AI for Prediction and Risk Modelling in Public Health

From predictive performance to better public health decisions

Gemma Postill | Population Health Analytics Lab | University of Toronto

AI Methods in Public Health | AI4PH | Public Health Agency of Canada (PHAC)

Population Health Analytics Lab





Professor
Laura Rosella

Focus of Today's Presentation

01 Introduction to prediction and risk modelling in public health

02 Statistics versus machine learning for prediction

03 Building & evaluating a machine learning prediction model

04 Responsible use of prediction in public health

1. Introduction to Prediction and Risk Modelling

What are we trying to predict, and why does it matter?

Types of epidemiology questions

DESCRIPTIVE

What does the population with diabetes look like?

Ex. What is the prevalence of diabetes in our community?

PREDICTION

Who is likely to experience the outcome?

Ex. What is the 10-year risk of type 2 diabetes?

CAUSATION

What would happen if we intervene?

Ex. Would changing the food environment reduce diabetes incidence?

Descriptive, prediction, and causation ask different questions

Types of epidemiology questions - 2

DESCRIPTIVE

What does the population with diabetes look like?

Ex. What is the prevalence of diabetes in our community?

PREDICTION

Who is likely to experience the outcome?

Ex. What is the 10-year risk of type 2 diabetes?

CAUSATION

What would happen if we intervene?

Ex. Would changing the food environment reduce diabetes incidence?

Descriptive, prediction, and causation ask different questions; today we are focusing on prediction

What public health questions can prediction models answer?

Individual-level risk prediction

- Who might develop disease?
- Who could be offered prevention?
- How does risk vary across groups?

Population-level forecasting

- How many cases might occur?
- Where could service demand rise?
- How should capacity be planned?

Example: Diabetes Population Risk Tool (DPoRT)

Prediction modelling can operate at different levels, answering different public health questions.

Level	Example prediction	Public health application
Individual	Who is likely to develop diabetes in the next 10-years?	Targeted prevention
Population subgroup	Which groups have the highest predicted burden of diabetes burden?	Identifying priority populations
Geographic	Which communities are expected to experience greater burden of diabetes?	Geographic resource allocation
Temporal	How will the future burden of diagnoses, or complications change across the next 20-years?	Surveillance and capacity planning

Example from Laura Rosella's work: The Diabetes Population Risk Tool (DPoRT) estimates future type 2 diabetes risk using population survey information. Individual risk estimates can be aggregated to characterize population risk and inform prevention planning.

Rosella et al. J Epidemiol Community Health. 2011 doi: 10.1136/jech.2009.102244.

2. Statistical and Machine Learning Models

Does a more complex algorithm improve the decision?

