Trustworthy AI (Tai): An Open-Source Framework for Interpretable Machine Learning for Risk Factor Datasets

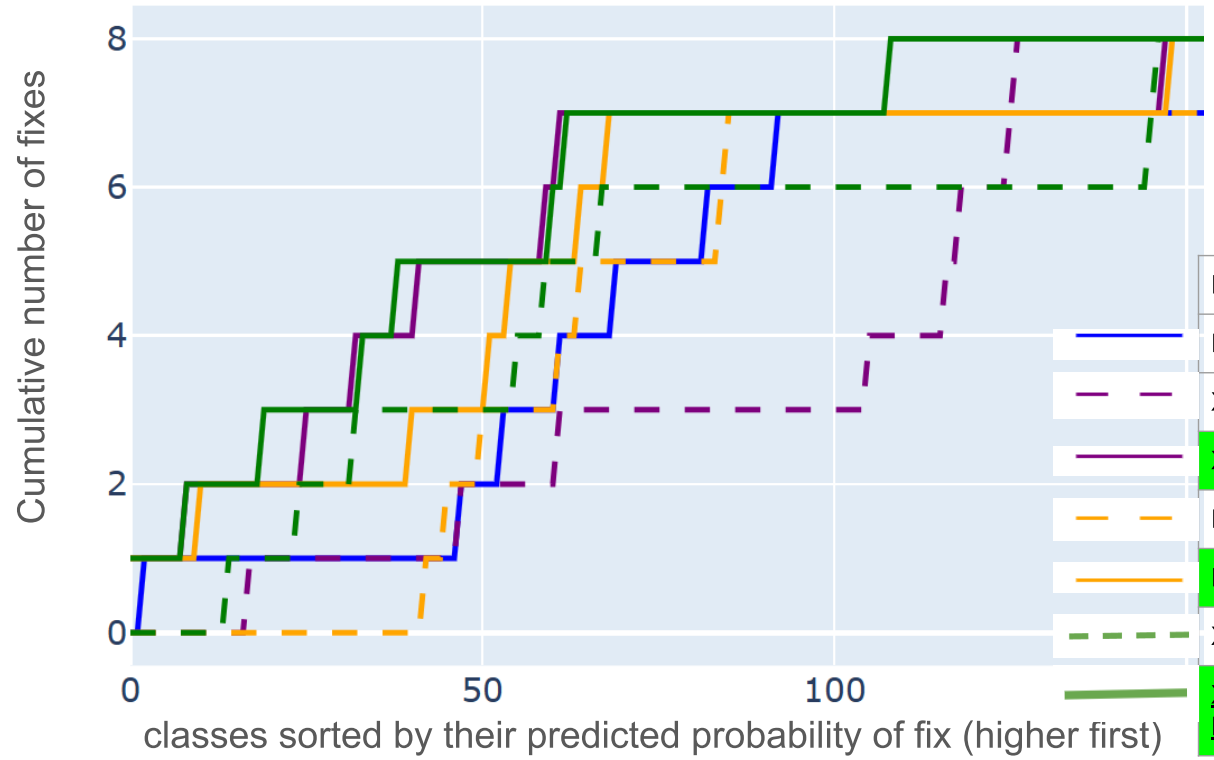
Ofir Pele, Ph.D.

Liat Pele, Ph.D.

Why Interpretable?

- More accurate, especially with small risk factor data
- Identify target leakage
- Test data differs from training data, e.g., extrapolation scenarios
- Insights, especially actionable ones
- Encourage adoption
- Meet regulatory demands

