CARDIOLOGY 2024

DATA COLLECTION AND MACHINE LEARNING TO IMPROVE PERIOPERATIVE CARE IN THE PEDIATRIC POPULATION

J. “Nick” Pratap MB BCHIR, MRCPCH, FRCA

February 15, 2024

Children's Hospital of Philadelphia
Cardiac Center

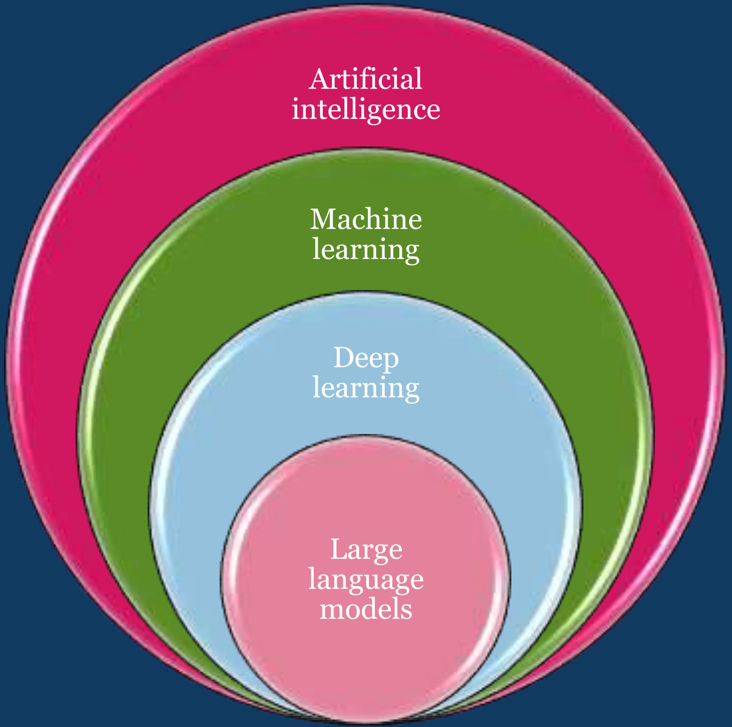
DISCLOSURES

- AHRQ R21 HS024983 (PI Pratap)
    - "Using the electronic health record to identify children likely to suffer last-minute surgery cancellation"

GOALS

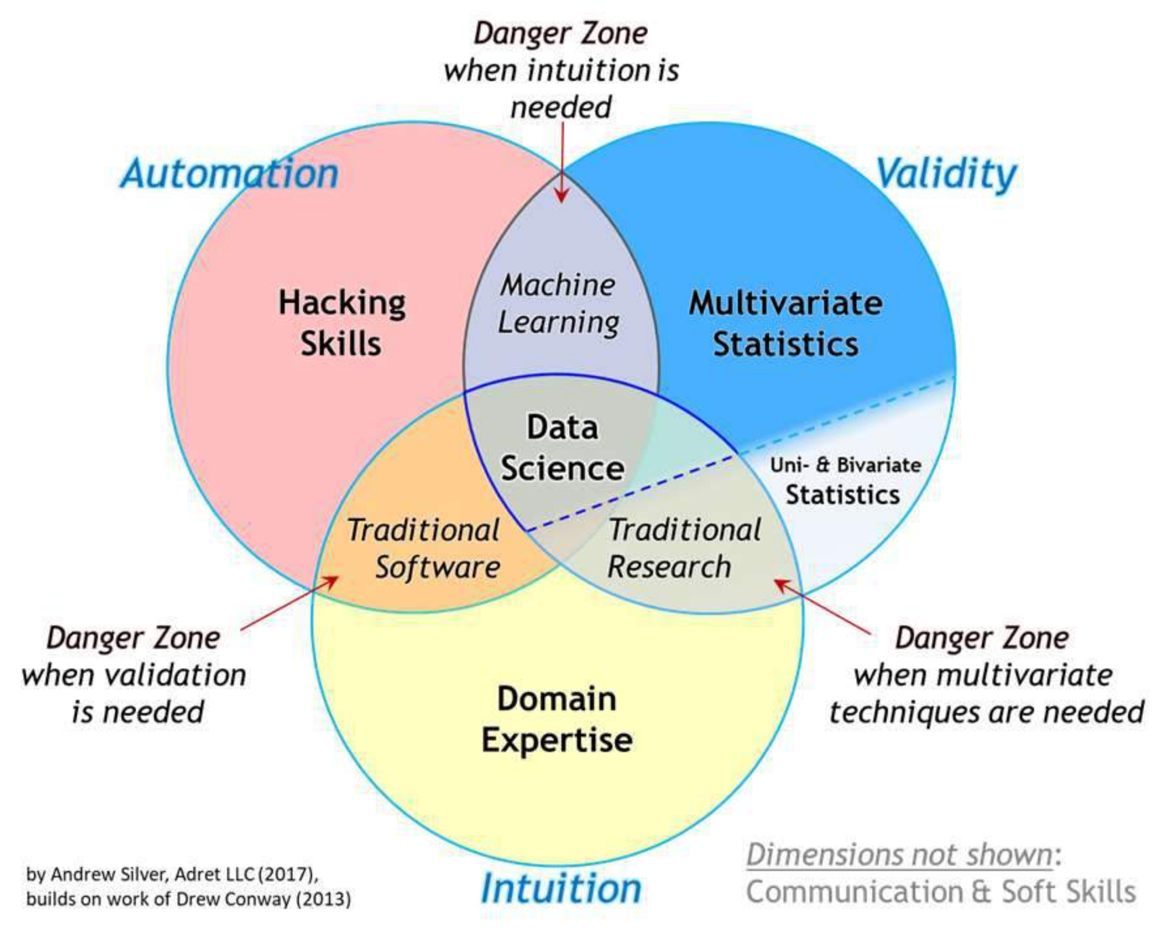
Understand	Fundamental machine learning (ML) concepts
Appreciate	Role of ML in perioperative outcomes research
Develop	Competence to critique research applying ML to perioperative care of congenital heart disease patients

WHAT IS ARTIFICIAL INTELLIGENCE?