Machine Learning is an approach within AI

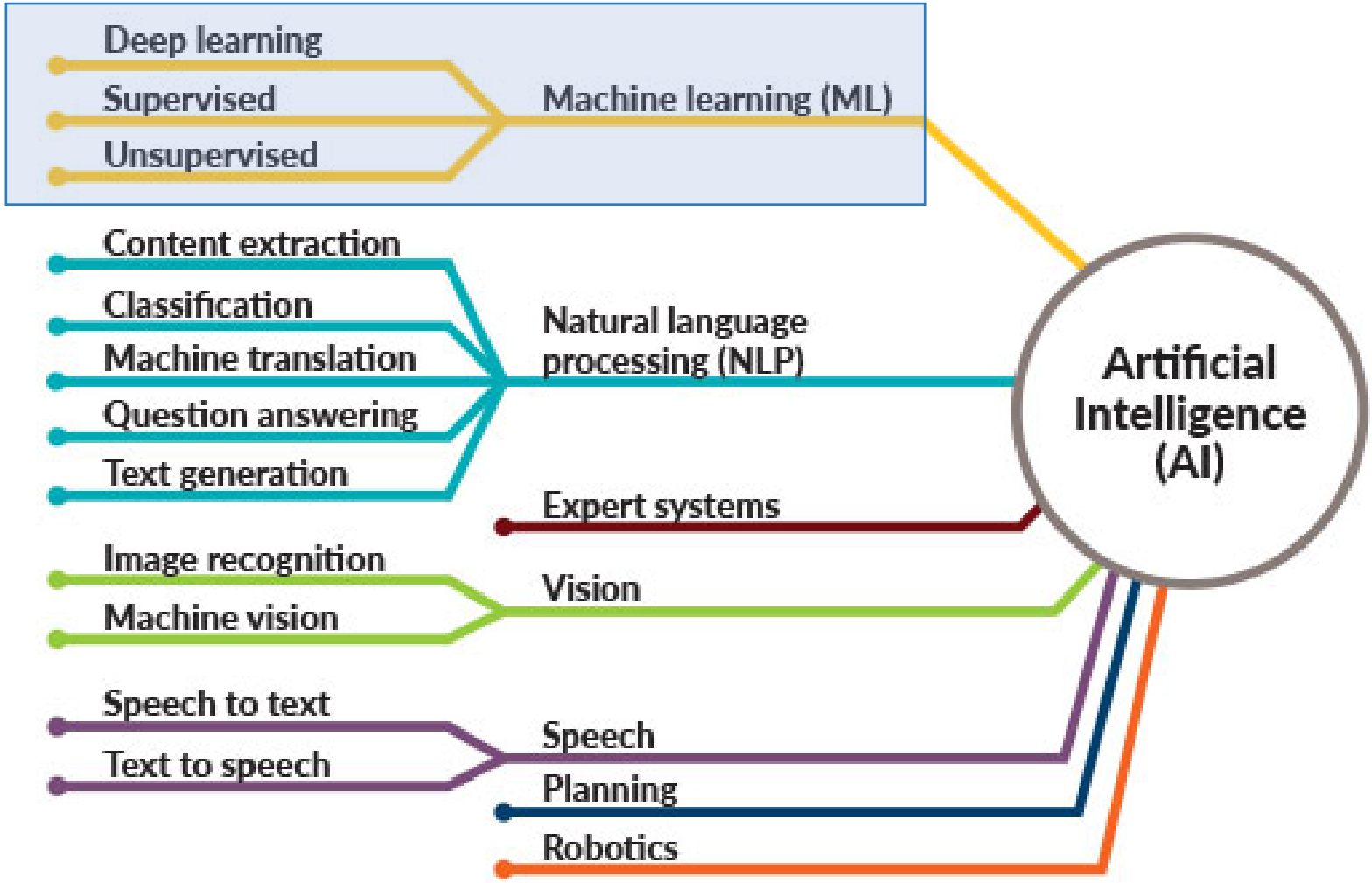


Image source: https://www.automation.com/en-us/articles/august-2022/ai-machine-learning-human-intelligent-systems

Modelling approaches: complexity is a choice, not a goal

	Traditional models Ex. Logistic, Poisson, Cox Regression	Machine learning Ex. Random forests, XGBoost, neural networks
Model specification	▪ Relationships between predictors and outcomes are **explicitly specified**.	▪ Relationships are **learned more flexibly** from the data.
Nonlinearity and interactions	▪ Typically require **explicit specification** (e.g., splines, interaction terms).	▪ Can capture complex patterns without explicitly specifying each relationship.
Interpretability	▪ Coefficients often have familiar statistical interpretations.	▪ Often requires additional methods to explain predictions.
Complexity	▪ Often fewer tuning requirements and lower computational demands.	▪ Typically more hyperparameter tuning and computational demands.

"Machine Learning" does not refer to one model

Different ML approaches can be used for risk prediction

Penalized regression
LASSO, ridge, elastic net

Useful when there are many predictors; shrinks or selects variables.

Tree-based methods
Decision trees, random forests

Capture interactions and nonlinearities; often robust and flexible.

Boosting methods
XGBoost, LightGBM, CatBoost

Sequentially improve prediction; often strong performance on tabular health data.

Support vector machines

Can classify complex patterns using different kernels.

Neural networks / deep learning

Learn highly complex relationships; often require larger datasets and more tuning.

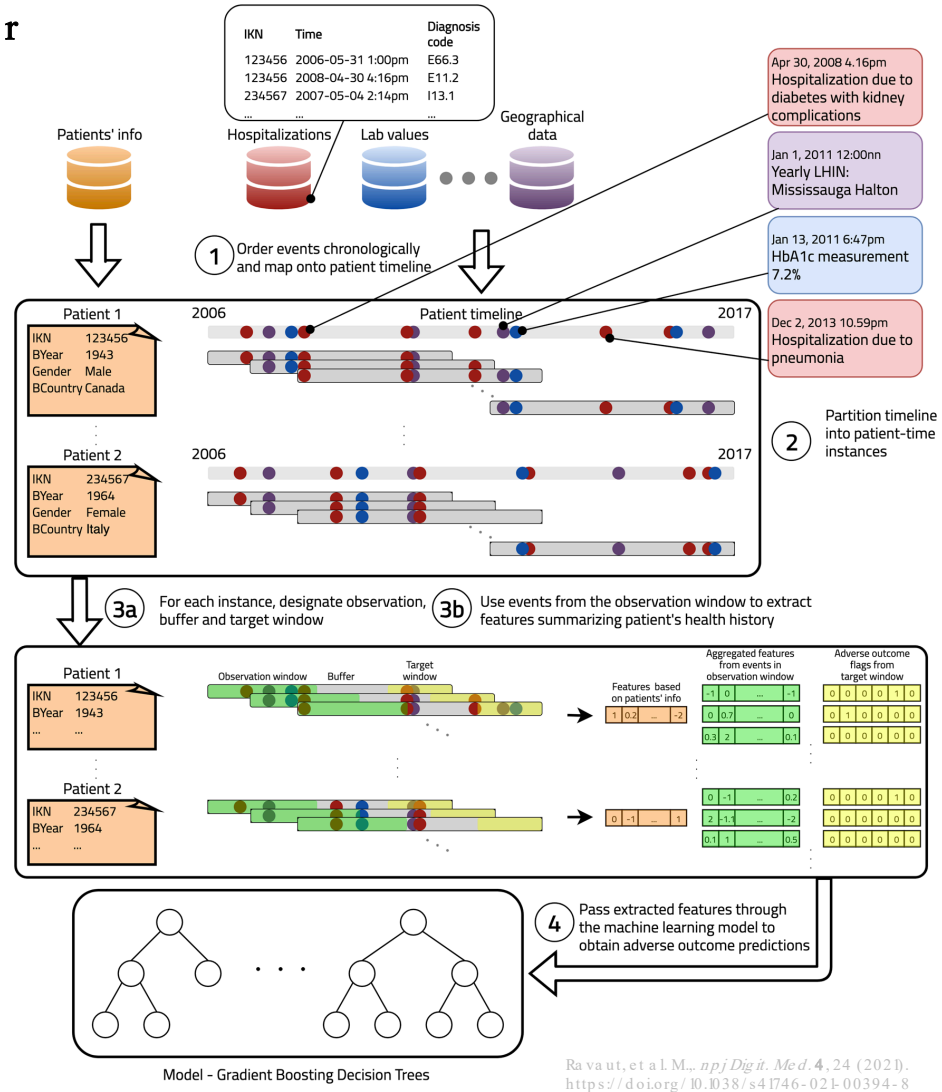
Survival machine learning
Random survival forests, DeepSurv

