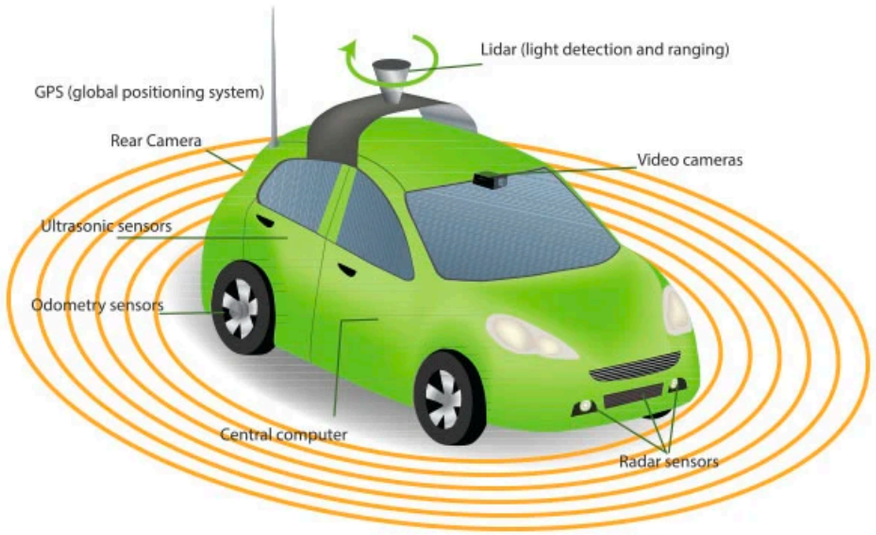
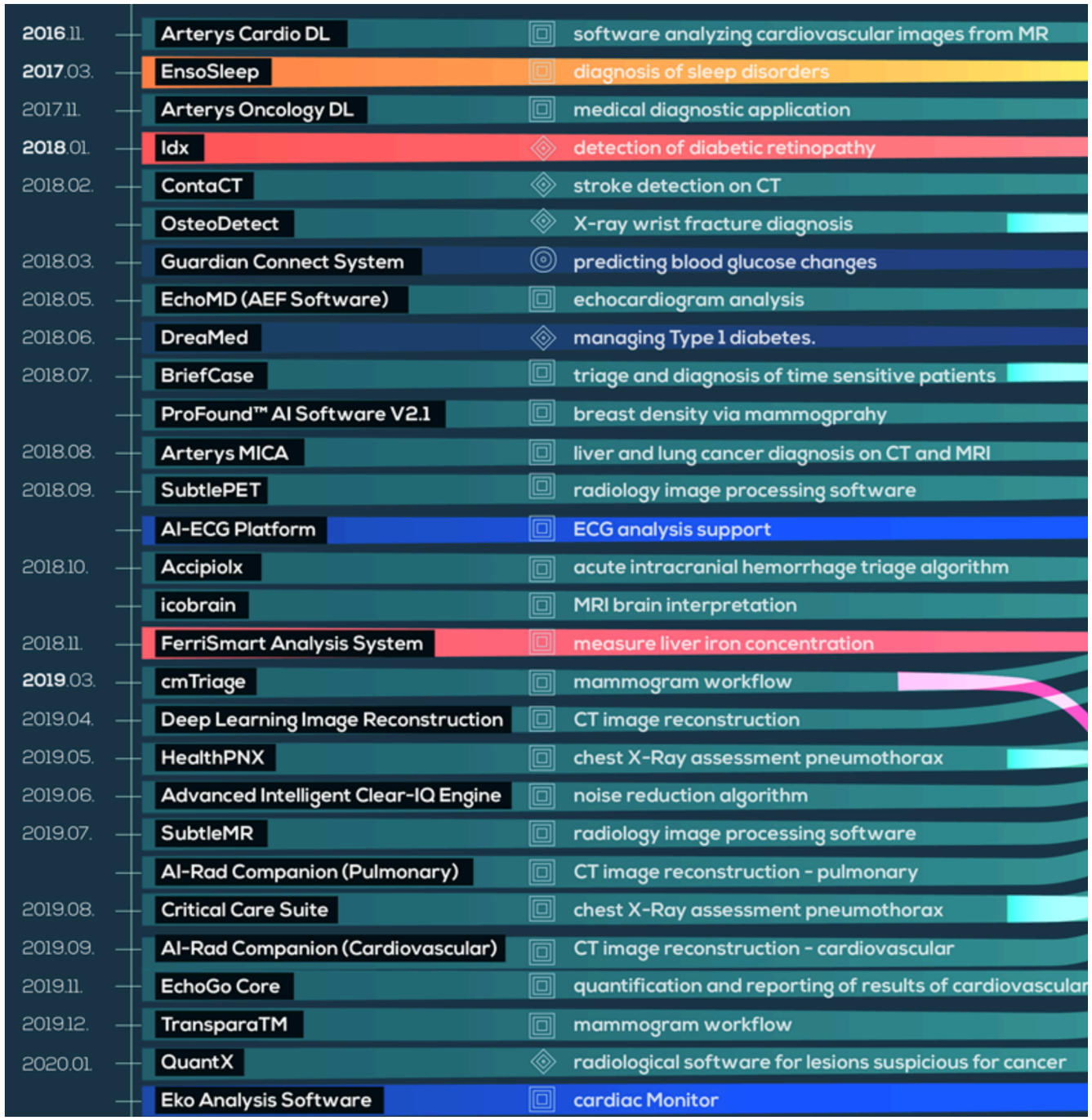
# Opportunities at the intersection of Machine Learning and Epidemiology

Jean Feng
University of California, San Francisco


*IEEE*



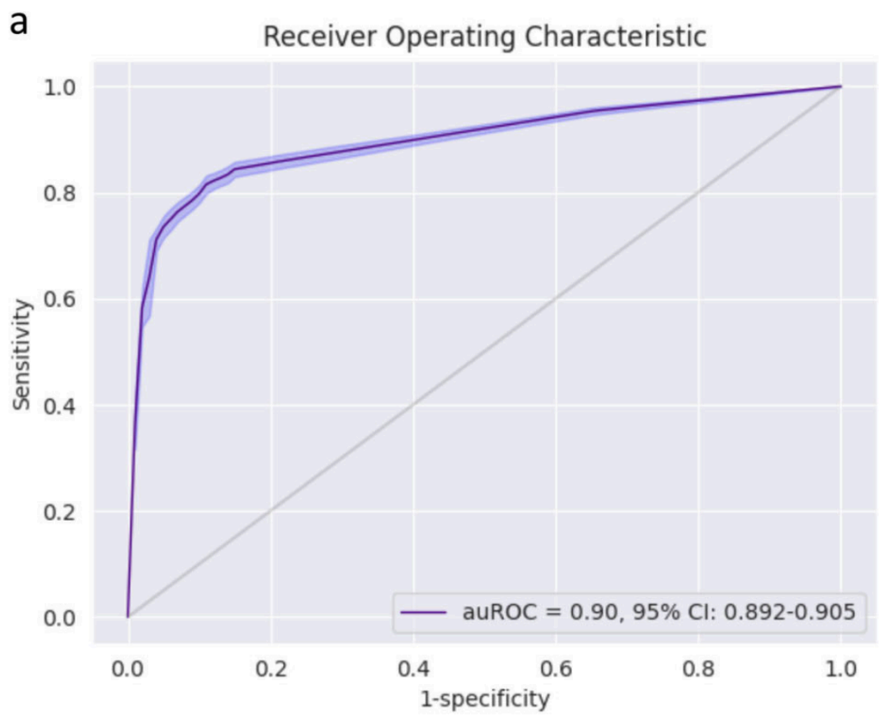

*Benjamens 2020*

ARTICLE OPEN

Check for updates

# Machine learning-based prediction of COVID-19 diagnosis based on symptoms

Yazeed Zoabi[1], Shira Deri-Rozov[1] and Noam Shomron[1]✉





ARTICLE OPEN

Check for updates

# Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19

Sonu Subudhi[1], Ashish Verma[2,10], Ankit B. Patel[2,10], C. Corey Hardin[3], Melin J. Khandekar[4], Hang Lee[5], Dustin McEvoy[6], Triantafyllos Stylianopoulos[7], Lance L. Munn[8], Sayon Dutta[6,9]✉ and Rakesh K. Jain[8]✉