Model Performance - Code Health

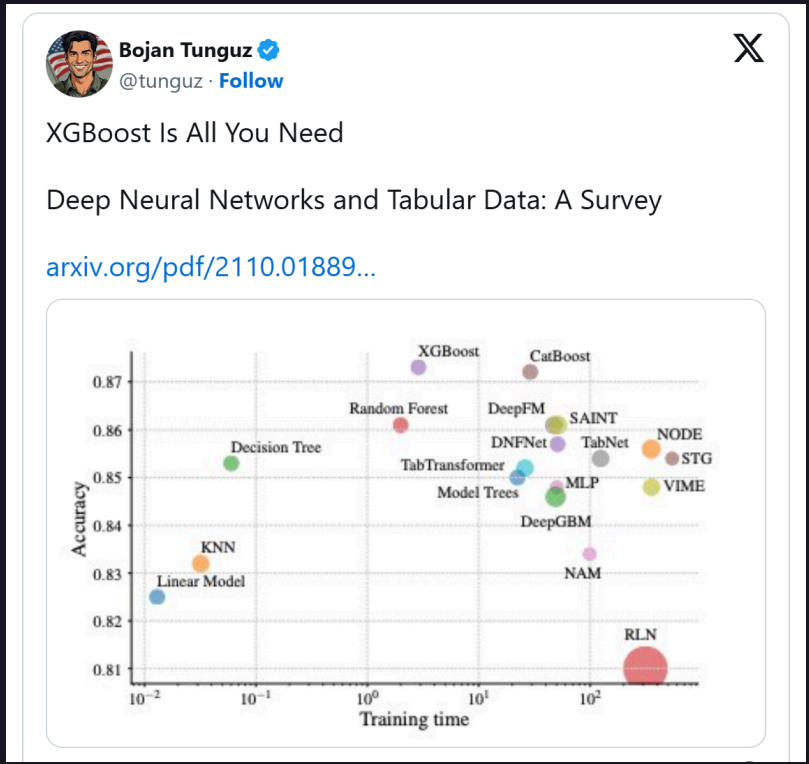


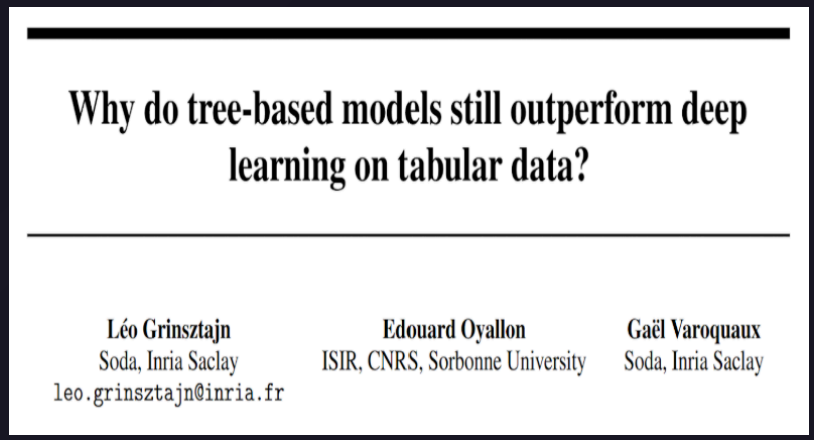
Model	Monotone	Test AUC
Linear	V	55.
XGB	X	42
XGB +Monotone	V	71
Lattice	X	52
Lattice +Monotone	V	65
XGB+Lattice	X	50
XGB+Lattice + Monotone	V	73

Visualization

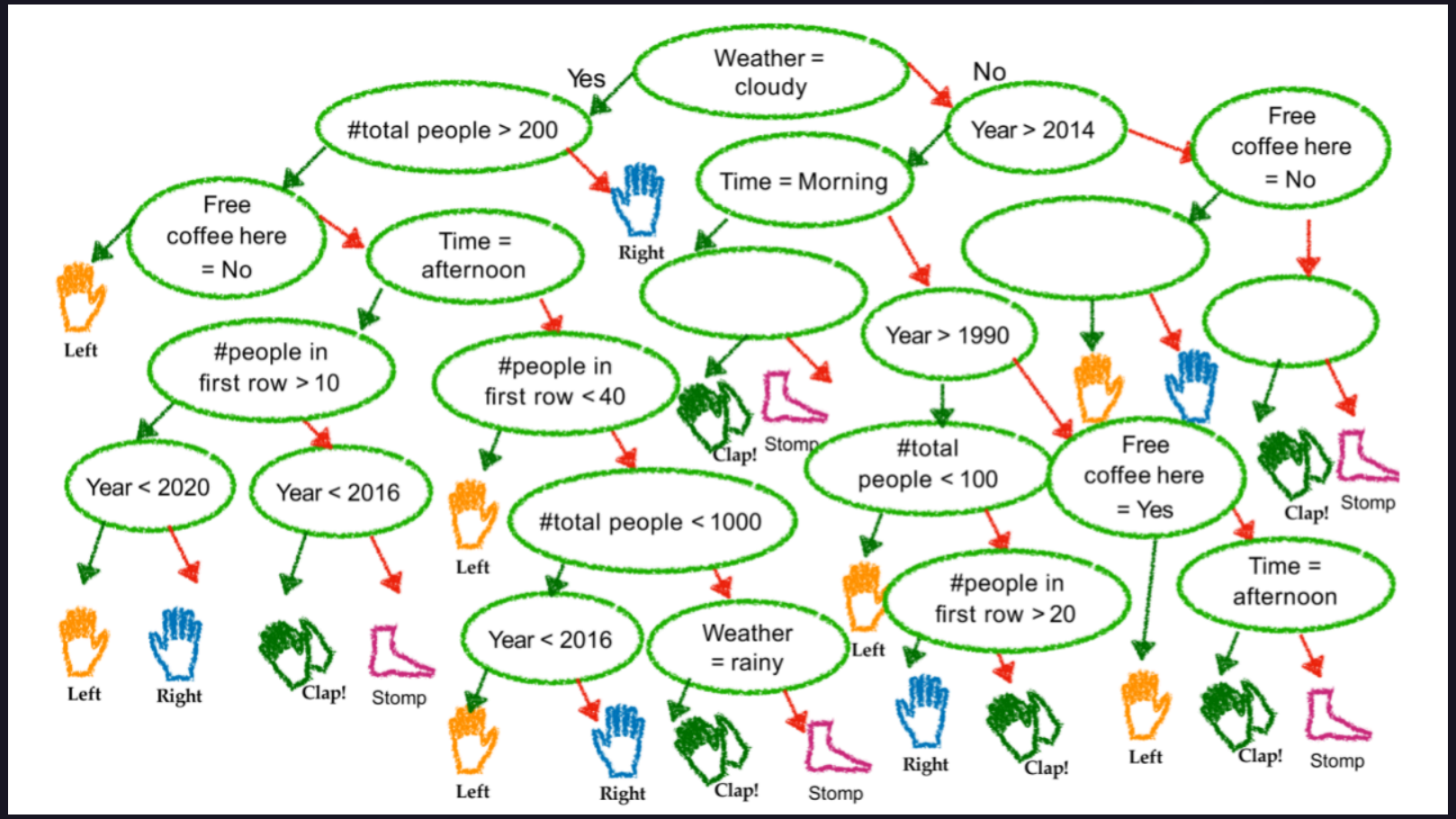
Visualization - Insights

ML Models for Tabular Data - Trees





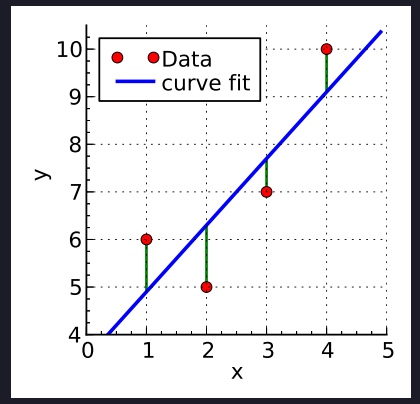
Trees are Not Interpretable & Need Lots of Data