Useful when predicting time-to-event risk.

The appropriate method depends on the prediction task, the data, and the need for interpretability.

Example: Population Diabetes Risk Prediction

DPORT- MLPaper



DPoRT—Diabetes Population Risk Tool

- Population-based algorithm estimating 10-year risk of developing type 2 diabetes.
- Uses a traditional regression-based approach with routinely collected risk factors.

DPoRT-ML—Machine-Learning Approach

- Extends population-level diabetes risk prediction using machine learning.
- Explores more flexible relationships, including nonlinearities and interactions among predictors.
- Illustrates opportunities to incorporate richer, more complex health data.

The key difference: ML can capture more complex relationships in the data, but greater complexity does not necessarily mean better prediction.

Ravaut, et al. M., *npj Digit. Med.* **4**, 24 (2021). https://doi.org/10.1038/s41746-021-00394-8

3. Building & Evaluating a Prediction Machine Learning Model

How do we move from data to trustworthy risk estimates?

A prediction modelling workflow

cmaj manuscript

Example code

Specify use case → **Build cohort** → **Choose predictors** → **Split & train** → **Validate** → **Evaluate action**

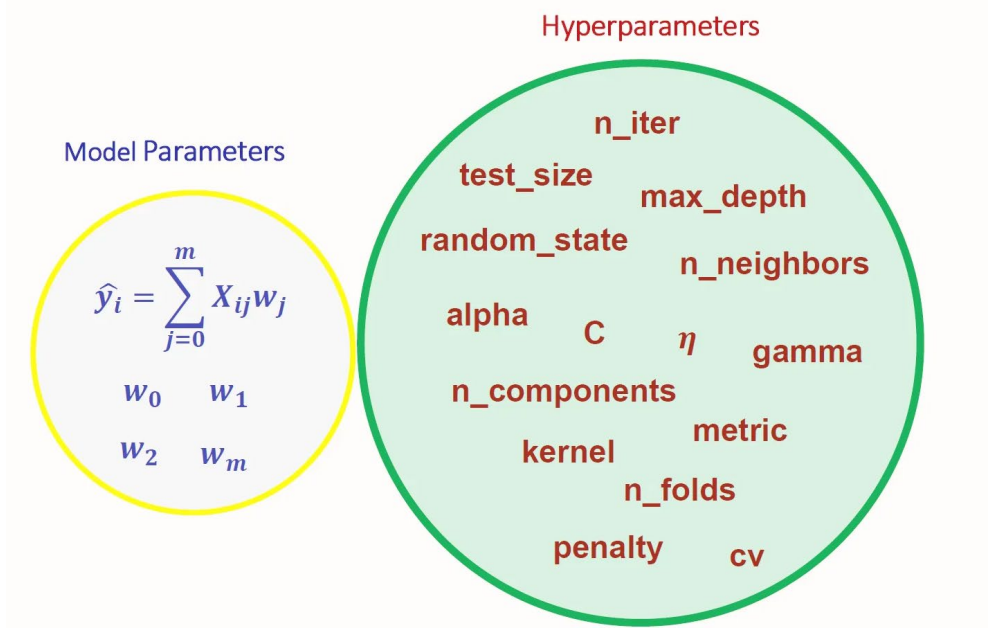
- Predictors must be available when the prediction is made
- Separate training from temporal/geographic external validation
- Prevent leakage, assess missingness and provenance, and preserve the target population

What makes ML model training unique?

cmaj manuscript

Example code

Model Hyperparameters



Source: https://towardsdatascience.com/model-parameters-and-hyperparameters-in-machine-learning-what-is-the-difference-702d30970f6