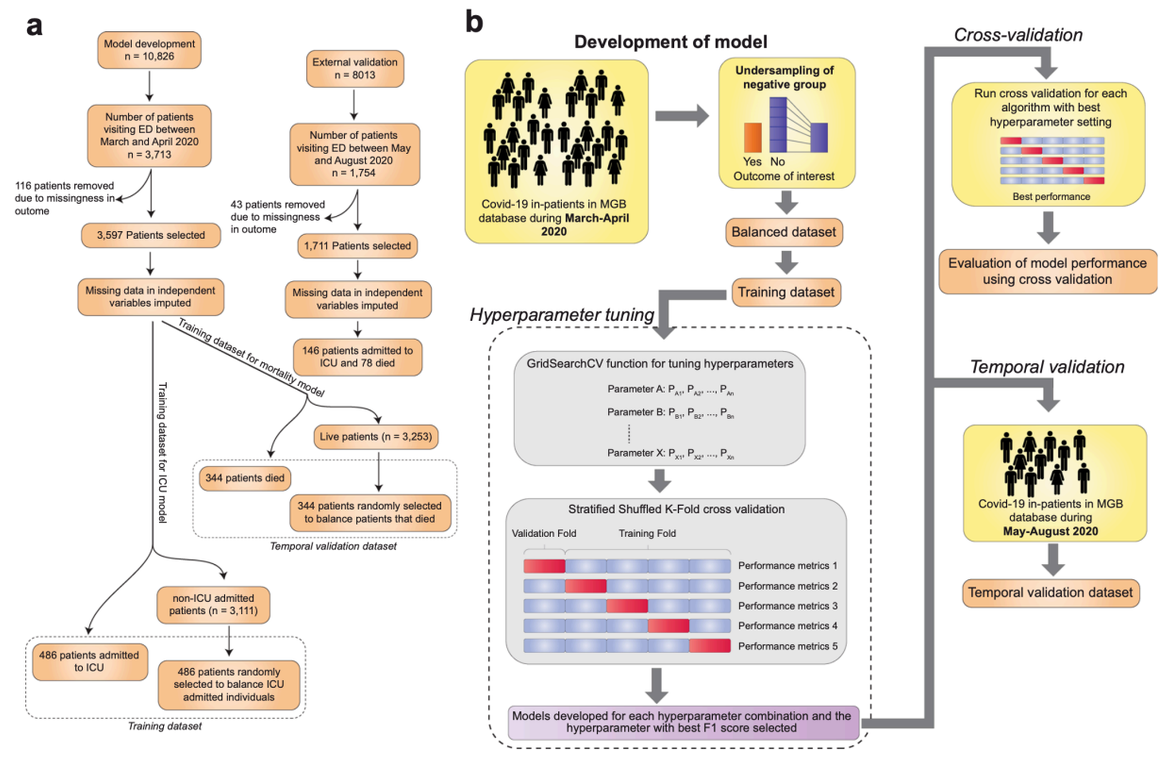
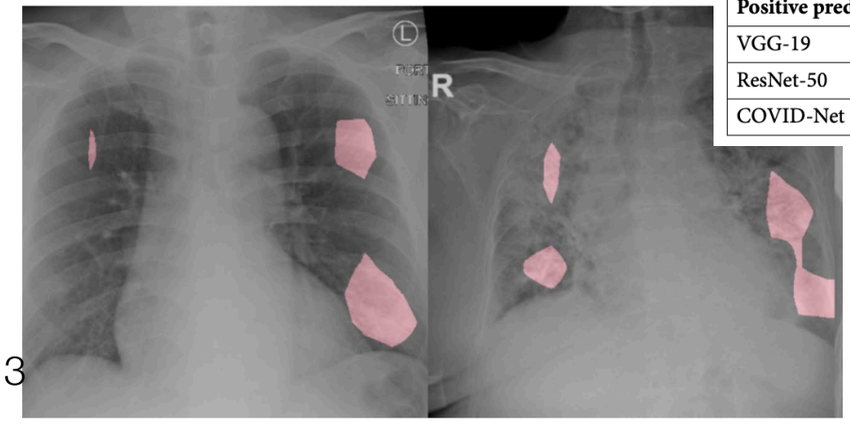



**Fig. 1 Schematic diagram representing the process of machine learning model development. a** Flow diagram depicting steps in obtaining the training and temporal validation datasets (with patient numbers in each step). **b** The process of patient selection, dataset balancing, hyperparameter tuning, cross-validation and temporal validation are shown.

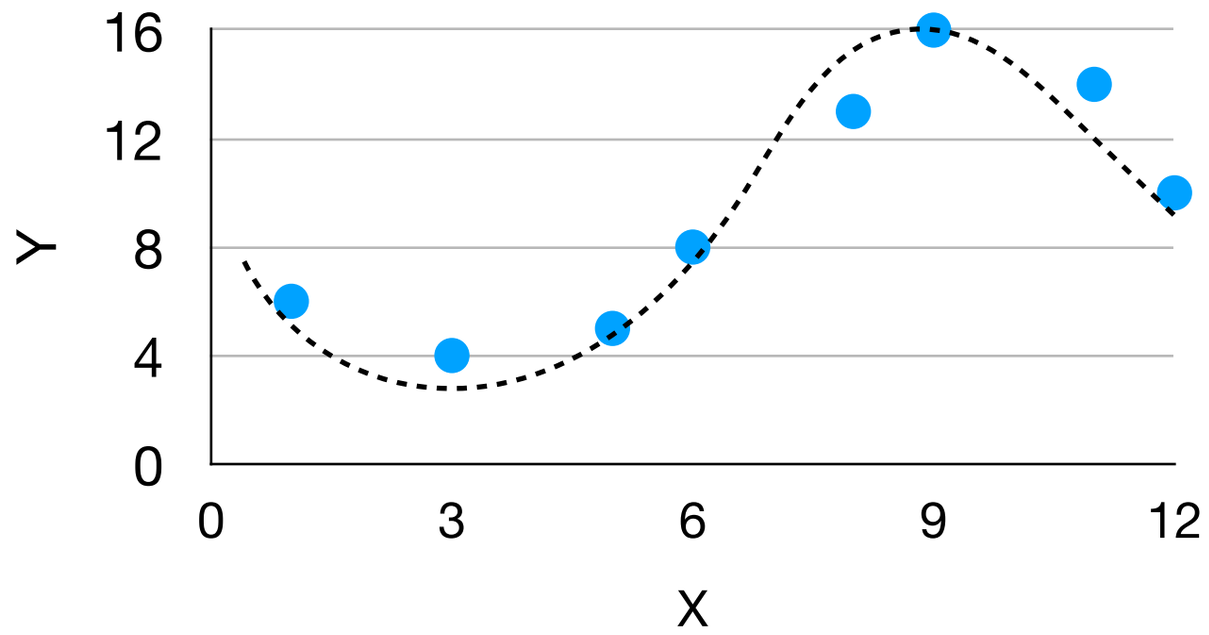
# COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images

