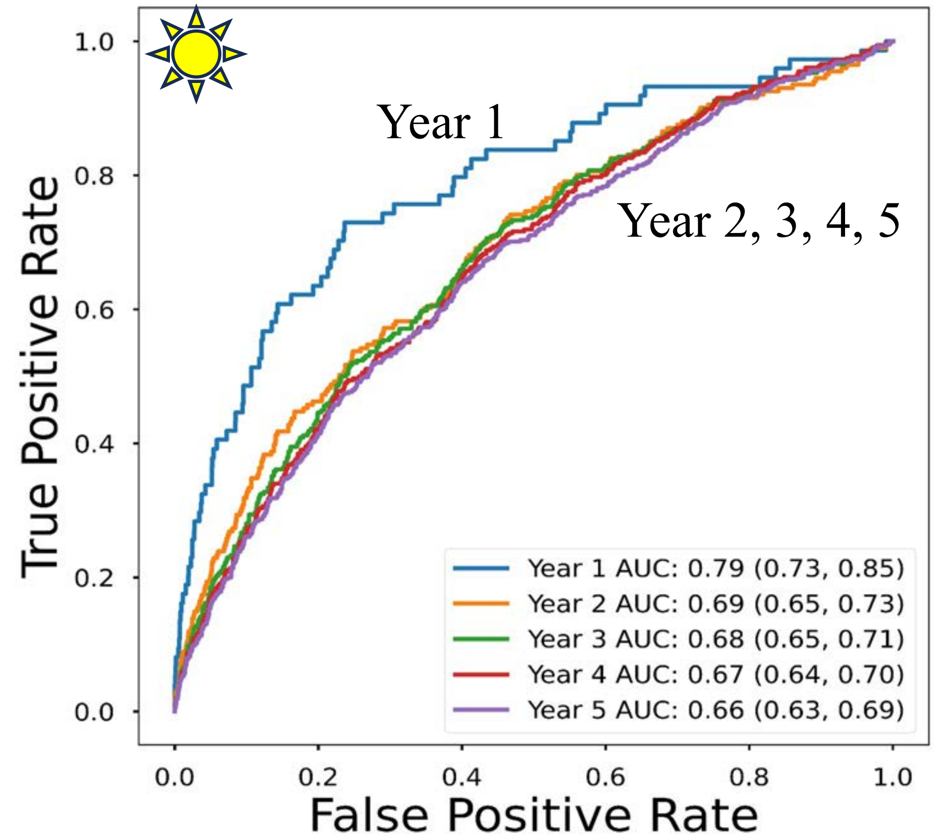
Mirai ROC

EMBED Validation Screening Exams

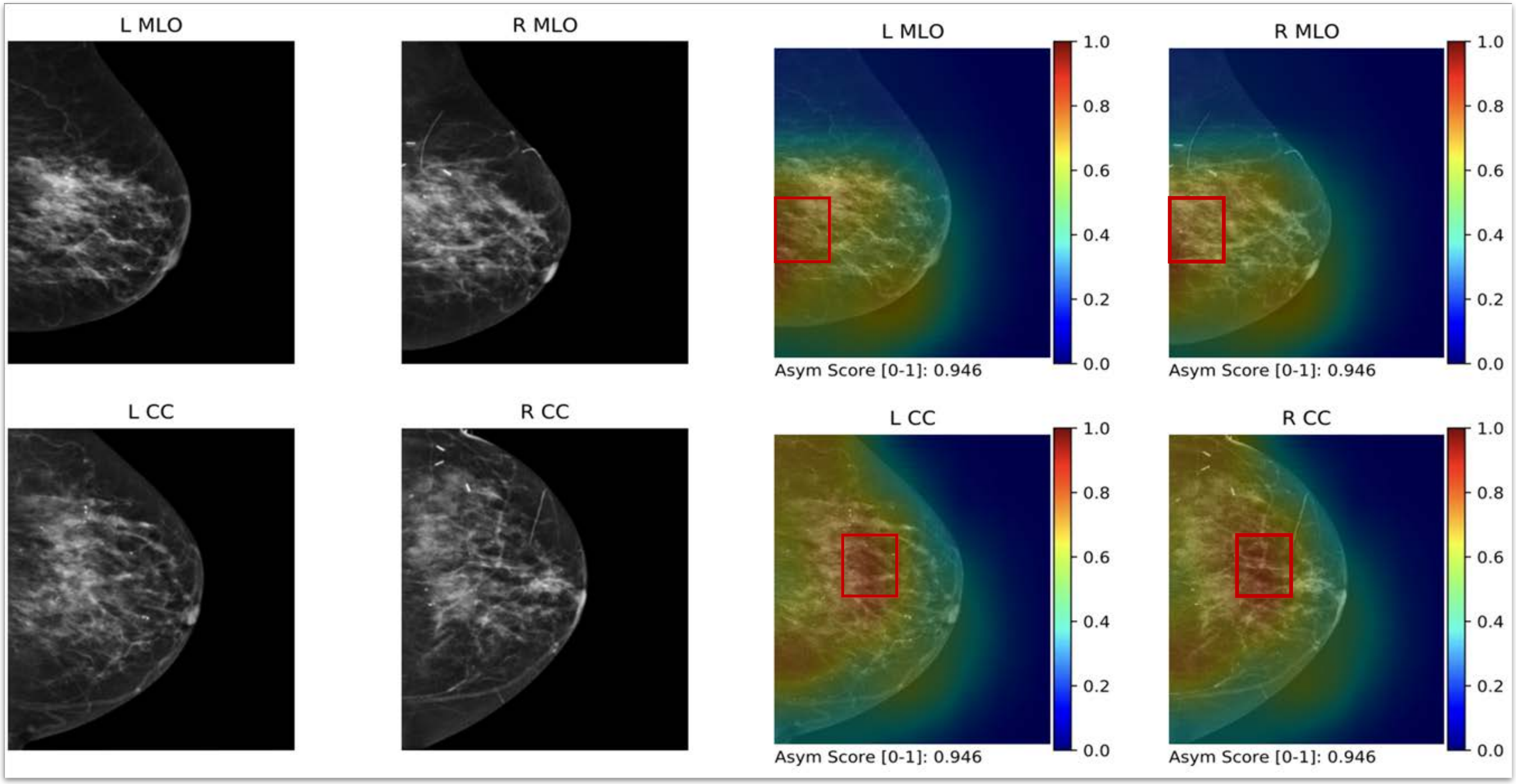


AsymMirai ROC

EMBED Validation Screening Exams

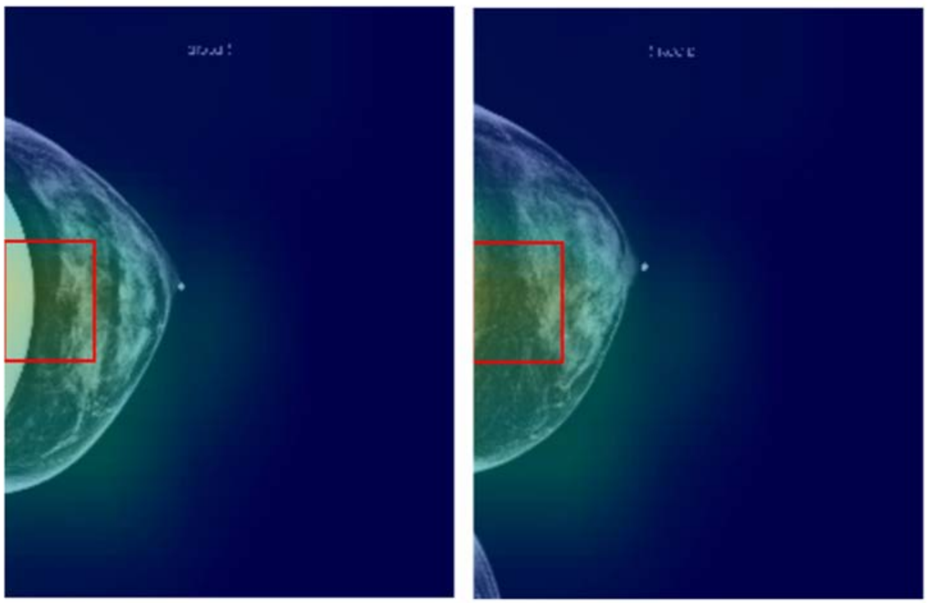