Approaches for Selecting Hyperparameters

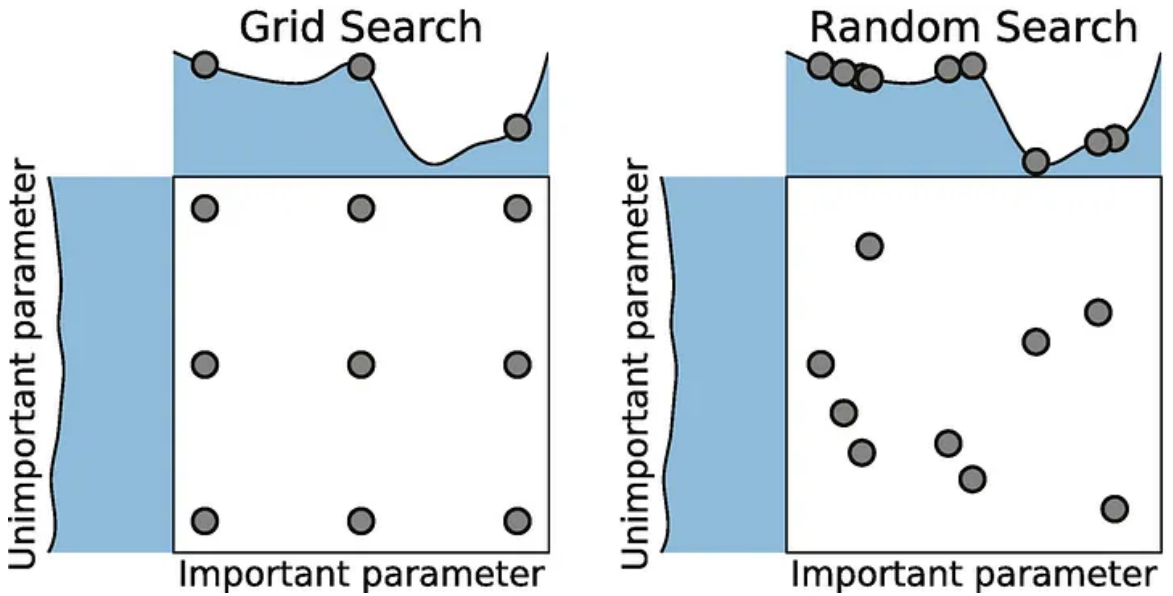


Image Source: Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. Journal of Machine Learning Research 13, 281–305 (2012)

Model Evaluation: Does added flexibility always mean better?

cmaj manuscript

Example code

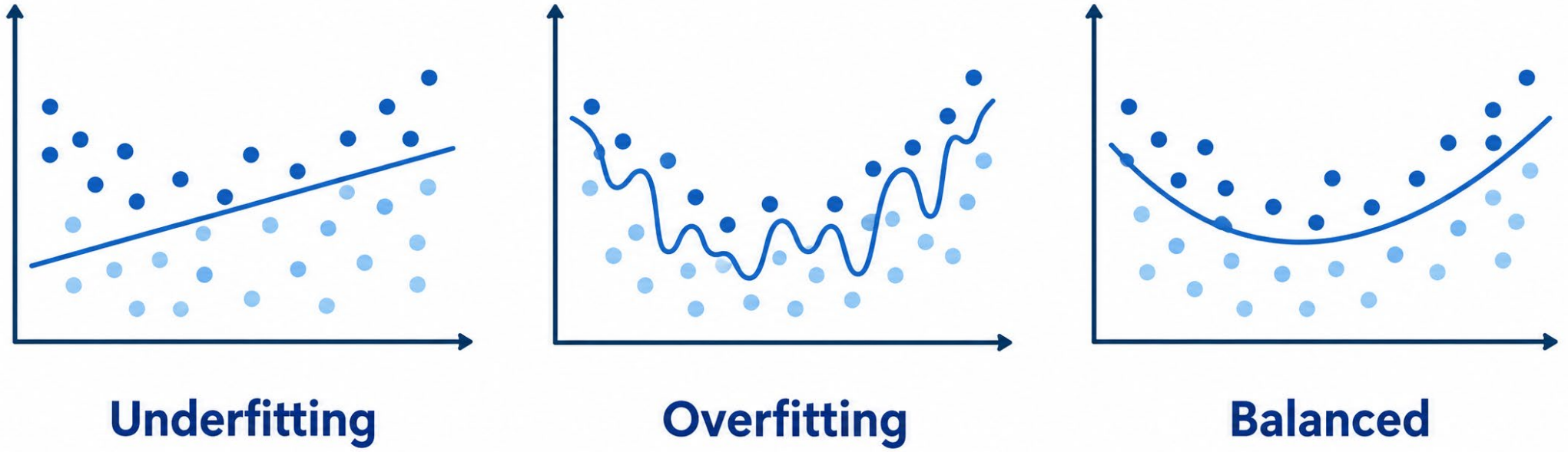
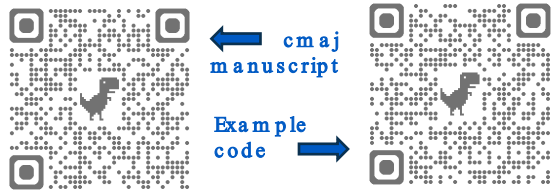


Image source: https://www.r-bloggers.com/2022/09/how-to-avoid-overfitting/

Model Evaluation: Does added flexibility always mean better? - 2



Discrimination

Can the model rank higher-risk people above lower-risk people?