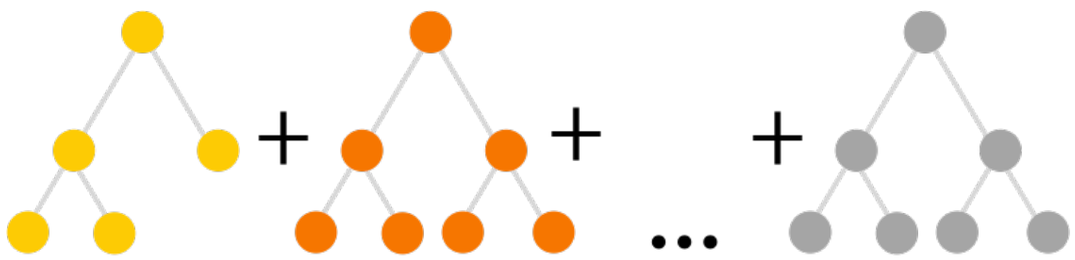
Linda Wang[1,2,3]✉, Zhong Qiu Lin[1,2,3] & Alexander Wong[1,2,3]

| Architecture | Normal | Non-COVID19 | COVID-19 |
|---|---|---|---|
| **Positive predictive value (%)** | | | |
| VGG-19 | 83.1 | 75.0 | 98.4 |
| ResNet-50 | 88.2 | 86.8 | 98.8 |
| COVID-Net | **90.5** | **91.3** | **98.9** |

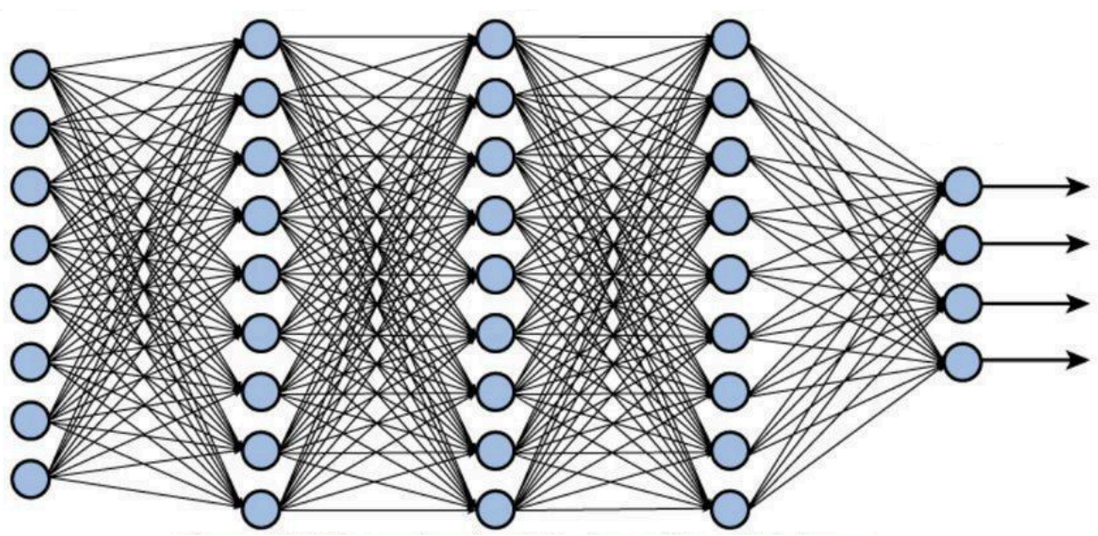


$$X \mapsto Y$$



Gradient boosted trees

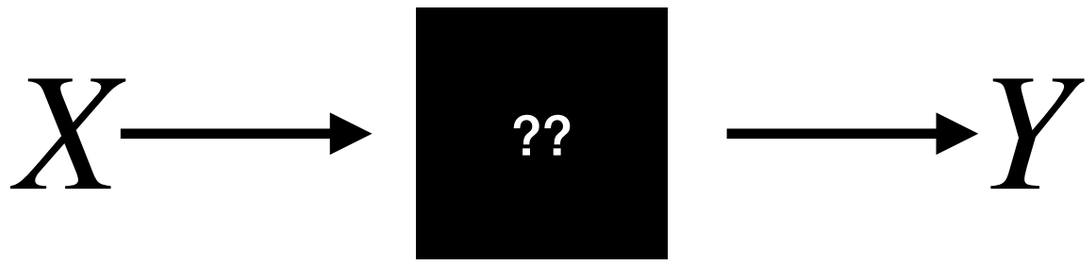


Neural network

**ML for Epidemiology:** Interpretable ML

**Epidemiology for ML:** Surveillance of ML

# Interpretable ML




original report

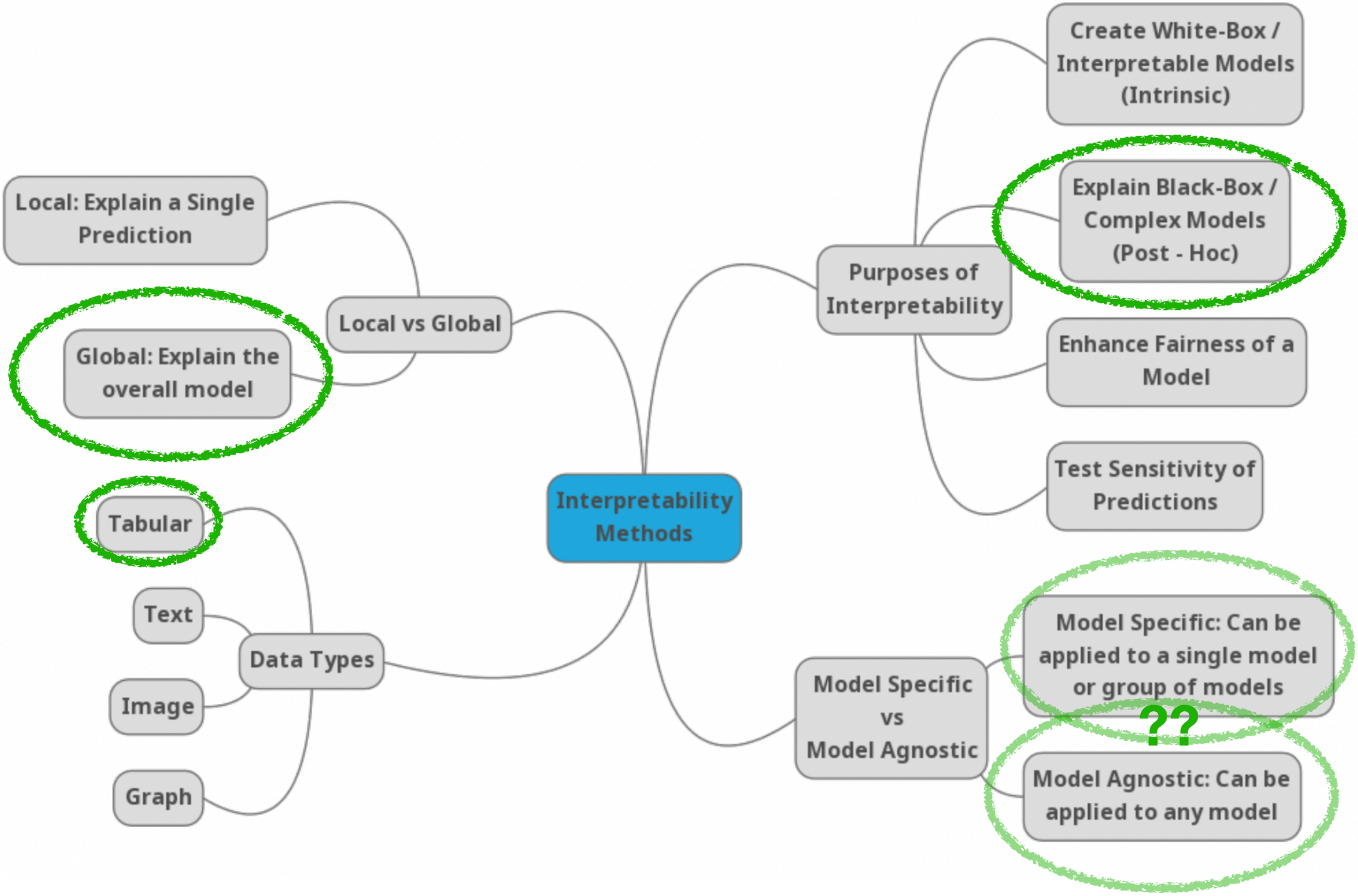
## Predicting Emergency Visits and Hospital Admissions During Radiation and Chemoradiation: An Internally Validated Pretreatment Machine Learning Algorithm