AsymMirai's Interpretable Outputs



Confounding Analysis with AsymMirai's Interpretability

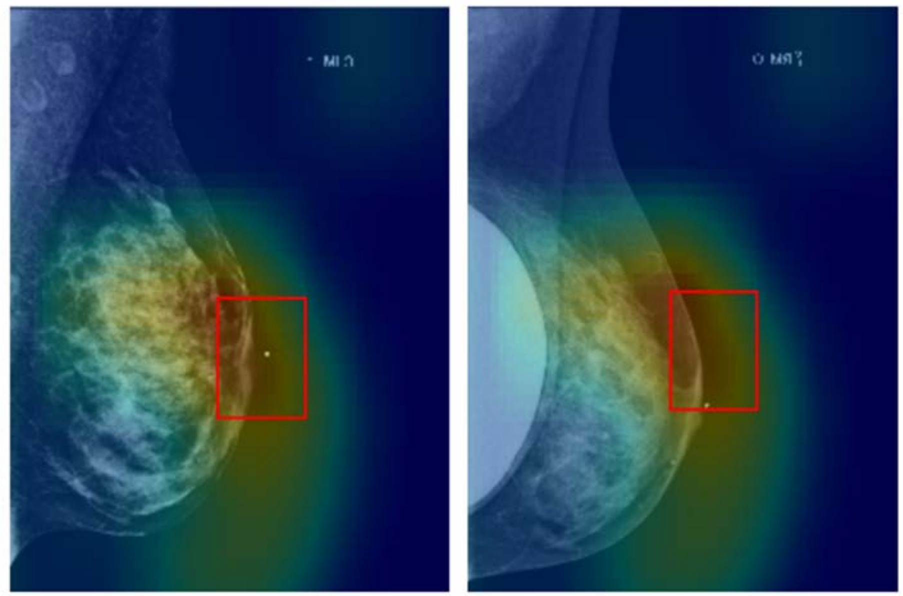
Moderate Risk Prediction
Confounded by Implant