AUROC / AUPRC

Calibration

Are predicted probabilities consistent with observed risk?

Calibration plot / slope

Threshold performance

Who is flagged for action at a chosen cutoff?

Sensitivity / specificity / PPV

Decision utility

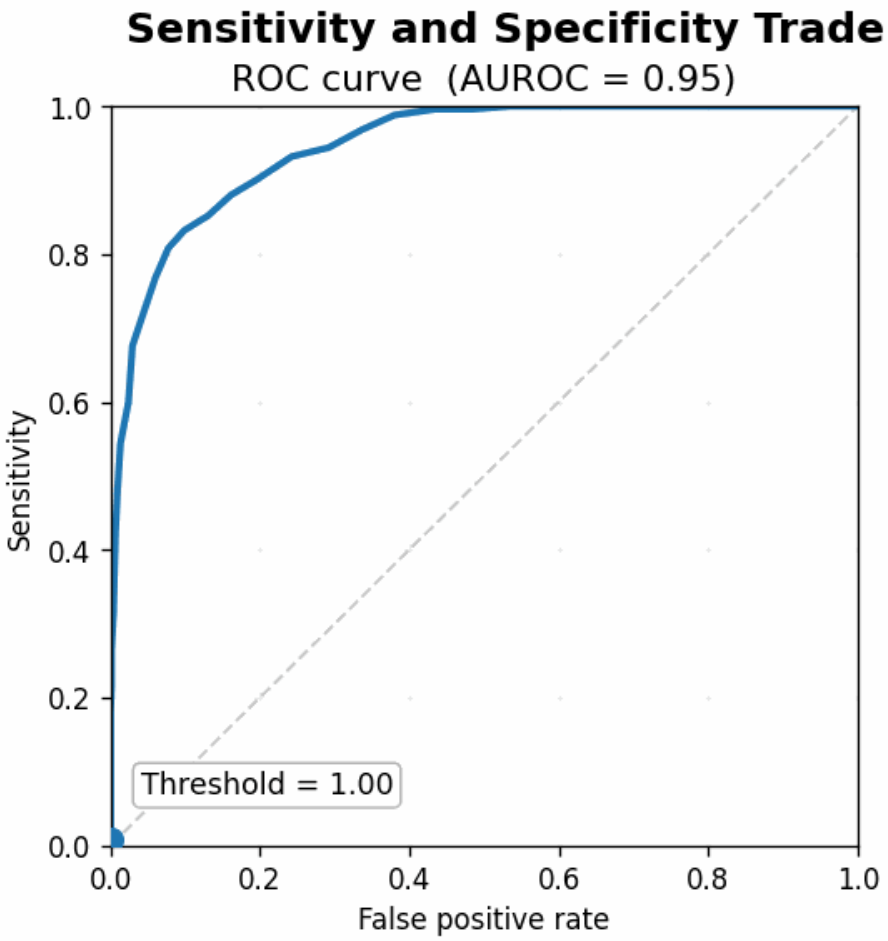
Does acting on the model improve decisions?

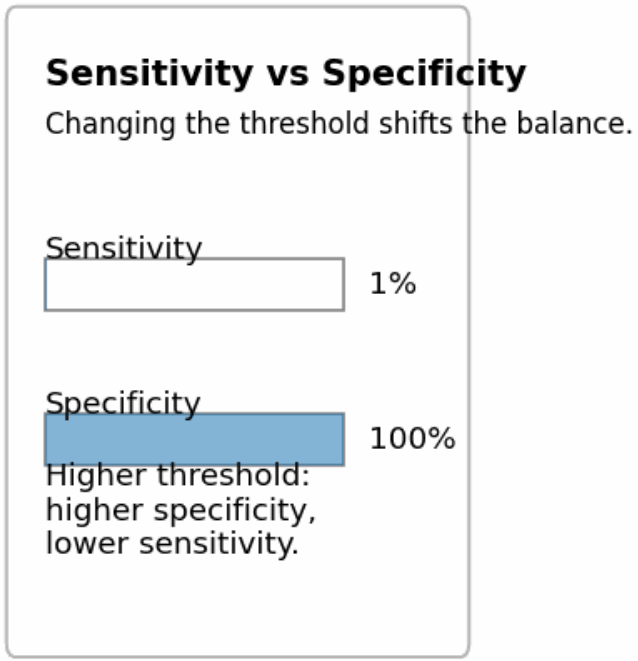
Net benefit / impact evaluation

Model Evaluation: Discrimination

cmaj manuscript

Example code

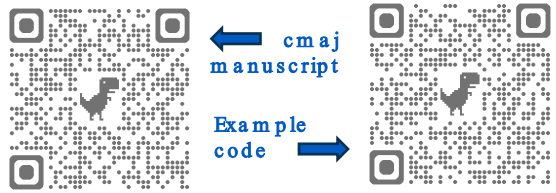




ML model discriminations is typically evaluated by AUROC.