Julian C. Hong
Donna Niedzwiecki
Manisha Palta
Jessica D. Tenenbaum

**Table 1.** Variables Used to Train Machine Learning Algorithms

| Variable | No. of Levels or Variables |
|---|---|
| Demographic | |
| Sex (male, female) | 2 |
| Race | 12 |
| Age at start of treatment | Continuous |
| Ethnic group | 8 |
| Marital status | 7 |
| Religion | 46 |
| Zip code | 1,248 |
| Disease and treatment | |
| Primary treatment diagnosis (by subchapter/by three-digit ICD code with metastatic sites based on full ICD code) | 59/172 |
| Planned RT dose (Gy) | Continuous |
| Planned No. of RT fractions | Continuous |
| RT techniques used (2D or 3D conformal RT, intensity-modulated RT or volumetric modulated arc therapy, stereotactic radiosurgery/stereotactic body RT, total skin irradiation) | 5 |
| Any concurrent antineoplastic drugs (first 2 weeks of radiation) | Indicator |
| Concurrent antineoplastic drugs by MeSHPA class/RxNorm agent | 51/86 |
| Any recent antineoplastic drugs (6 months before radiation) | Indicator |
| Recent antineoplastic drugs by MeSHPA class/RxNorm agent | 58/109 |
| Treating radiation oncologist | 26 |
| Recent encounters before treatment in EHR | |
| Time since most recent admission and emergency visit before start of radiation | Continuous |
| No. of admissions in the month and year before start of radiation | Continuous |
| No. of days admitted in the year before start of radiation | Continuous |
| No. of emergency visits in the month and year before start of radiation | Continuous |
| Started RT as inpatient | Indicator |
| Medical history known at start of radiation | |
| All prior diagnosis and problem list ICD history (by ICD subchapter) | 269 |
| All prior CPT history | 9,236 |
| All prior level-3 Agency for Healthcare Research and Quality category history | 323 |
| Medications before and at start of therapy | |
| All recent medications (6 months before radiation; MeSHPA class) | 298 |
| All active medications at start of radiation (MeSHPA class) | 295 |
| Social history | |
| Reported tobacco use | 5 |
| Reported alcohol use | 3 |
| Reported illicit drug use | 3 |
| Reported sexually active | 4 |
| Recent laboratory values | |
| Presence of any abnormally flagged laboratory studies in the 4 weeks before start of radiation | 737 |
| Recent vital signs in the year before start of treatment | |
| Weight loss from maximum weight | Continuous |
| Presence of hypertension (SBP ≥ 130 mm Hg, DBP ≥ 80 mm Hg) | Indicator |
| Presence of hypotension (SBP < 90 mm Hg, DBP < 60 mm Hg) | Indicator |

(Continued on following page)




*Linardatos 2021*

# Variable importance (VI)

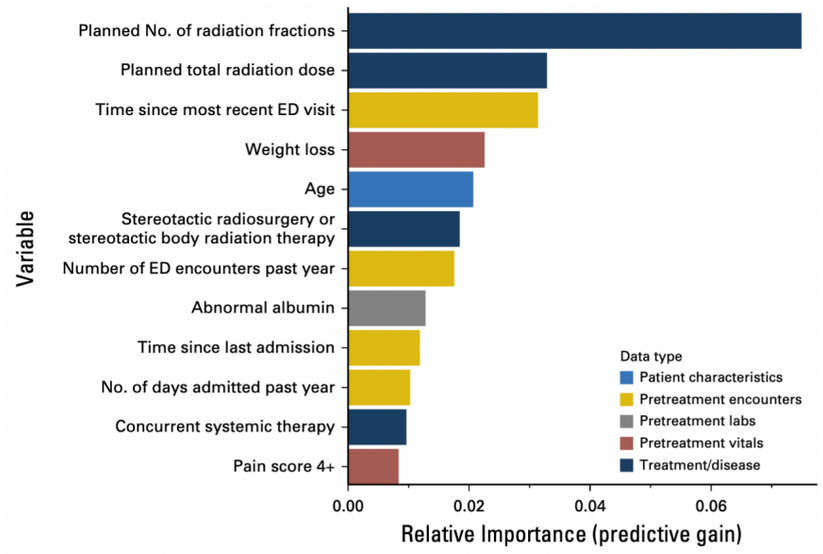
## Statistics/Epidemiology

1950s: $R^2$ and ANOVA for parametric models

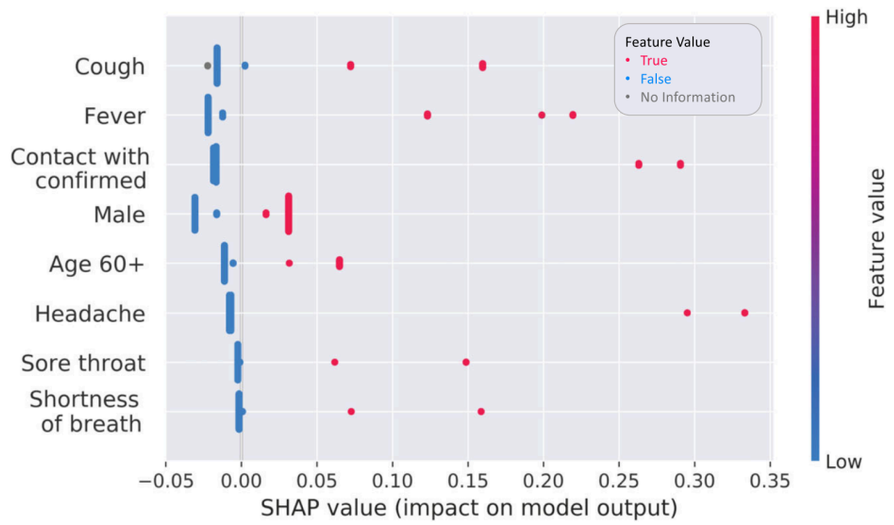
2010s: Nonparametric population-level $R^2$ and variable importance

## Machine learning

2000s: Model-specific VI




2017: Shapley Model-specific VI



# What is the “right” Variable Importance measure?

## Statistics/Epidemiology

### Population-level VI is a property of $\mathbb{E}[Y|X=x]=\mu(x)$

- The population-level VI is an **estimand** and **must be estimated** its value, i.e. using ML. Moreover, we should quantify the uncertainty of our estimates using confidence intervals.

- Population-level VI is **model-agnostic**, so its estimates using different ML algorithms should be similar.

- Although the true Shapley population-level VI is computationally intractable, **we can efficiently calculate estimates** for Shapley population-level VI.

## Machine learning

### Model-specific VI is a property of $\hat{f}(x)$

- Model-specific VI **does not need to be estimated** because we have access to the entire model.

- Model-specific VI measures **vary across different models**.

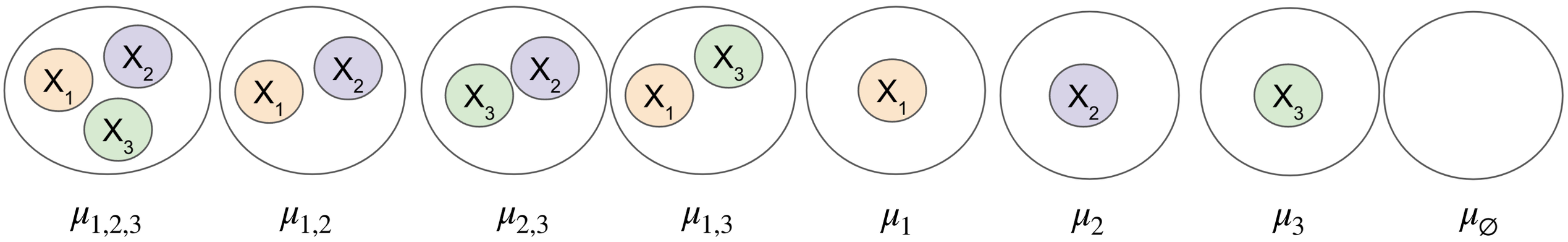
- In general, calculating the Shapley model-specific VI is **computationally infeasible**. Instead, one may rely on approximations.