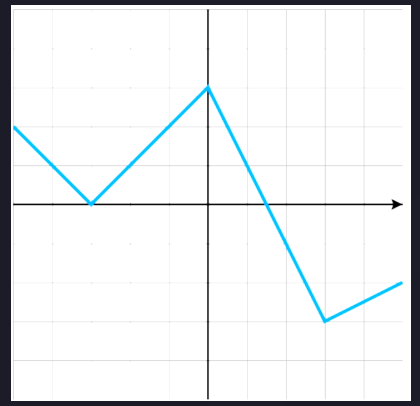
Taken from Been Kim's presentation

Interpretable Models

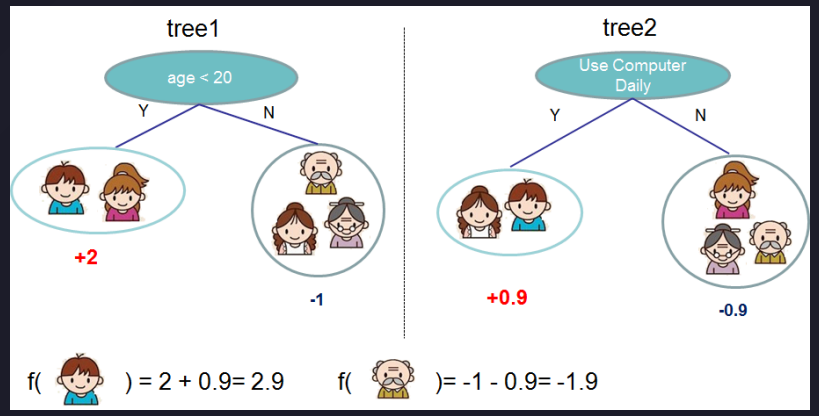
Linear



Piecewise



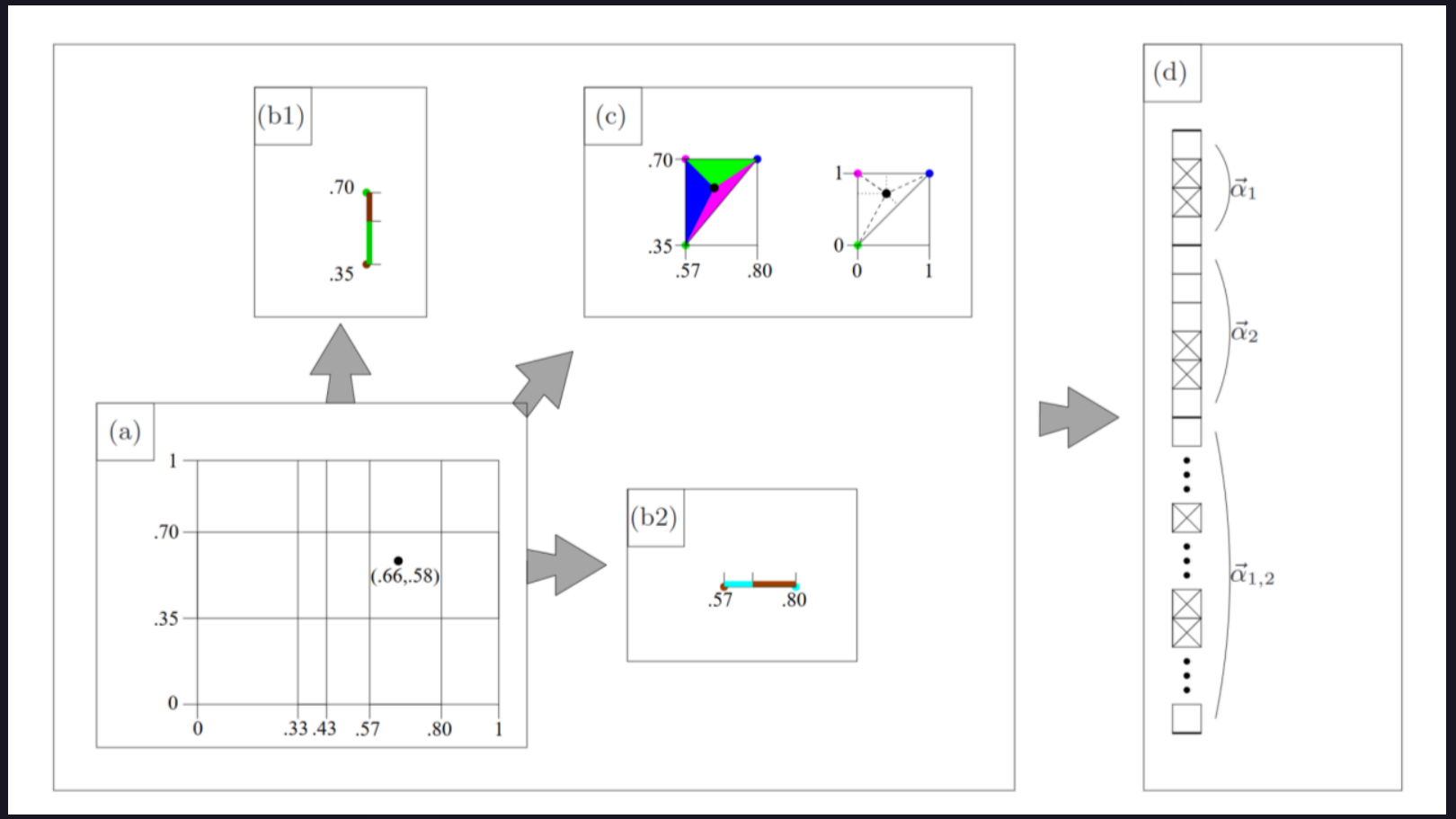
Small Trees



Tai - Trustworthy AI

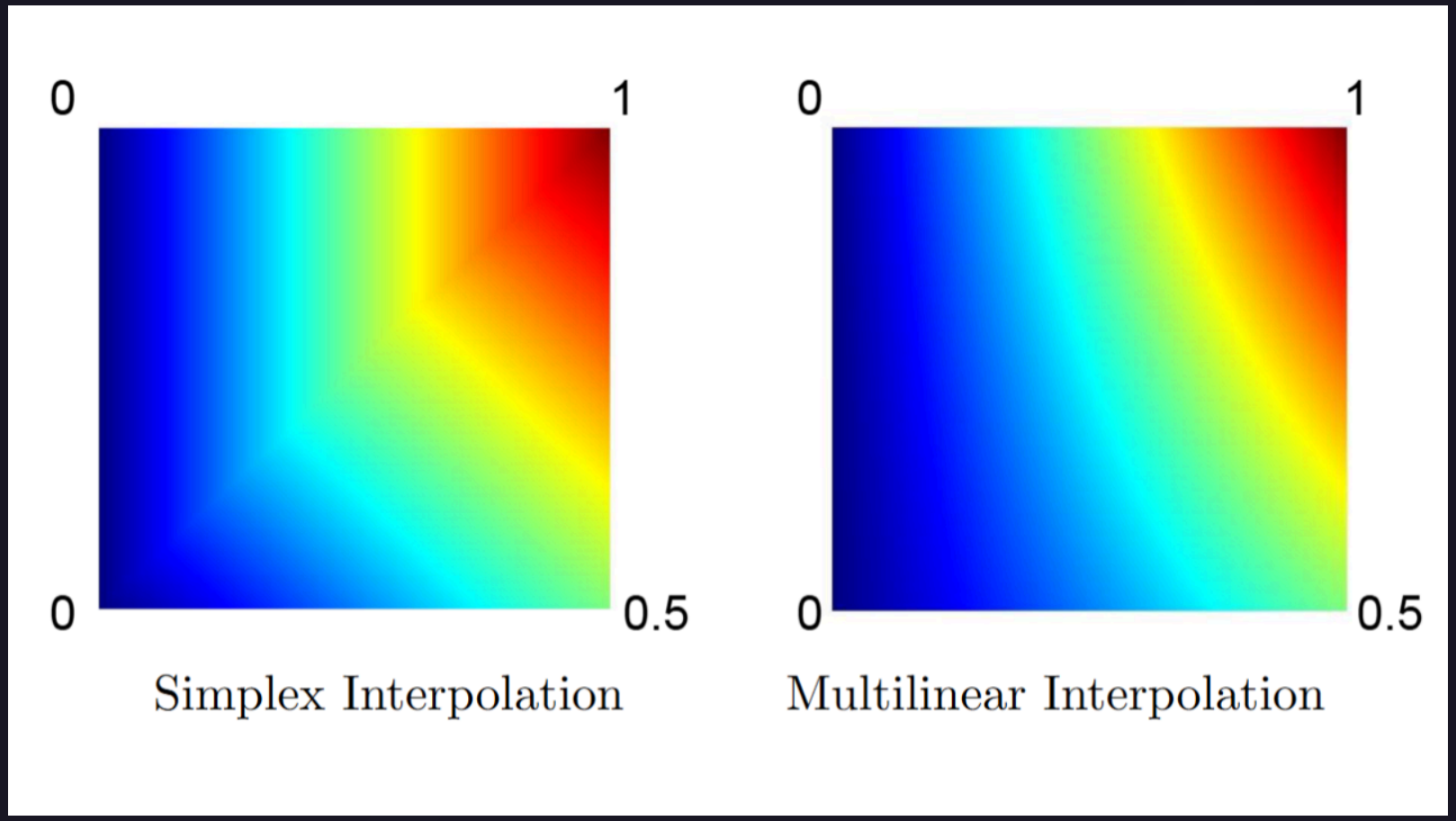
- A "monotone universal approximator"
  - Every feature either monotonically increases the risk or monotonically decreases the risk
- Instead of enforcing constraints, we learn them from data
  - Currently using a linear model
- We use models that enable monotonic constraints:
  - Trees and Lattices