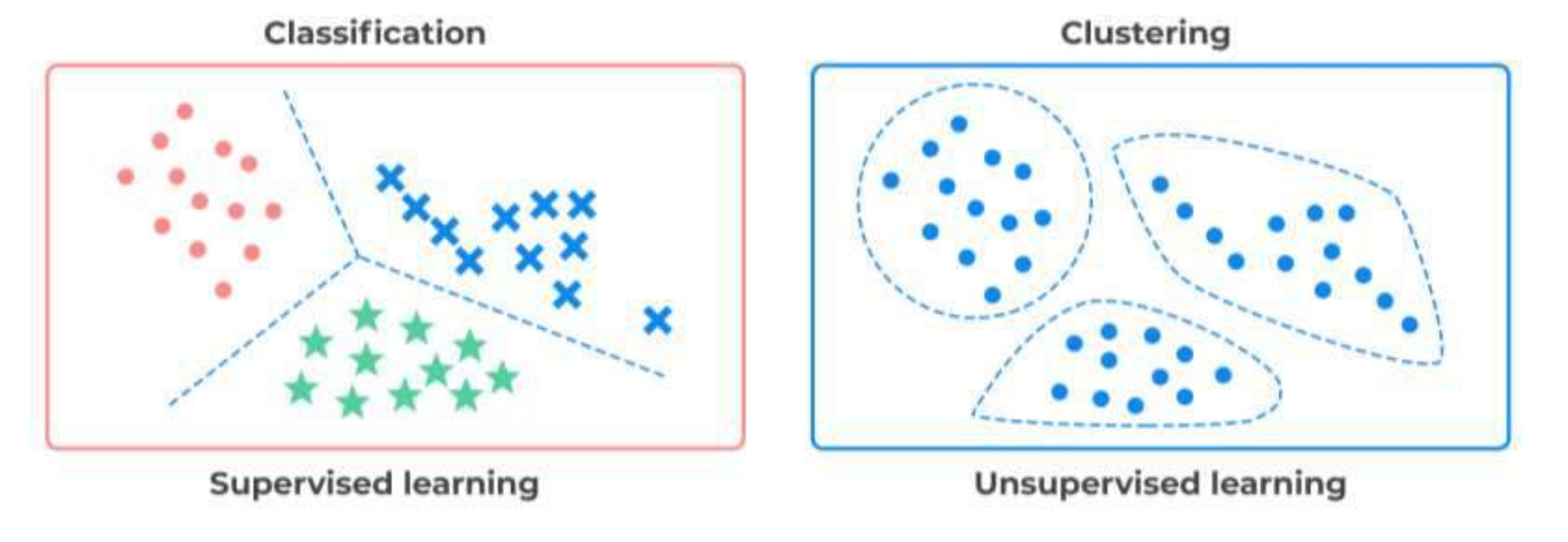


- **Artificial intelligence** refers to the general *ability of computers to emulate* human thought and *perform tasks in real-world environments.*

- **Machine learning** refers to the *technologies and algorithms* that enable systems to *identify patterns, make decisions, and improve themselves through experience and data.*


Danger Zone
when intuition is
needed
Automation
Validity
Hacking
Skills
Machine
Learning
Multivariate
Statistics
Data
Science
Uni- & Bivariate
Statistics
Traditional
Software
Traditional
Research
Danger Zone
when validation
is needed
Danger Zone
when multivariate
techniques are needed
Domain
Expertise
Intuition
by Andrew Silver, Adret LLC (2017),
builds on work of Drew Conway (2013)
Dimensions not shown:
Communication & Soft Skills