# Shapley Population-level VI measures (SPVIM)

**Q: How important is variable $X_j$ at explaining the variability of the outcome $Y$ in the population?**

<u>Estimand:</u> $$\psi_j = \sum_{s \in \{1,\cdots,p\} \setminus \{j\}} \frac{1}{p} \binom{p-1}{|s|}^{-1} \underbrace{\left\{ \mathbb{E}\left[(Y - \mu_s(X))^2\right] - \mathbb{E}\left[(Y - \mu_{s \cup \{j\}}(X))^2\right] \right\}}$$

*Change in mean squared error when information on $X_j$ is available*

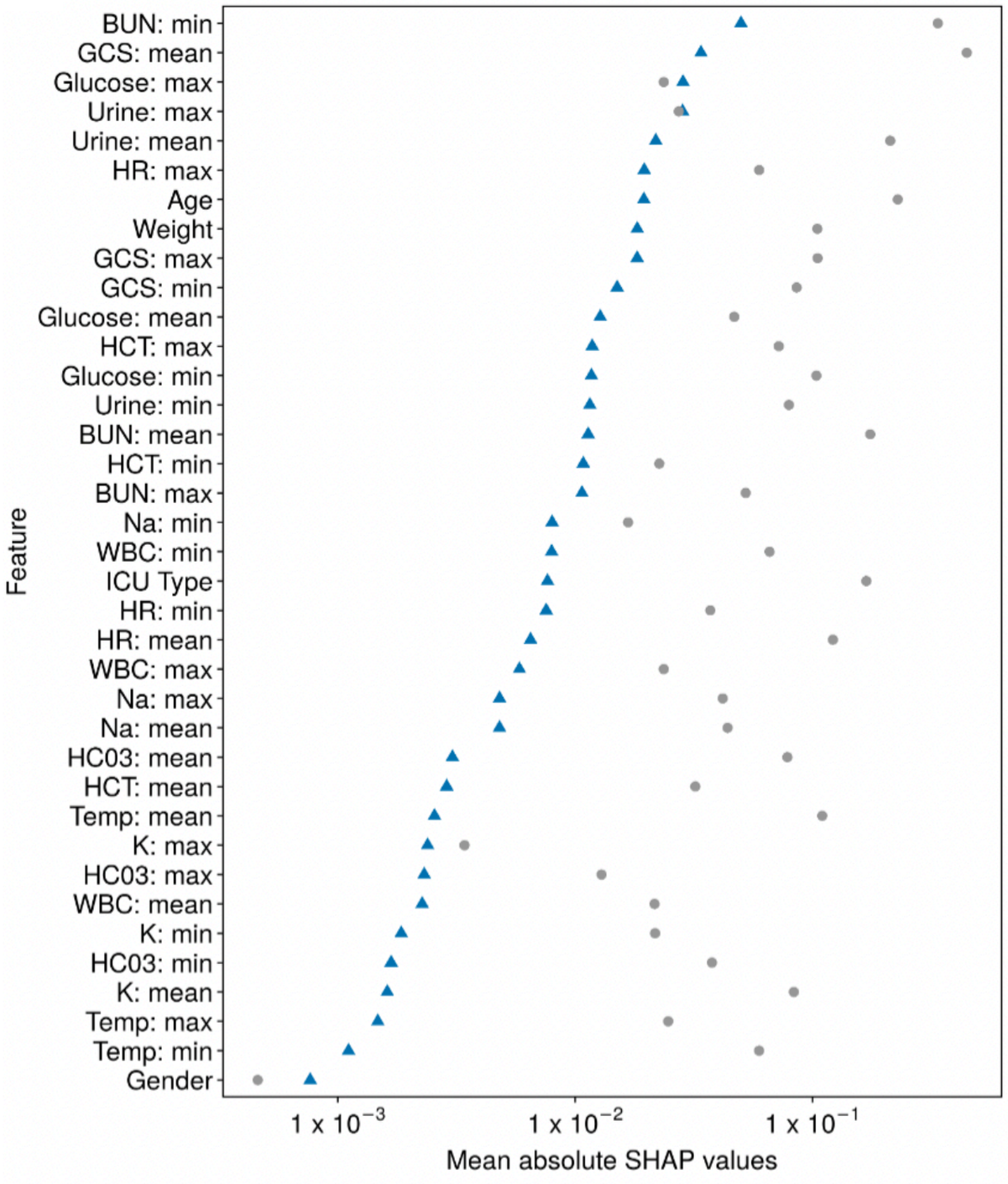



<u>Estimation procedure:</u> For randomly sampled subsets $s$, use ML to estimate $\mu_s$. Employ sample-splitting to estimate the population-level VI and prevent over-fitting.