Left CC   Right CC

Patient **does not** develop
cancer within 5 years

High Risk Prediction **not**
Confounded by Implant

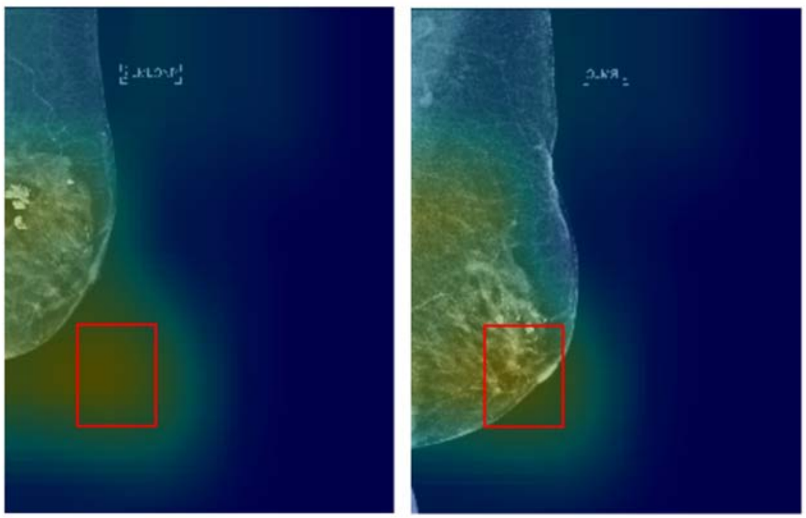


Left MLO   Right MLO

Patient **does** develop
cancer within 5 years

Confounding Analysis with AsymMirai's Interpretability

Moderate Risk Prediction
Confounded by Misaligned Image

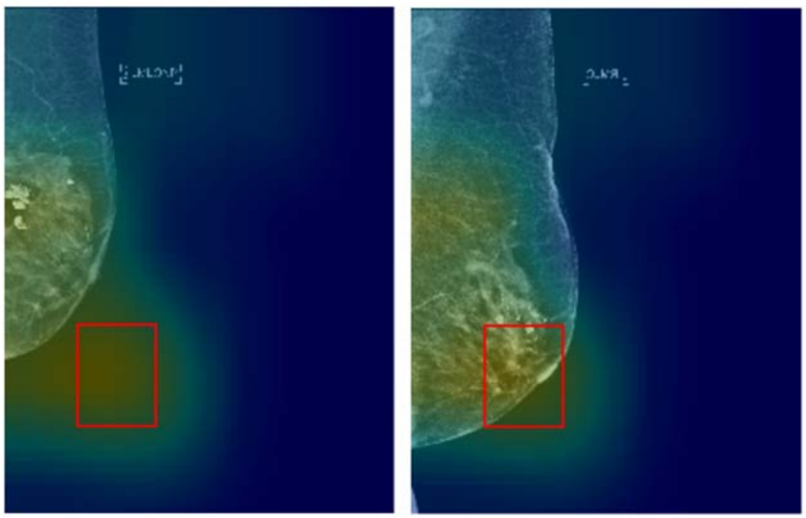


Left MLO | Right MLO

Patient **does not** develop
cancer within 5 years

Confounding Analysis with AsymMirai's Interpretability

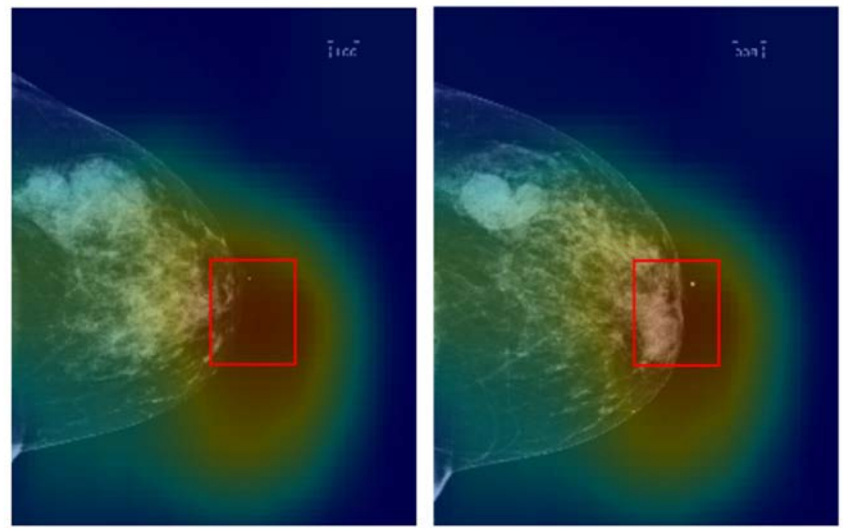
Moderate Risk Prediction Confounded by Misaligned Image



Left MLO   Right MLO

Patient **does not** develop cancer within 5 years

High Risk Prediction **not** Confounded by Pronounced Lymph Nodes



Left CC   Right CC

Patient **does** develop cancer within 5 years

Challenge: Make a discovery with interpretable AI that wasn’t possible with a black box.

Predict breast cancer up to 5 years in advance?

Yes, and we made a discovery.

Radiology

2024

AsymMirai: Interpretable Mammography-Based Deep Learning Model for 1- to 5-year Breast Cancer Risk Prediction

Jon Donnelly, Luke Moffett, Alina Barnett, Hari Trivedi, Fides Regina Schwartz, Joseph Lo, Cynthia Rudin

Jon Donnelly

Luke Moffett

Alina Barnett

Fides Regina Schwartz (Radiologist)

Joseph Lo (Professor of Radiology)

Amazing Things Come From Having Many Good Models

Cynthia Rudin [1 *] Chudi Zhong [1] Lesia Semenova [1] Margo Seltzer [2] Ronald Parr [1] Jiachang Liu [1]
Srikar Katta [1] Jon Donnelly [1] Harry Chen [1] Zachery Boner [1]