SUPERVISED VS UNSUPERVISED MACHINE LEARNING



MACHINE LEARNING APPROACHES

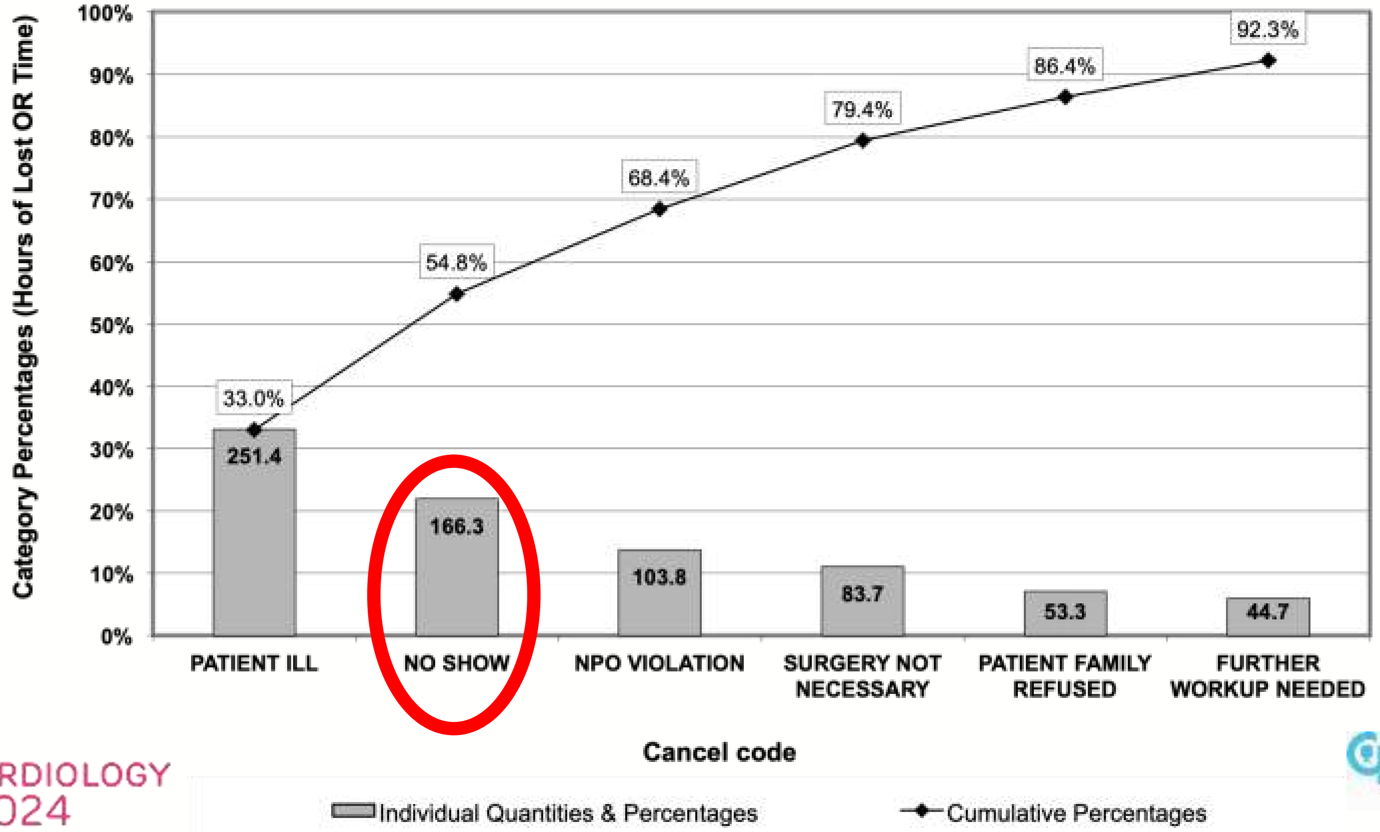
- Supervised learning
  - start with data where we know outcome (“labeled data”)
  - algorithms “learn” how to predict “unlabeled” data
  - e.g. risk assessment, prediction, fraud detection, image recognition
- Unsupervised learning
  - hoping to uncover previously unrecognized patterns and insights in unlabeled data
  - e.g. recommendation engines, customer segmentation, image segmentation
- Reinforcement learning
  - trains algorithms using a system of ‘reward’ and ‘punishment’, such that an agent can takes actions in a future specific environment to reach a predetermined goal
- Semi-supervised
  - Combination approach, includes generative adversarial networks (GANs)

TYPES OF SUPERVISED ML

- Classification
  - Binary outcome categories, e.g. yes vs. no
  - Multiple outcome categories
- Regression
  - Numeric output

- *Importance of a **specific** and **meaningful and/or recognized** target of prediction (outcome)*
  - *What do you hope to accomplish with ML model?*

EXAMPLE: SURGERY CANCELLATION AT CINCINNATI CHILDREN'S HOSPITAL



CARDIOLOGY 2024

Children's Hospital of Philadelphia Cardiac Center

IS THERE REASON TO BELIEVE SURGERY CANCELLATION IS PREDICTABLE?

- Yes! Cancellation rates vary across subpopulations
  - Privately-insured white children 1.9%
  - African American Medicaid children 9.4%

- *Cancellation of children's surgery is a source of disparity in health care access*

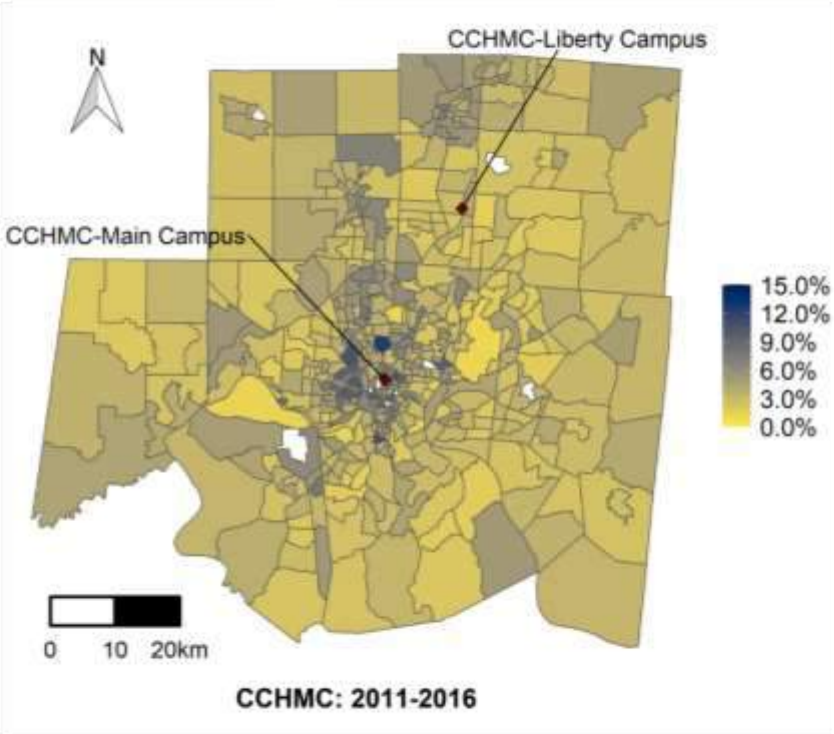
WHY MODEL SURGERY CANCELLATIONS? WHAT DO WE HOPE TO ACHIEVE?