The clinical value of AUROC depends on the selected threshold and thus, the trade-off between sensitivity and specificity

Choose risk thresholds for the public health action



Lower threshold

- More people eligible
- Greater sensitivity (typically)
- Higher program demand
- More false positives

Higher threshold

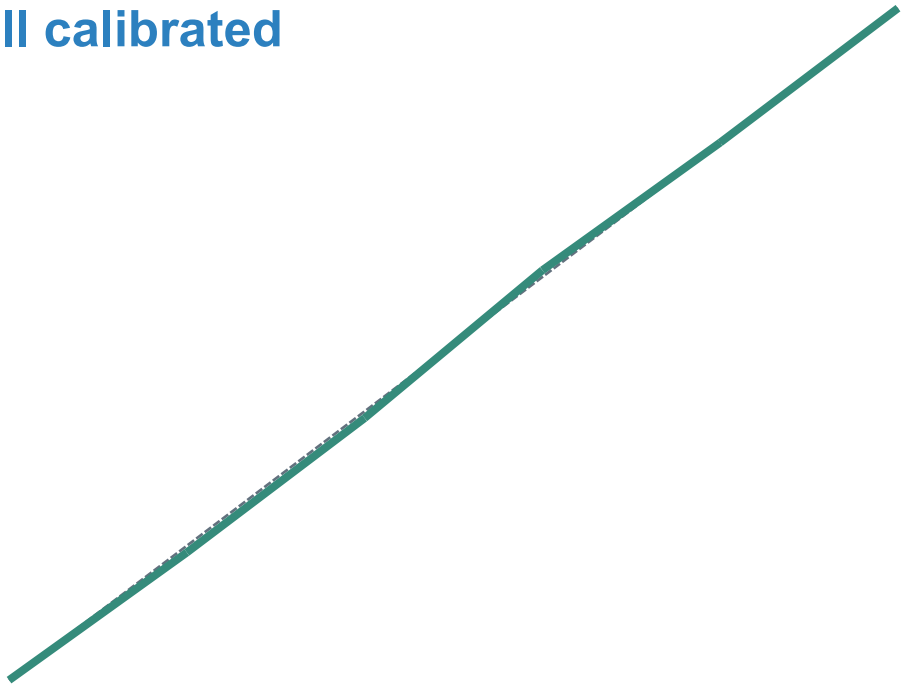
- Fewer people eligible
- Greater specificity (typically)
- Lower program demand
- More missed cases

The best threshold depends on intervention benefit, capacity, costs and equity.

Model Evaluation: Discrimination ≠ calibration

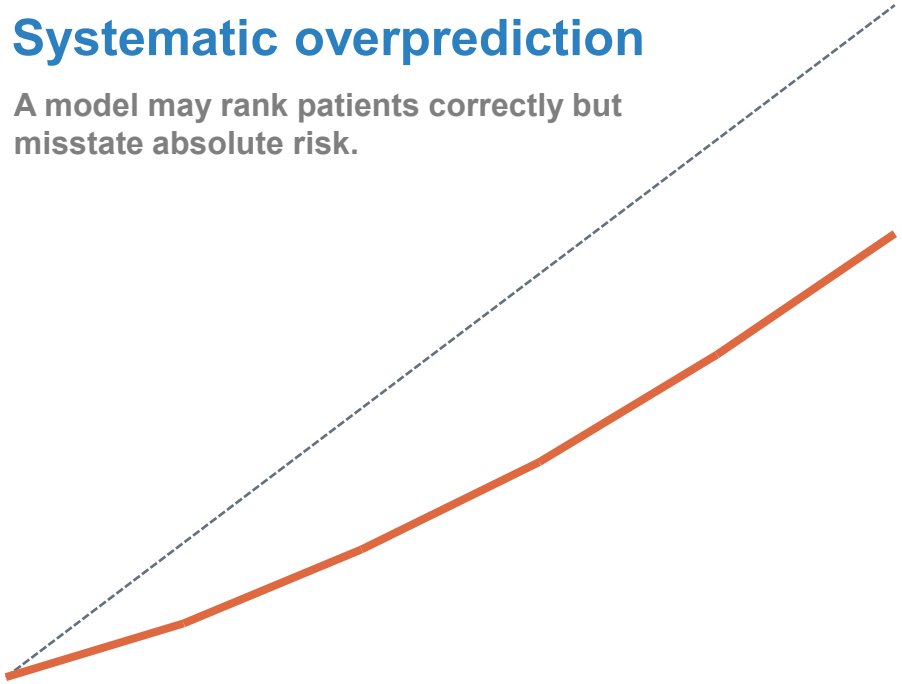
The following models might have the same discrimination

Well calibrated



Systematic overprediction

A model may rank patients correctly but misstate absolute risk.