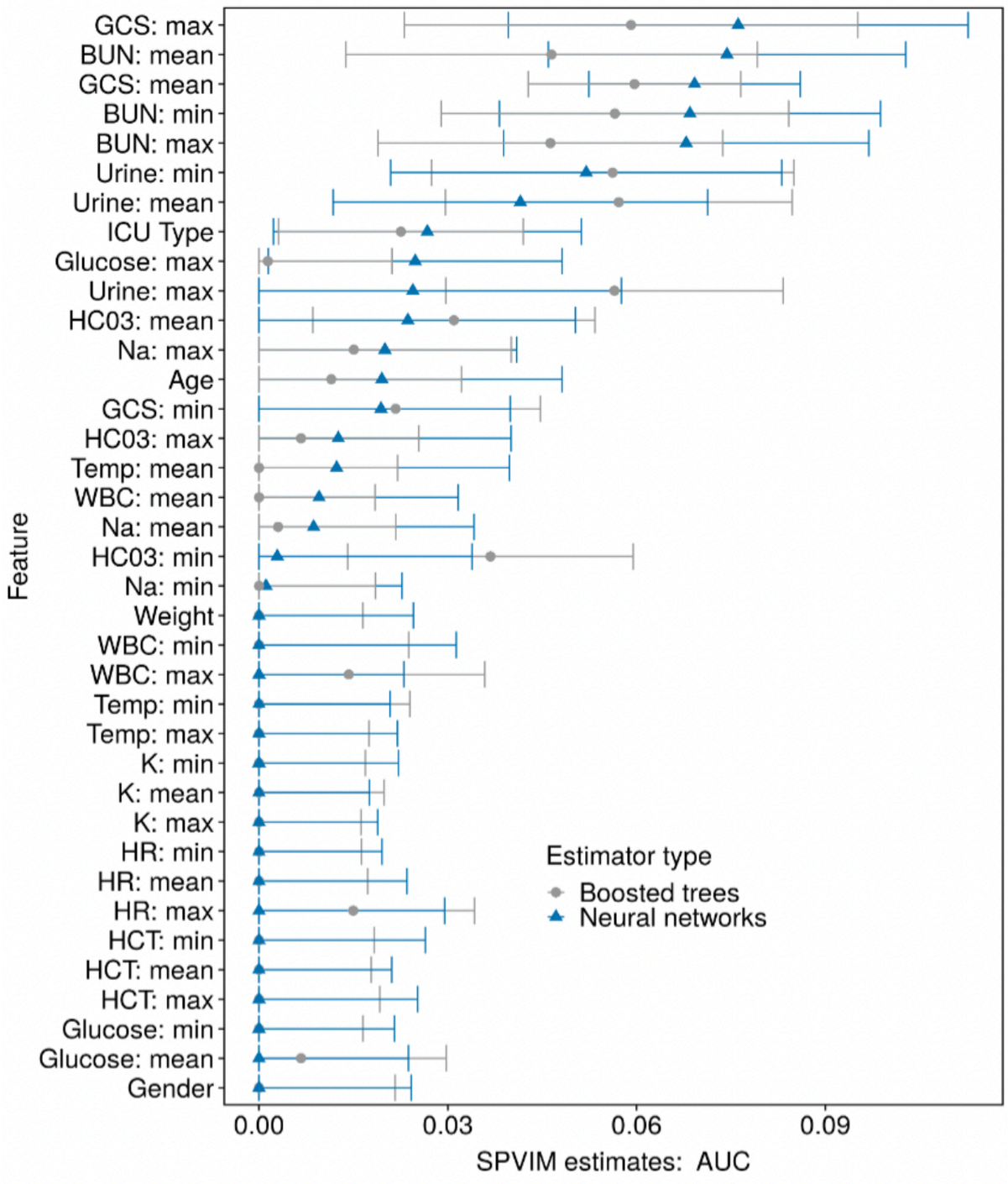
*Williamson and Feng 2020*

# From model-specific VI to population-level VI

## Shapley Model-specific VI




## Shapley population-level VI

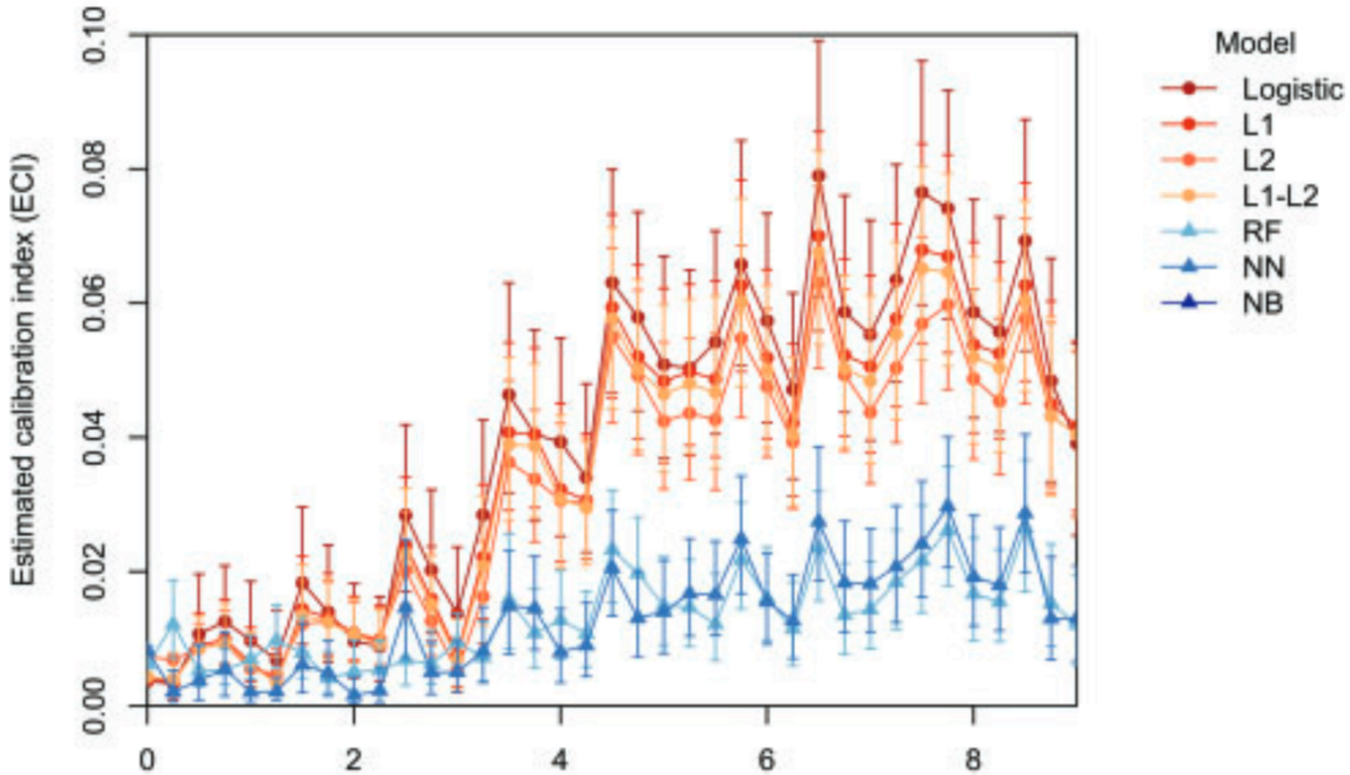


**ML for Epidemiology:** Interpretable ML

**Epidemiology for ML:** Surveillance of ML

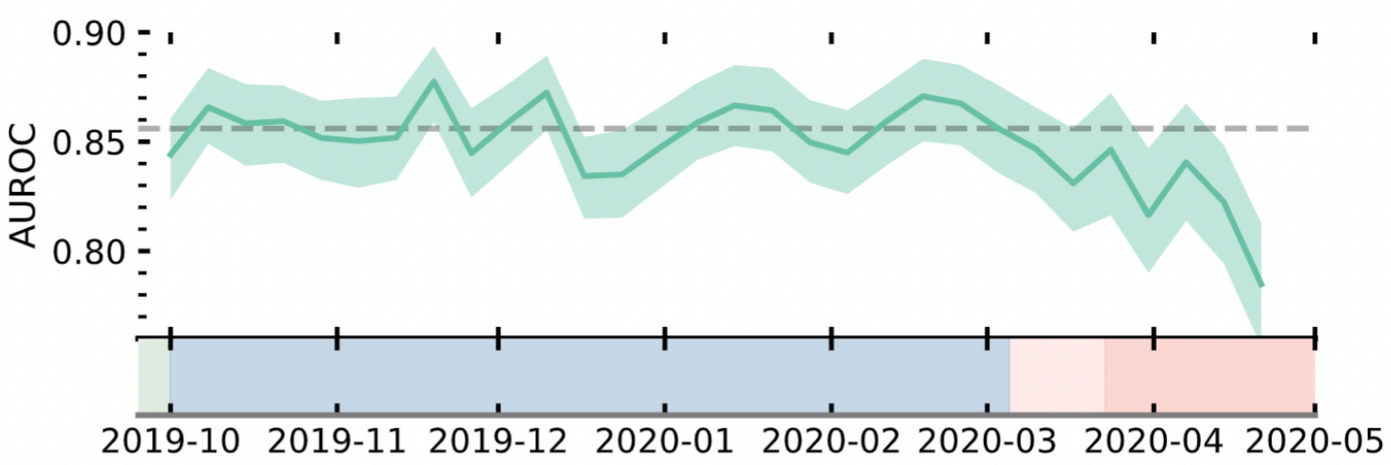
## Calibration drift in regression and machine learning models for acute kidney injury

Sharon E Davis,[1] Thomas A Lasko,[1] Guanhua Chen,[2] Edward D Siew,[3,4] Michael E Matheny[1,2,3,5]




## Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19

Christopher Duckworth[1]✉, Francis P. Chmiel[1], Dan K. Burns[1], Zlatko D. Zlatev[1], Neil M. White[1], Thomas W. V. Daniels[2,3], Michael Kiuber[4] & Michael J. Boniface[1]