- Insight into why cancellations happen
  - "Academic interest"
- Predict cancellation of ***individual*** cases
  - Prevention – targeting support to "at risk" families
    - Interventions informed by our research using qualitative interviewing
  - Mitigation of effects of cancellation on OR utilization and revenue
    - Identifies slots likely to be opened up for add-ons

SURGERY CANCELLATION AS A MACHINE LEARNING PROBLEM

- Supervised ML problem
  - Binary ***classification*** - case either "completed" or "cancelled on day"
  - Standardized definition of outcome and cohort (non-inpatients)
- Marked class imbalance
  - ~25 times more completed than cancelled cases
- Many variables suspected to have non-linear effects
- Intrinsically 'noisy' data related to cases re-booked after cancellation

GEOSPATIAL DISTRIBUTION OF CANCELLATION RATE



Outcome variable is ***cancellation rate*** for each census tract

Regression problem

AIMS OF ML MODELS

- Prediction
  - Typically choose best performing model
- Inference
  - Equivalent to hypothesis driven research
  - Some models easier to interpret than others
    - e.g. regularized regression-type models (ridge, lasso)
    - Challenge of "black boxes"
- *How much do we – and our patients – trust a model we cannot understand?*
  - Interpretable ML or "eXplainable AI" (XAI) e.g. iterative feature selection, Boruta algorithm, LIME, Shapley

OBTAINING DATA FOR ML

- *"Big data"*
  - Lots of data needed to find subtle nuances with confidence
  - Same human effort to code for large numbers

- Manual data collection
- EHR data
  - Messy, so needs careful attention to cleaning
  - Hard to access
- Billing data
  - Abundant, challenging to link between datasets, opaque (e.g. bundled ICU care)
- Registry data
  - Already cleaned up

DOWNSTREAM EFFECTS OF ERRONEOUS DATA

PREPARING DATA FOR ML

- Selection of variables as candidate predictors
  - Different schools of thought on selection – consider aims
    - Some (but not all) ML algorithms efficiently select for relevant variables
  - Inclusivity and equity issues
- Pre-processing - tedious but vitally important and ~90% of effort!
  - Cleaning
    - Inspect/visualize, detect errors, issue of “missingness”
  - Transformation
    - Statistical
      - Arithmetic numbers vs. ordered rank, non-linearity
    - Feature generation
      - Days of week, months of year, geocoding, linking to external data etc

SURGERY CANCELLATION DATASETS

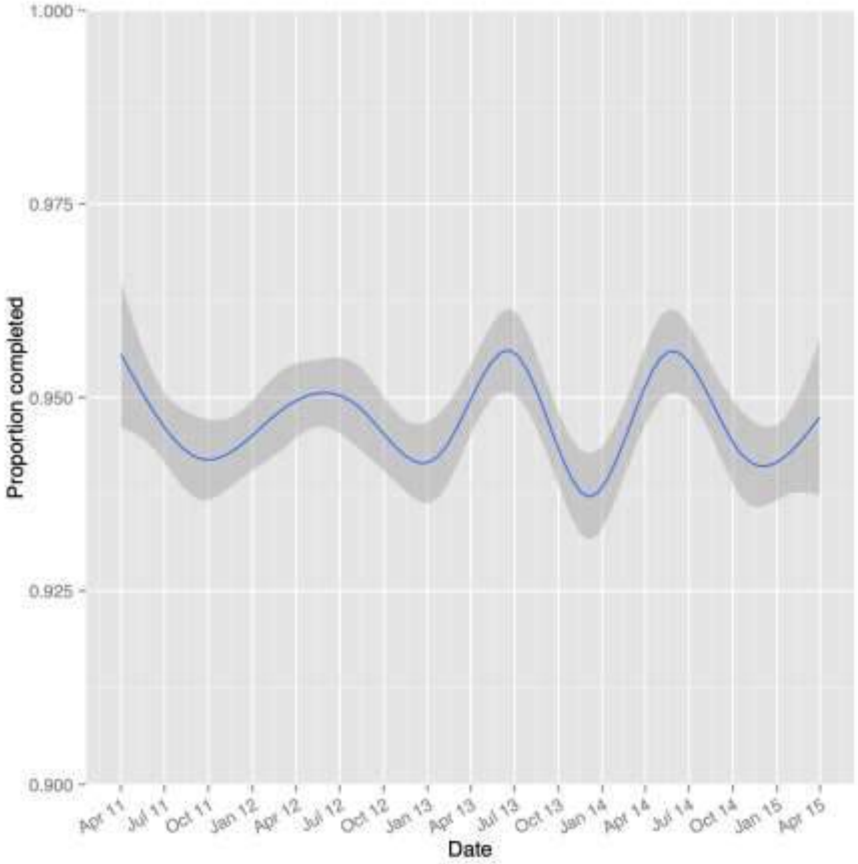
- 5 years of completed and cancelled surgeries
  - Removed rescheduled cases to reduce 'noise'
- Primary (Burnet) and secondary (Liberty) campuses of same children's hospital
  - Different range of procedures – less invasive/complex at Liberty
  - Some differences in patient population
    - Socioeconomic
    - No medically complex patients at Liberty