Lattices



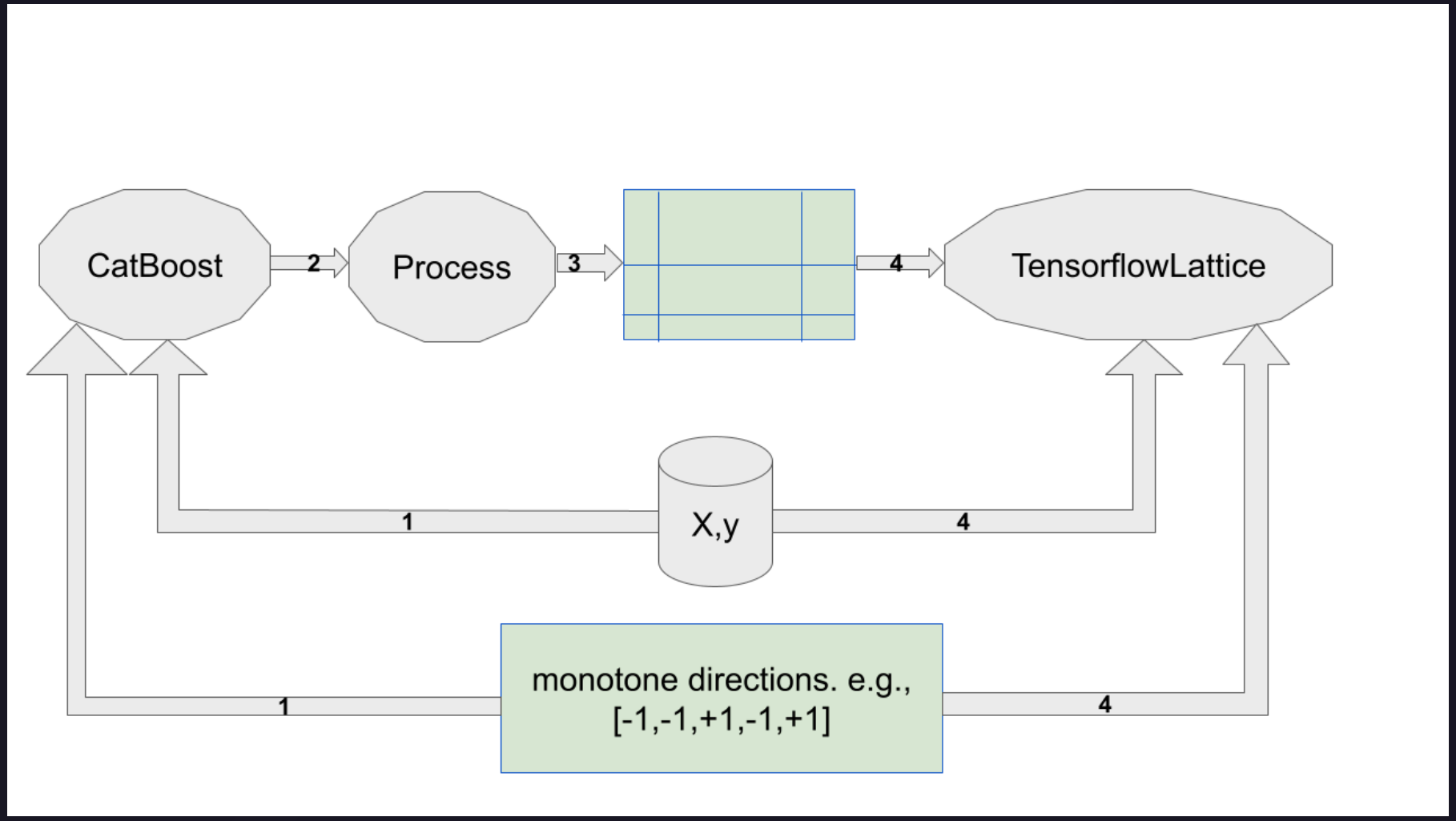
Taken from https://ofirpele.github.io/papers/ICML2013.pdf