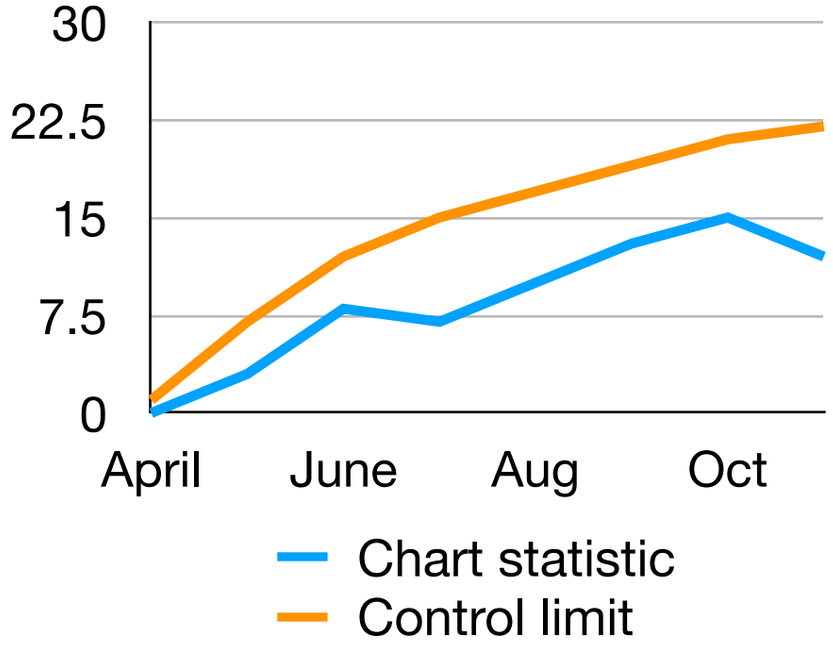


# Tools for monitoring machine learning models

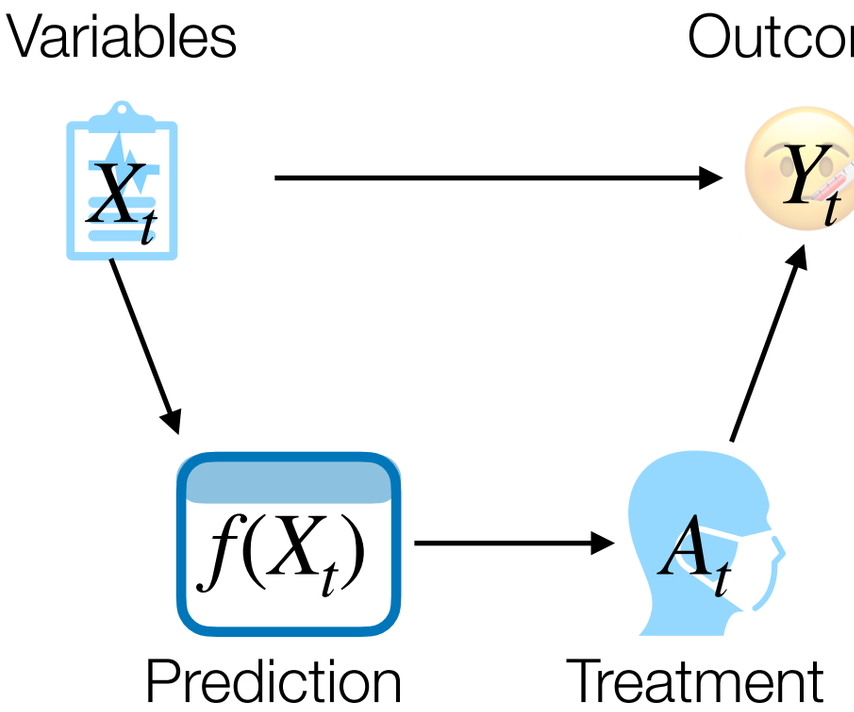
**Table 1.** Methods from statistical process control (SPC) and their application to monitoring ML algorithms.

| Method(s) | What the method(s) detect and assumptions | Example uses |
|---|---|---|
| CUSUM, EWMA | Detects a shift in the mean of a single variable, given shift size. Assumes the pre-shift mean and variance are known. Extensions can monitor changes in the variance. | • Monitoring changes in individual input variables |
| | | • Monitoring changes in real-valued performance metrics (e.g. monitoring the prediction error) |
| MCUSUM, MEWMA, Hotelling's $T^2$ | Monitor changes in the relationship between multiple variables | • Monitoring changes in the relationship between input variables |
| Generalized likelihood ratio test (GLRT), Online change point detection | Detects if a change occurred in a data distribution and when. Can be applied if characteristics of the pre- and/or post-shift distributions are unknown. GLRT methods typically make parametric assumptions. Parametric and nonparametric variants exist for online change point detection methods. | • Detecting distributional shifts for individual or multiple input variables |
| | | • Detecting shifts in the conditional distribution of outcome *Y* given input variables |
| | | • Determining whether parametric model recalibration/revision is needed |
| Generalized fluctuation monitoring | Monitor changes to the residuals or gradient | • Detect when the average gradient of the training loss for a differentiable ML algorithm (e.g. neural network) differs from zero |