	Number of surgeries	***Number of Cancellations***
Main campus	84,615	3,088 (3.6%)
Liberty campus	43,162	1,940 (4.5%)

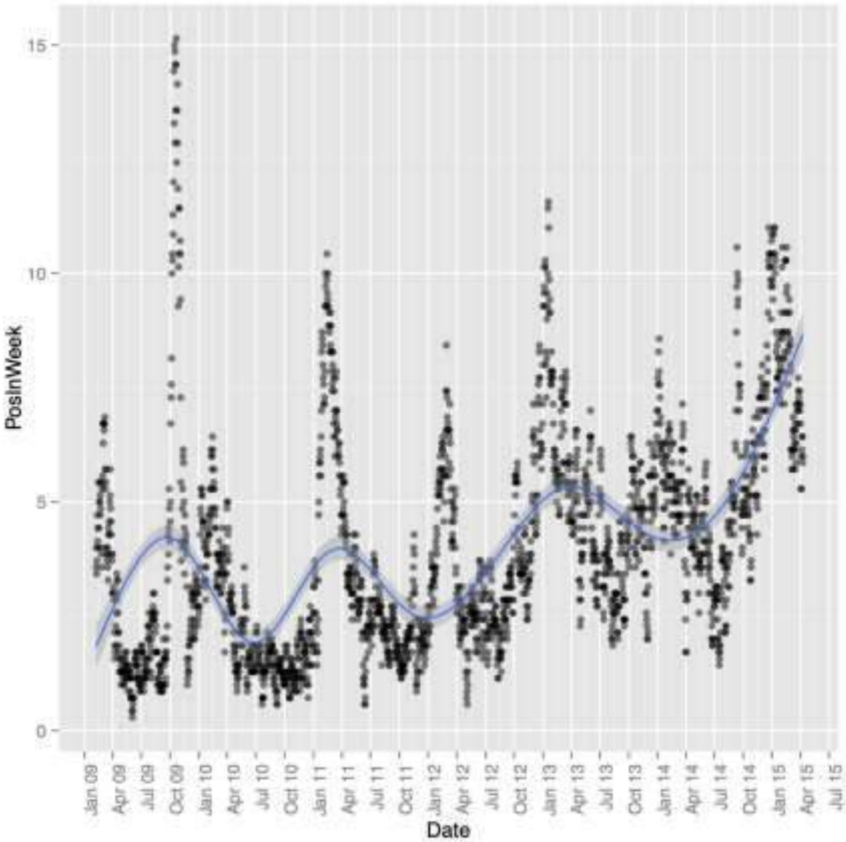
AVAILABLE EHR VARIABLES

Category	*Number of variables*	*Description*
Demographics	5	Patient age, sex, race, ethnicity, distance of home from hospital
Insurance info	2	Payer, payer type
Pre-op phone call	5	Number of call attempts, 'live' contact reached, first and final contacts, history & physical completed
Recent health care use	7	Number of medications taken as outpatient before surgery, recent ER attendance (4 timepoints), office visits, hospitalizations in 6 months
Prior cancellation behaviors	5	Numbers of previous cancellations, previous "no shows," previous other cancellations, clinic "no shows," previous surgeries
Scheduling/surgery related factors	9	Hour, day of week and month of surgery, lead time, "work in" case, surgical specialty, estimated case length, post-op disposition, time since original QI project
Weather	24	Detailed daily weather records from nearby airport (NOAA)
Infection risk	1	Local circulating load of respiratory, gastrointestinal and other febrile pathogens

SURGERY CANCELLATION RATE EXHIBITS SEASONALITY



POSITIVE RESPIRATORY/GI/FEBRILE PATHOGEN TEST RESULTS BY DATE



CARDIOLOGY 2024

Children's Hospital of Philadelphia Cardiac Center

CANDIDATE AMERICAN COMMUNITY SURVEY VARIABLES

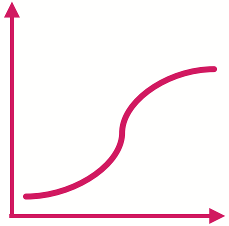
ACS table code	*Table description*	*Predictor*
B02001	Race	Black or African American race
B03003	Hispanic or Latino origin	Hispanic or Latino heritage
B17012	Poverty status of families by household type by number of related children aged <18 years	Families in poverty
B15002	Sex by educational attainment for the population aged ≥25 years	Population with low educational attainment
B16002	Language spoken at home and ability to speak English	Linguistic isolation
B06008	Place of birth by marital status in the United States	Adults never married
B08201	Household size by vehicles available	No car in household
B25003	Residential tenure	Rented houses
B25077	Median home value (US $)	Median home value
B19125	Median family income in the past 12 months by the presence of own children aged <18 years	Median household income
B25002	Residential occupancy status	Vacant houses
B25014	Tenure by occupants per room	Household overcrowding
B01003	Census tract total population	Total population