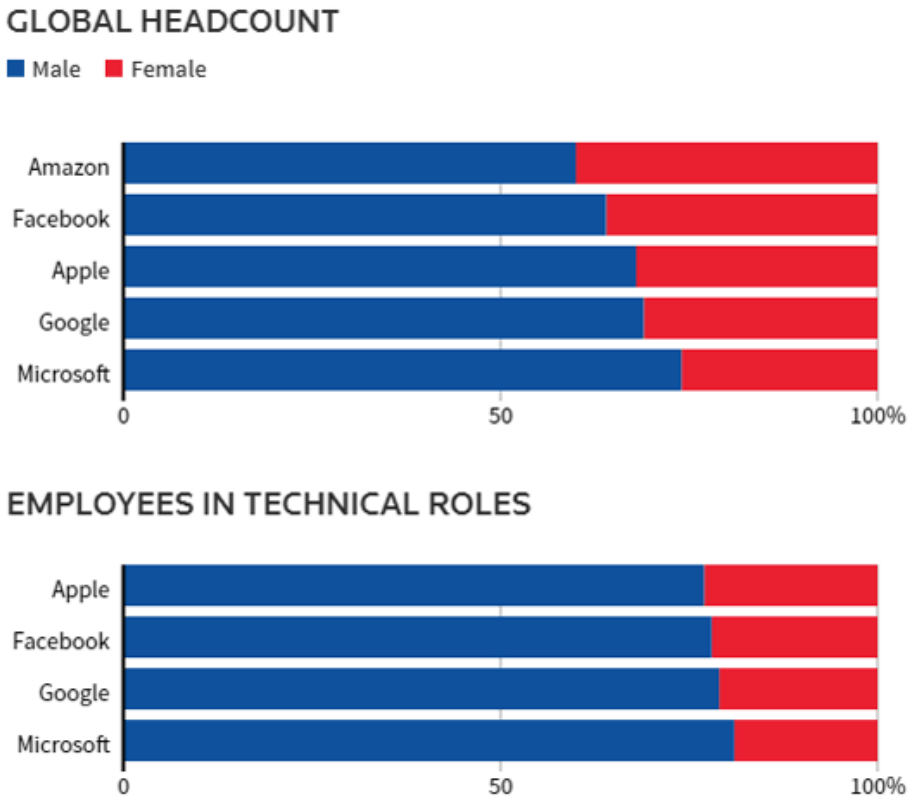


Illustrative calibration plots (not empirical study results).

Model Evaluation: Subgroup evaluation

In 2018, Amazon scraped its secret AI recruiting tool that showed bias against women!

The system penalized resumes that included the word "Female", as in "President of Female Business Leadership Group Chapter."



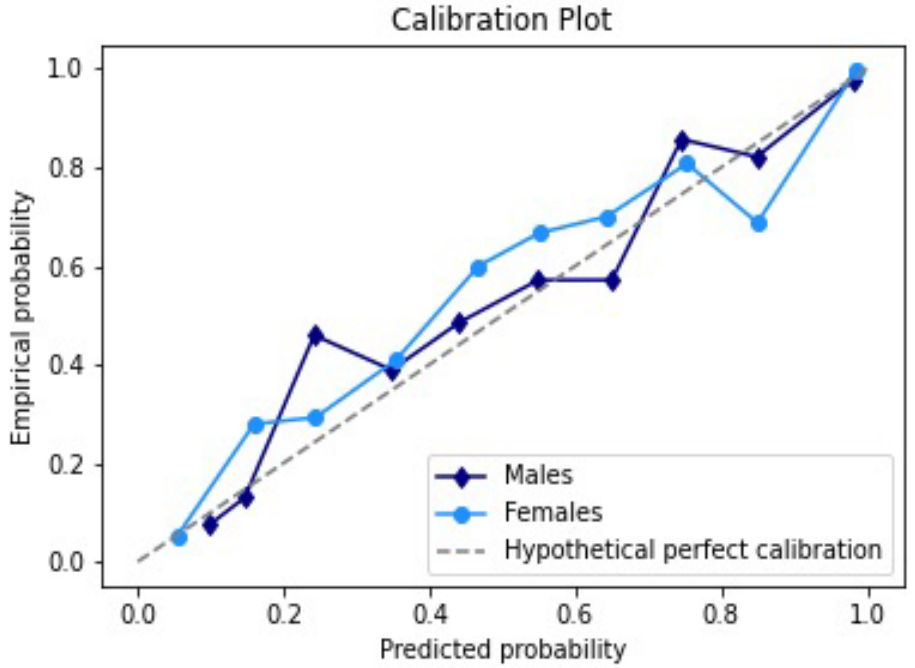
Model Evaluation: Subgroup evaluation - 2