Lattices

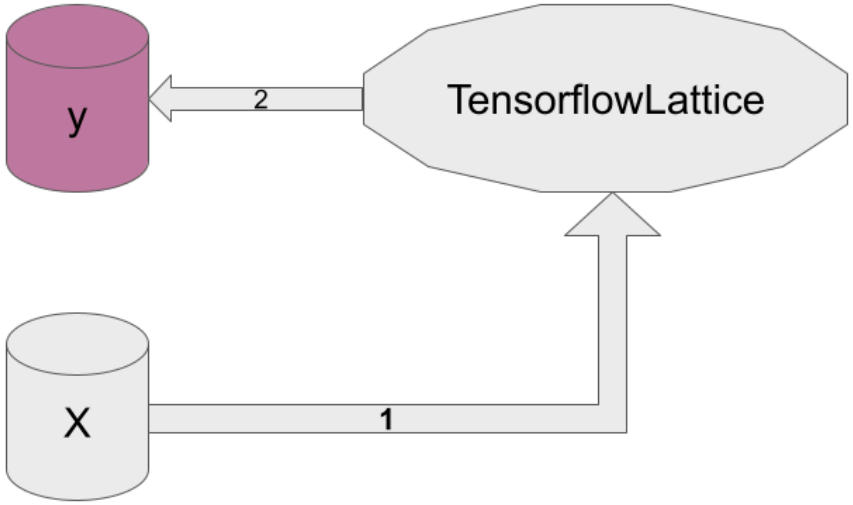


Taken from https://jmlr.org/papers/volume17/15-243/15-243.pdf

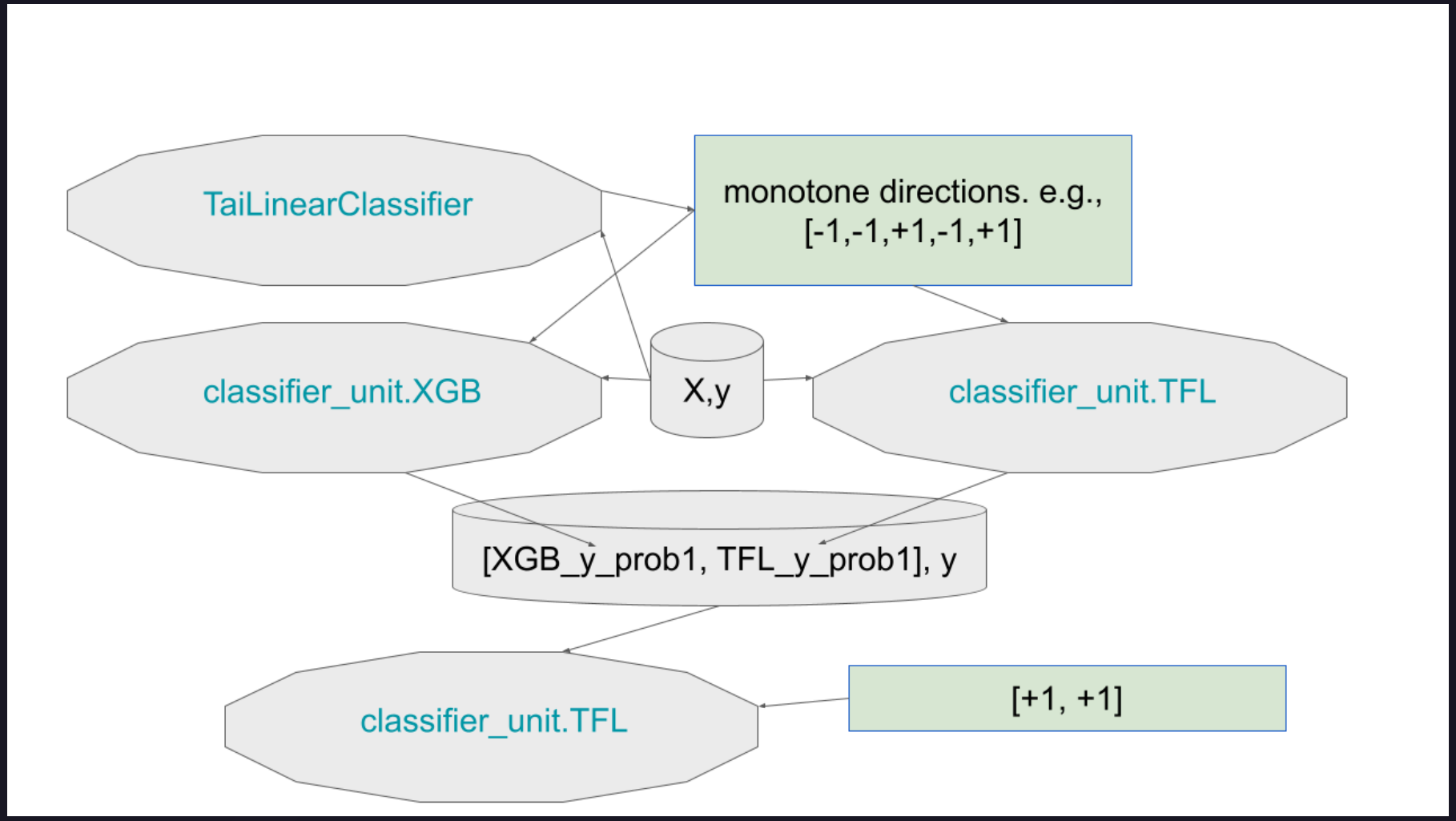
classifier_unit.TFL Training



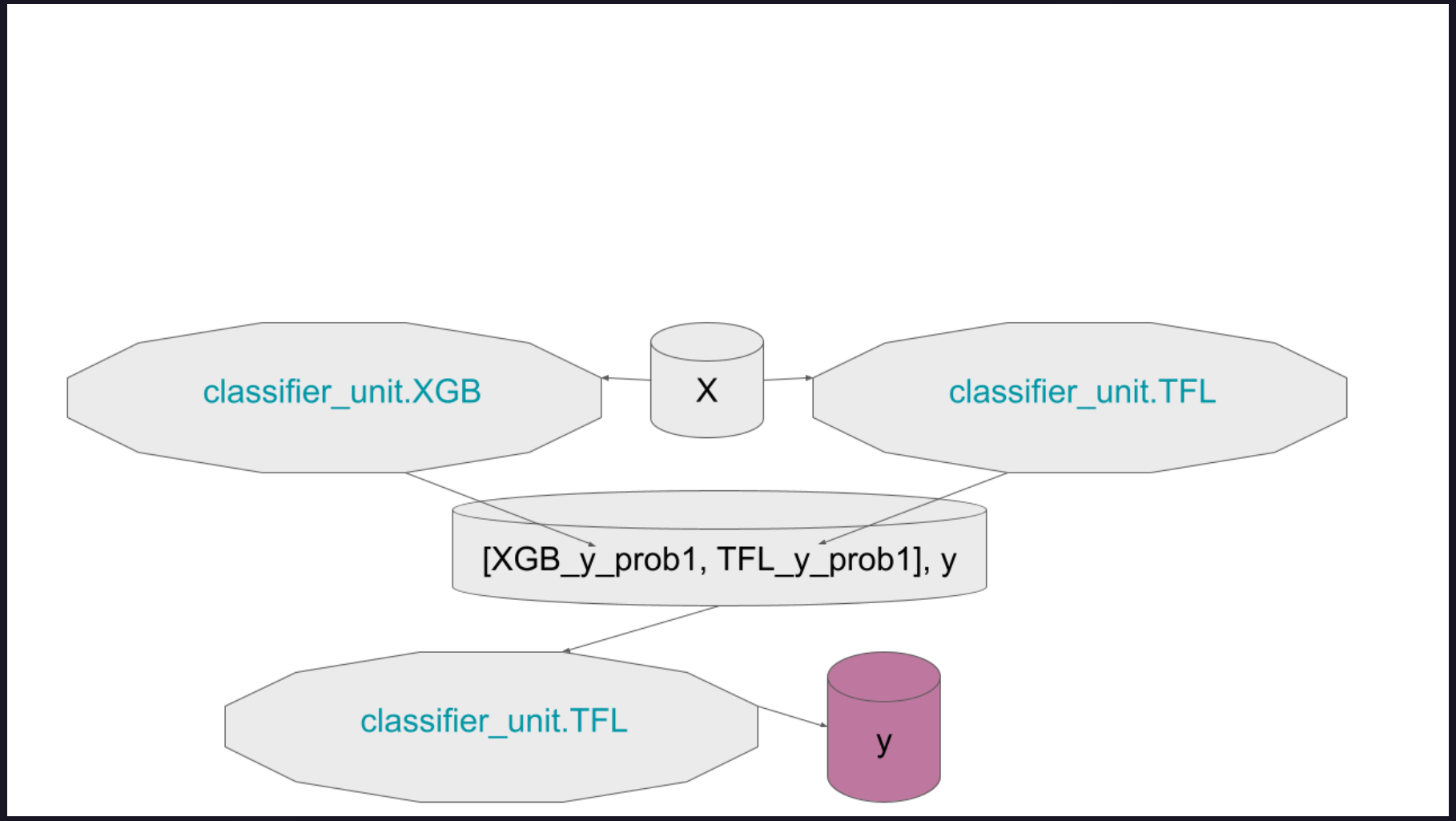
classifier_unit.TFL Inference



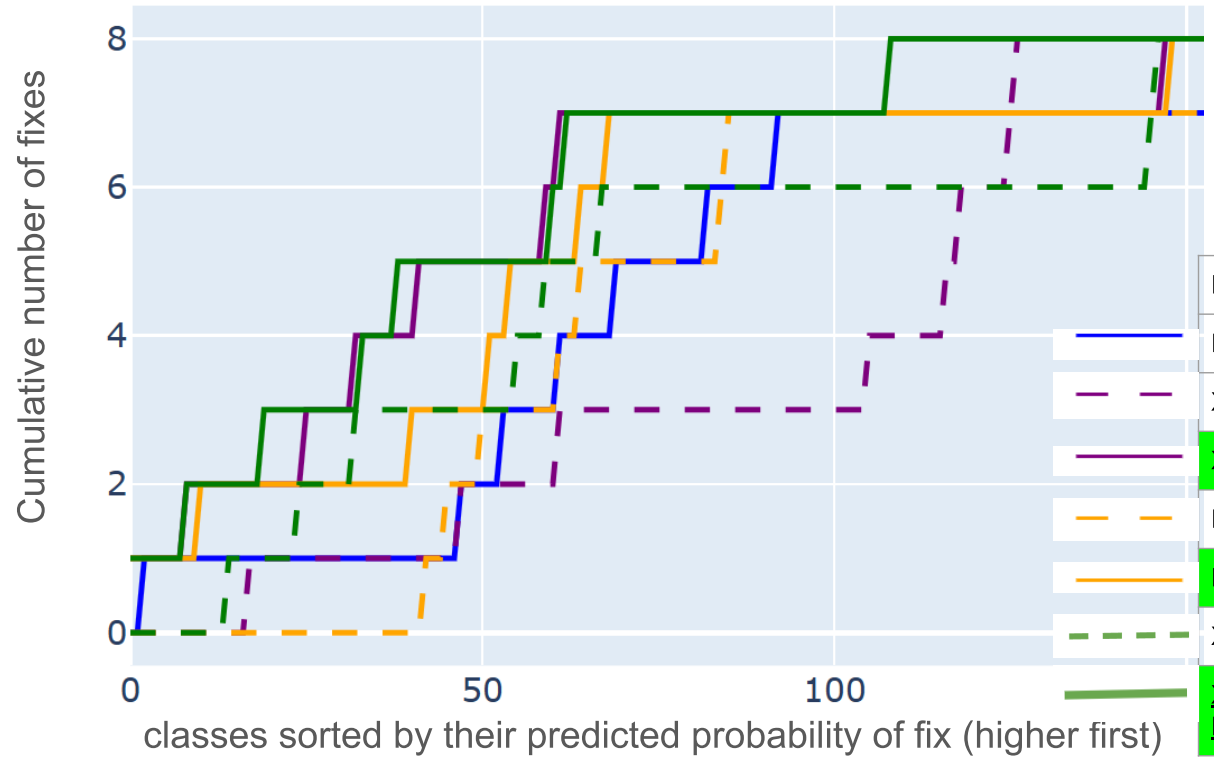
TaiClassifier Training



TaiClassifier Inference

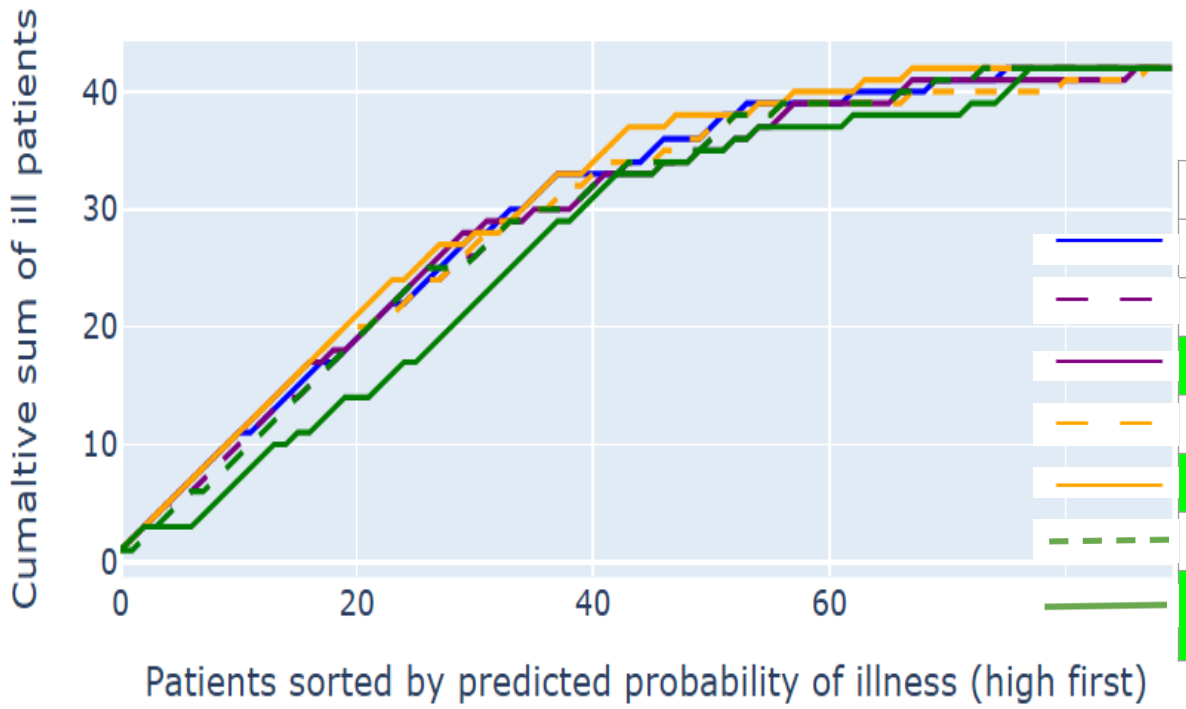


Model Performance - Code Health



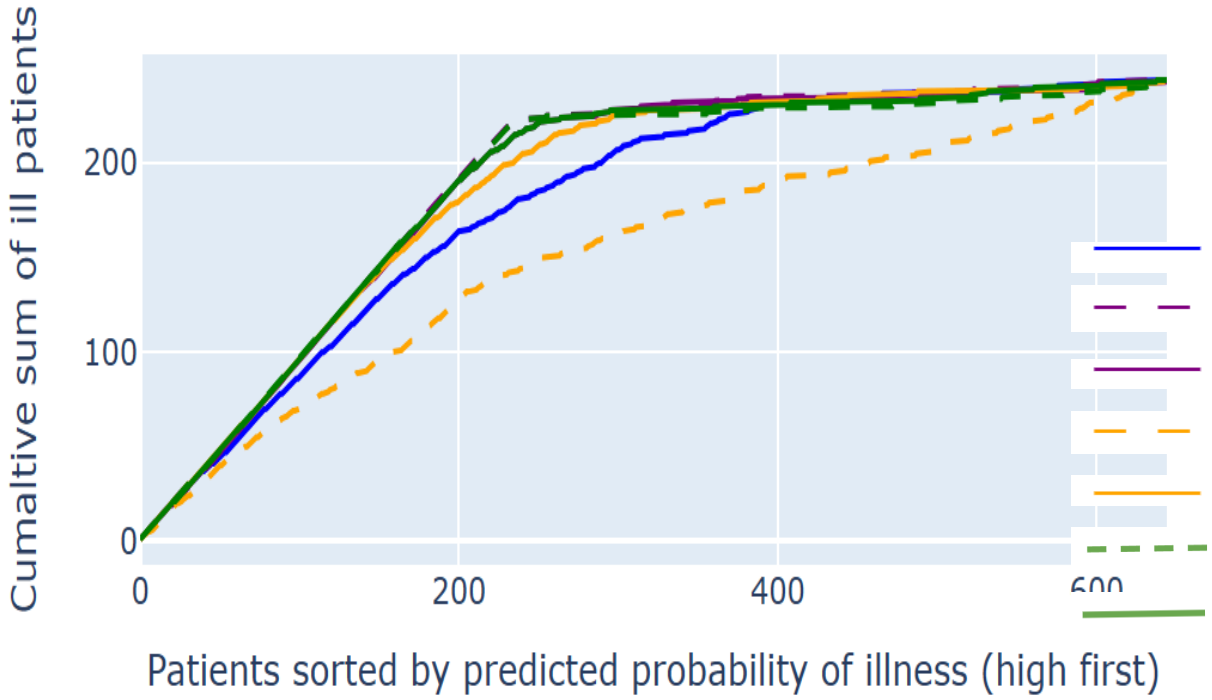
Model	Monotone	Test AUC
Linear	V	55.
XGB	X	42
XGB +Monotone	V	71
Lattice	X	52
Lattice +Monotone	V	65
XGB+Lattice	X	50
XGB+Lattice + Monotone	V	73

Model Performance - Heart Cleveland



Model	Monotone	Test AUC
Linear	V	89
XGB	X	87
XGB +Monotone	V	87
Lattice	X	87
Lattice +Monotone	V	92
XGB+Lattice	X	87
XGB+Lattice + Monotone	V	80

Model Performance - Alzheimer



Model	Monotone	Test AUC
Linear	V	80
XGB	X	95
XGB +Monotone	V	94
Lattice	X	70
Lattice +Monotone	V	93
XGB+Lattice	X	94
XGB+Lattice + Monotone	V	94

Future Directions

- More datasets
- More tasks: regression, RL, ...
- More shape constraints: conditional monotonicity, unimodal, ...
- Distill shape constraints from a Lattice&Tree learned model
- Report compatibility scores for learned shape constraints
- Research on the gap between Tai and non-constrained models
- Optimize code
- Use my implementation of lattices

Future Directions

- Learn Lattice&Tree model together
- Adaptive lattice for RL