*Feng et al 2022*



# The problem of Confounding Medical Interventions (and why causality matters)




1. Alert! Patient is at high risk of developing an adverse event

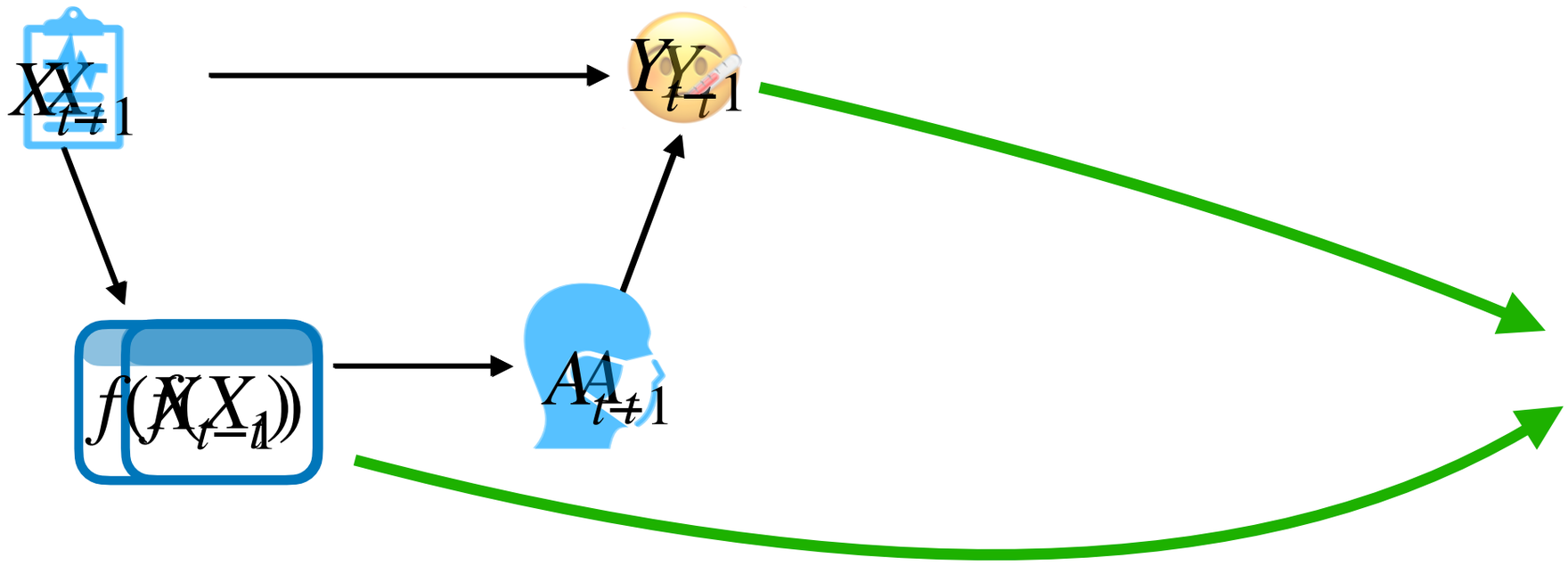
2. Administer prophylactic treatment

3. Patient doesn't develop the adverse event

*Was the model wrong or did the treatment make a difference?*

*Dyagilev and Saria 2016*

# The problem of Confounding Medical Interventions (and why causality matters)



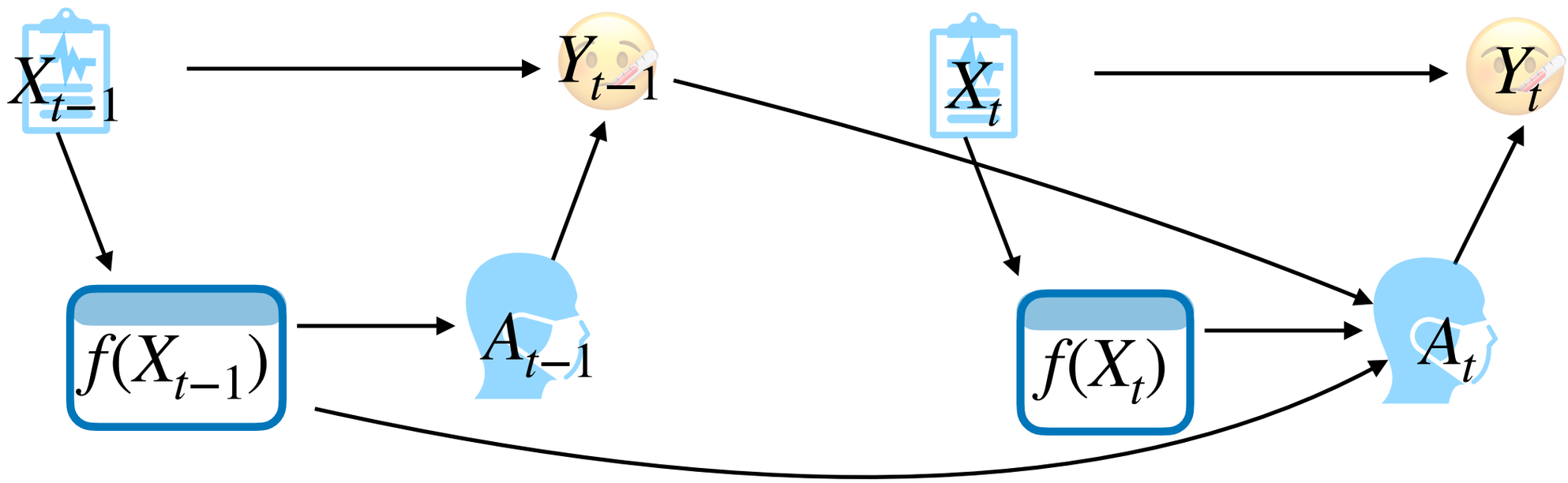

- Moreover, we need to think about if/how treatment propensities vary over time as clinicians interact with the ML-based clinical decision support system.

# Bringing in ideas from causal inference

- Key questions to answer:
  - What is the target of inference?
  - What types of bias arise in this situation?
  - What are the confounders in this problem? What is the adjustment set?

- For instance, we may assume conditional exchangeability, i.e. no unmeasured confounding:

  *A clinician's propensity to treat patient $X_t$ only depends on their prediction $f(X_t)$ and the clinician's past experiences interacting with the ML algorithm.*

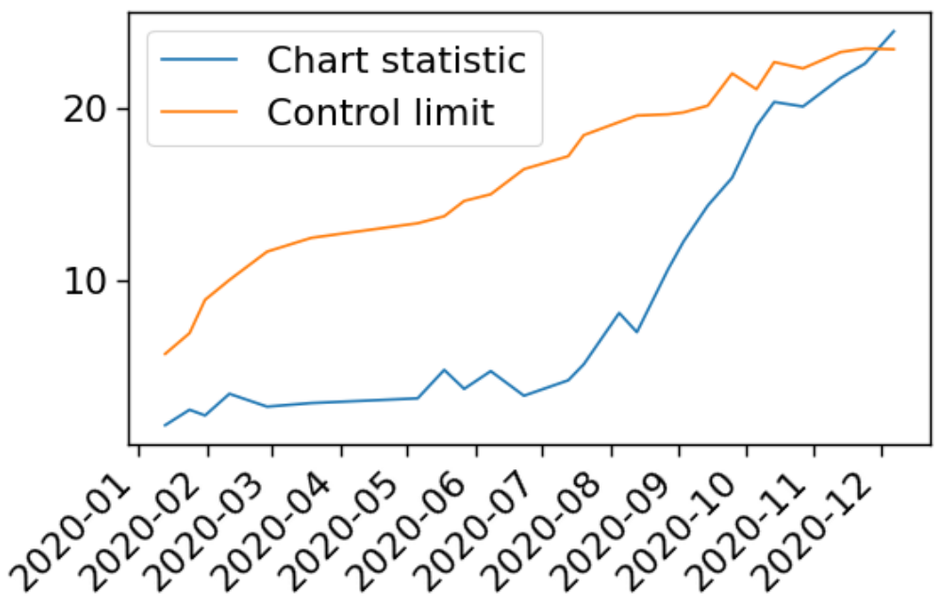


# Case study: Post-operative Nausea and Vomiting (PONV)

- Data: UCSF Multicenter Perioperative Outcomes Group (MPOG)
- ML algorithm: Random Forest using sex, smoking status, American Society of Anesthesiologists (ASA) classification, …

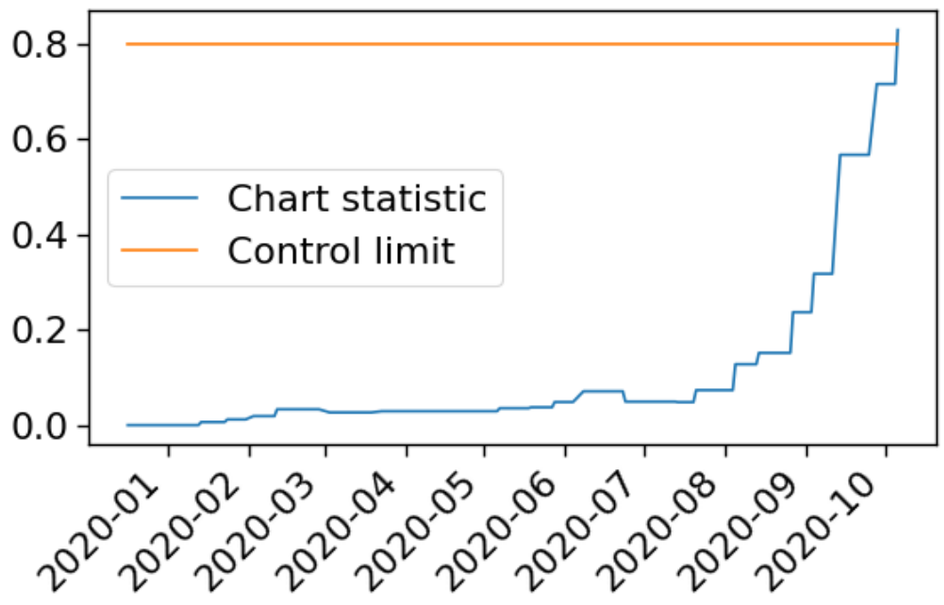
**Frequentist**

Score-based CUSUM




**Bayesian**

Bayesian changepoint monitoring



# Other opportunities at the intersection

ML for Epidemiology:

Interpretable ML

Using ML to unlock new data modalities, e.g. images, videos, audio, free text

Nonparametric treatment effect estimation

Heterogeneous treatment effects

Epidemiology for ML:

Surveillance of ML

Embedding causal reasoning into ML algorithms

Transportable ML algorithms

# Thanks!

Noah Simon

Brian Williamson

Romain Pirracchio

Leo Celi

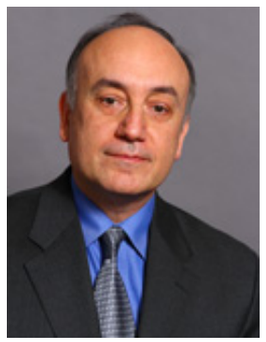
Rachael Phillips

Ivana Malenica

Andrew Bishara

Alan Hubbard

Alexej Gossmann



Berkman Sahiner

Nicholas Petrick

Gene Pennello

Support from the UCSF-Stanford CERSI program

*(Disclaimer: The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.)*