cmaj manuscript

Example code

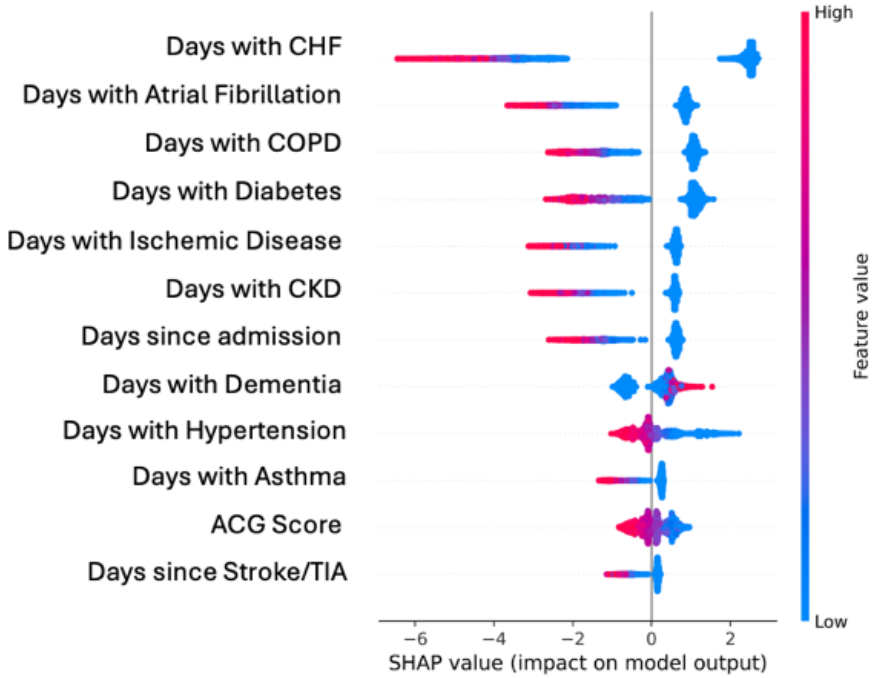
Evaluating across subgroups can help detect potential bias and support equitable model development

	Overall (N=1856)	Females (N=898)	Males (N=958)
Accuracy	0.89	0.90	0.88
Precision	0.89	0.91	0.88
Recall	0.89	0.90	0.88
F1score	0.89	0.90	0.88
AUC	0.89	0.91	0.88
False Positives, n(%)	57 (3.1)	24 (2.7)	33 (3.4)
False Negatives, n(%)	147 (7.9)	63 (7.0)	84 (8.8)

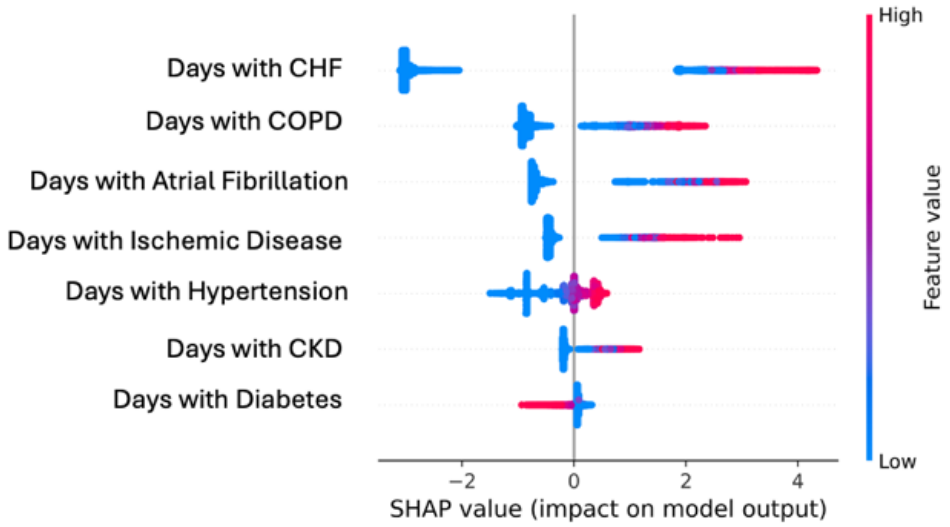


Model explainability – and our misinterpretation of explainability

(A) Feature importance of frail model – Cluster B
Measured as utility of features (and their values) to predict Cluster B (vs. Cluster A)



(B) Feature importance of *non*-frail model – Cluster C
Measured as utility of features (and their values) to predict Cluster C (vs. Cluster D or E)



4. Responsible Implementation

What needs to be true before a prediction tool is used?

Checklist for responsible use of AI prediction models in public health

- External validation in the intended populations, settings and calendar period
- Transparent reporting, appropriate data governance and privacy protections
- Calibration, subgroup performance, and assessment of inequitable downstream consequences
- Human oversight, interpretable decision pathways and clear accountability
- Prospective impact evaluation, drift monitoring, updating and retirement criteria

Responsible use = valid model + defensible action + ongoing governance.

Better metrics ≠ better health outcomes

Model

Higher discrimination and appropriate calibration

→

Decision

Action changes because risk is available

→

Impact

Prevention, better outcomes, equitable value

Evaluate prospective impact, costs, unintended consequences, and who benefits.

Five take-home messages if you want to build an ML prediction model

- Start with the public health decision—not the algorithm
- Compare ML with strong statistical baselines
- Assess calibration and decision utility, not discrimination alone
- Do not mistake predicted risk for preventable risk
- Real-world benefit requires implementation science, equity and monitoring

Thank you!

Let’s Connect

Email: gemma.postill@utoronto.ca
LinkedIn: https://www.linkedin.com/in/gemma-p-80481987/