American Community Survey data from the US Census Bureau measured at the census tract level

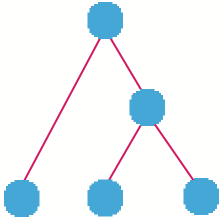
MACHINE LEARNING CLASSIFIERS

$$P(O|X)=\frac{P(X|O)P(O)}{P(X)}$$

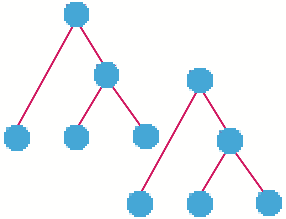
Naïve Bayes



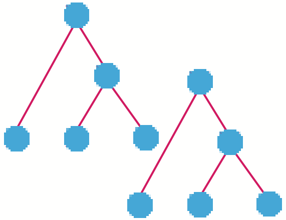
Logistic regression (LR) with L1 and L2 normalization



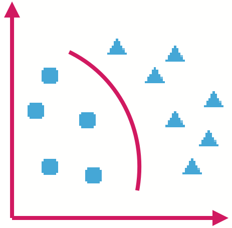
Decision tree



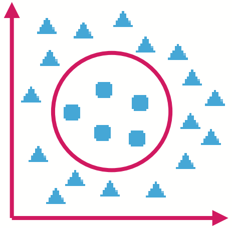
Random forest



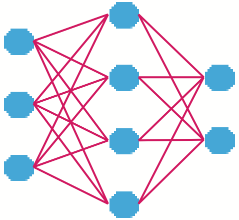
Gradient-boosted LR



Support vector machine (SVM) with polynomial kernel



SVM with radial basis function kernel



Artificial neural networks

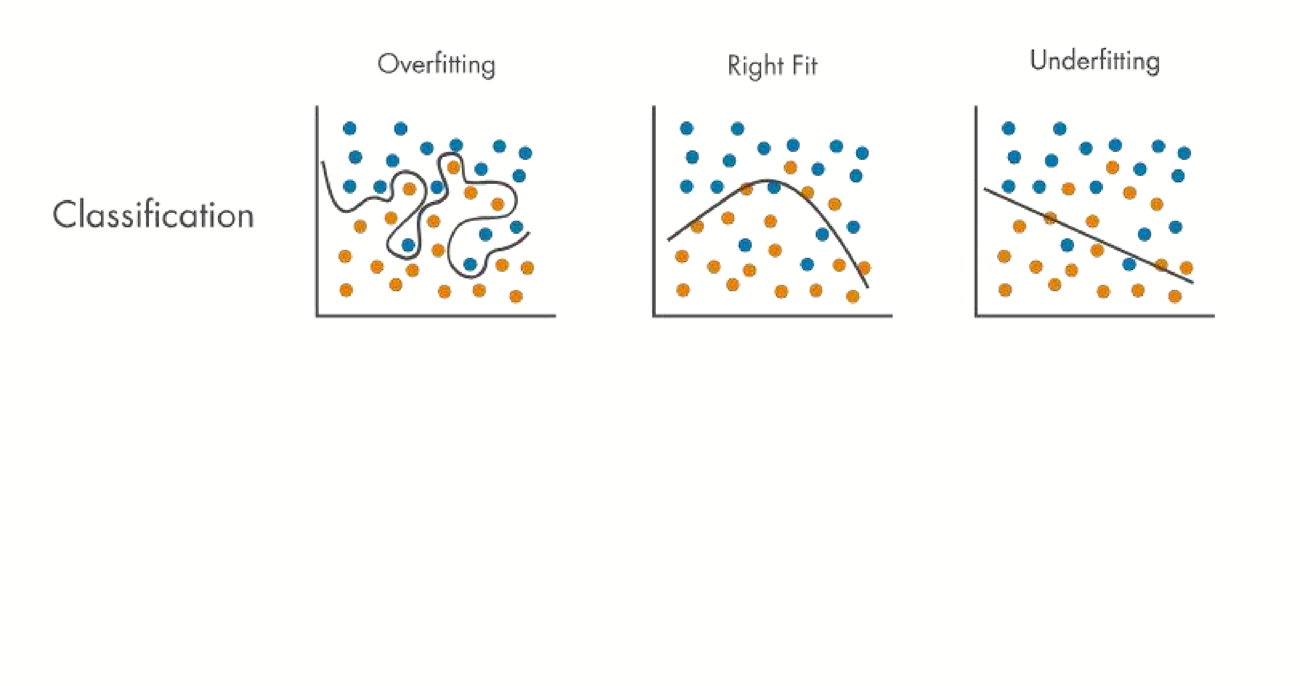
APPLYING ML ALGORITHMS

- Wide range of algorithms, including deep learning variants
  - Substantial computing power required for some
- Nature of candidate predictors may suggest most likely candidate algorithms
  - Typically, worth trying a substantial range of algorithms
- Most algorithms have tuning 'hyperparameters'
- Modifications to training for special circumstances may require tuning too
  - e.g. class imbalance, time series