We address how the Rashomon Effect impacts:

ICML 2024 spotlight

(1) the existence of simple-yet-accurate models

(2) flexibility to address user preferences, such as fairness and monotonicity, without losing performance

(3) uncertainty in predictions, fairness, and explanations

(4) reliable variable importance

(5) algorithm choice, specifically, providing advanced knowledge of which algorithms might be suitable for a given problem

(6) public policy

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

My biostatistician friends tell me, “Doctors can interpret logistic regression.” There is no way they can interpret a black box containing fifty trees hooked together. In a choice between accuracy and interpretability, they'll go for interpretability.

Policy Implications

- Interpretable models by default in high-stakes settings
- Put interpretable ML into AI education

Amazing Things Come From Having Many Good Models

Cynthia Rudin [1*] Chudi Zhong [1] Lesia Semenova [1] Margo Seltzer [2] Ronald Parr [1] Jiachang Liu [1]
Srikar Katta [1] Jon Donnelly [1] Harry Chen [1] Zachery Boner [1]

ICML 2024 spotlight

We address how the Rashomon Effect impacts:
(1) the existence of simple-yet-accurate models
(2) flexibility to address user preferences, such as fairness and monotonicity, without losing performance
(3) uncertainty in predictions, fairness, and explanations
(4) reliable variable importance
(5) algorithm choice, specifically, providing advanced knowledge of which algorithms might be suitable for a given problem
(6) public policy

Thanks