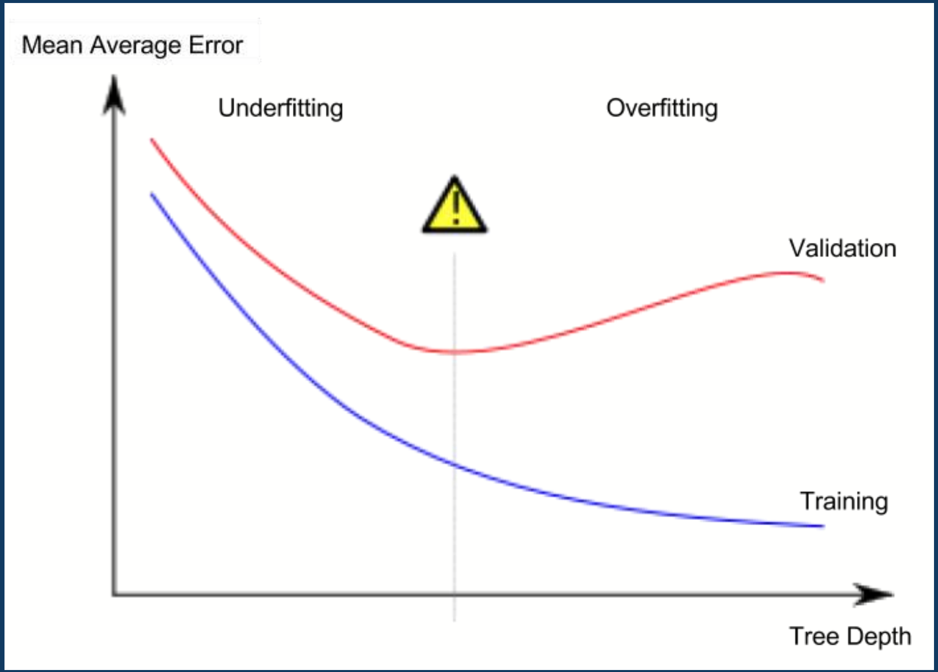
- Often ML algorithms are 'too good' – learn unhelpful irregularities in data

- *So how do you work out which ML model and combination of tuning paremeters is "best"?*

CHALLENGE OF "RIGHT" FIT IN ML



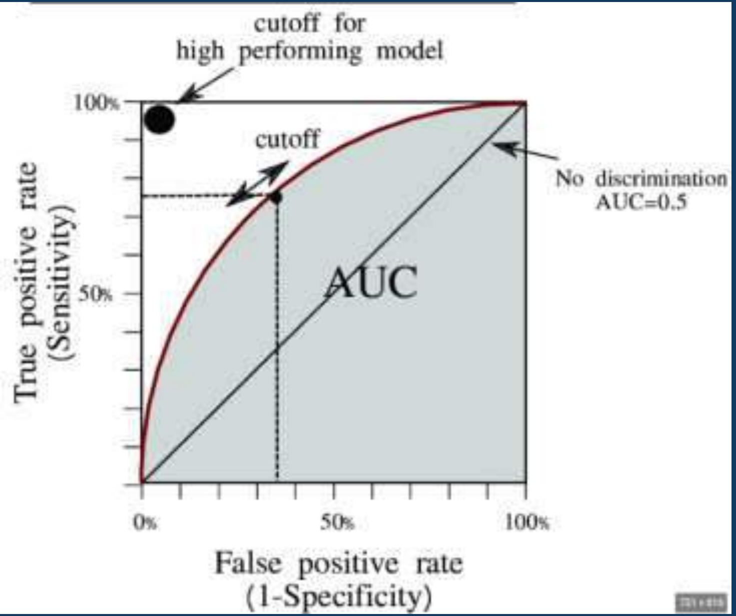
GENERALIZABILITY



MEASURING ML PERFORMANCE

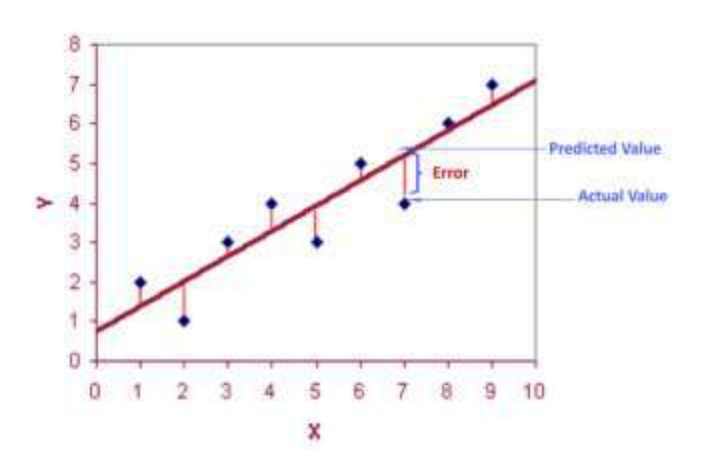
Classification

e.g. area under receiver-operator characteristic curve (AUROC or ROCAUC)



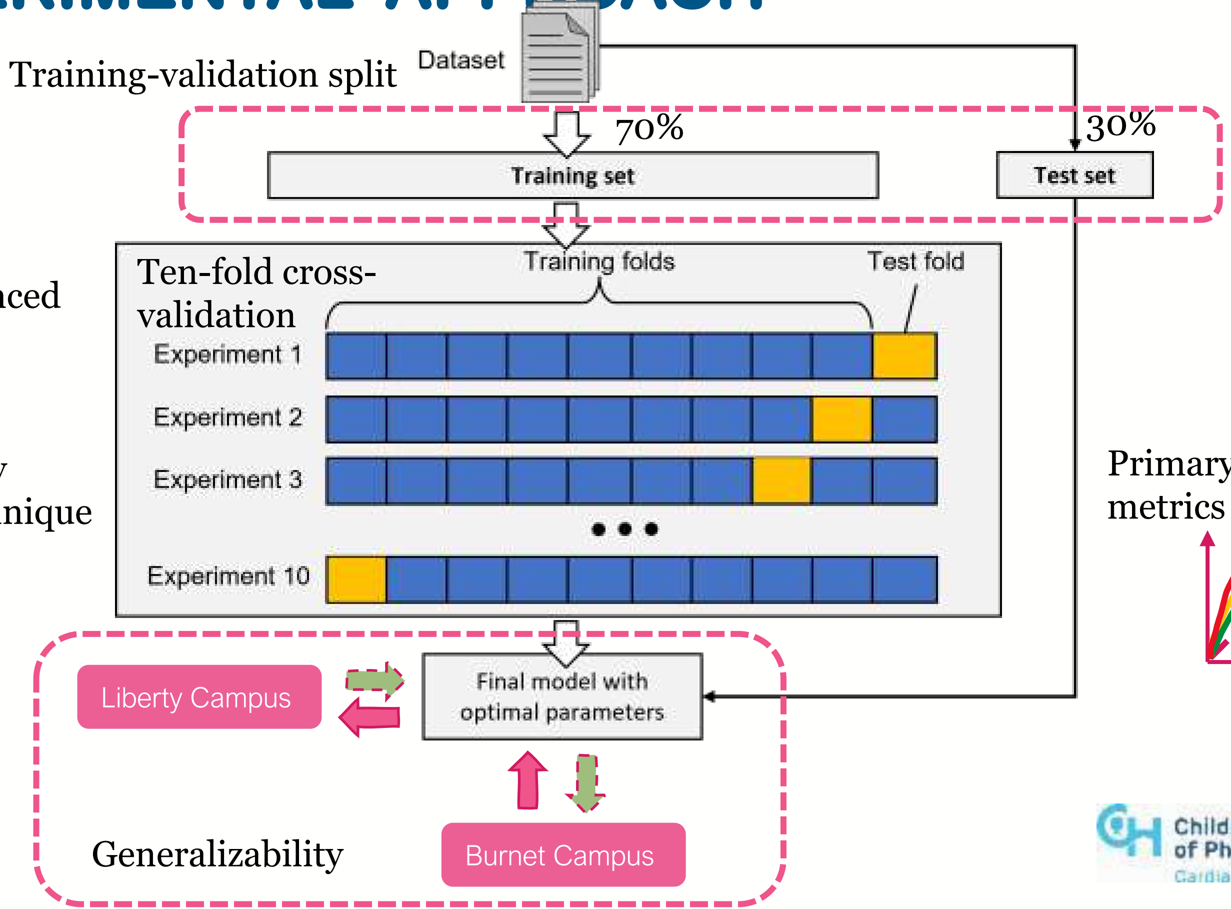
Regression

e.g. root mean squared error (RMSE)



$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2}$$

ML 'EXPERIMENTAL' APPROACH



Dealing with imbalanced data

- Up-sampling
- Down-sampling
- Synthetic minority oversampling technique (SMOTE)

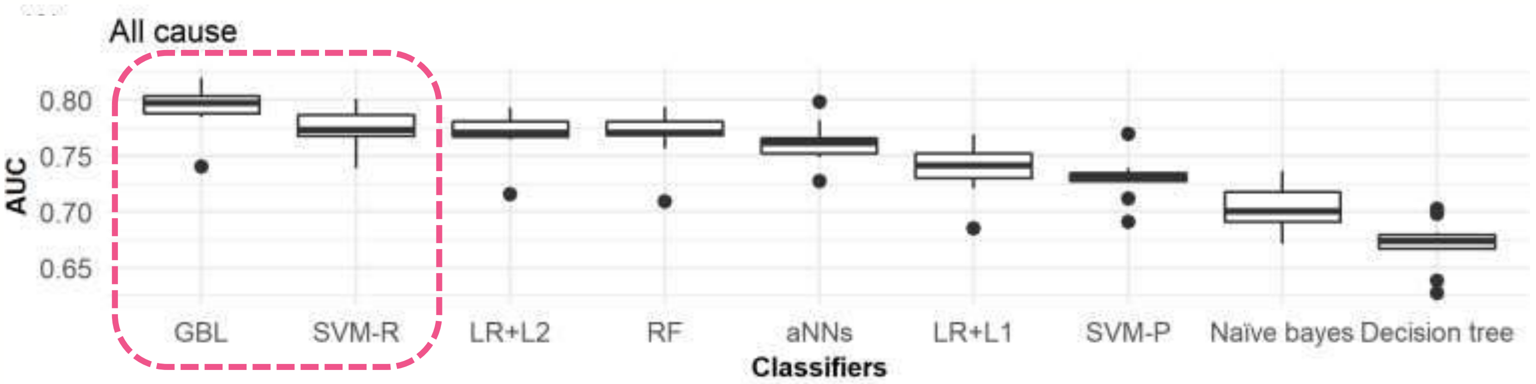
Primary evaluation metrics

CARDIOLOGY 2024

Children's Hospital of Philadelphia Cardiac Center

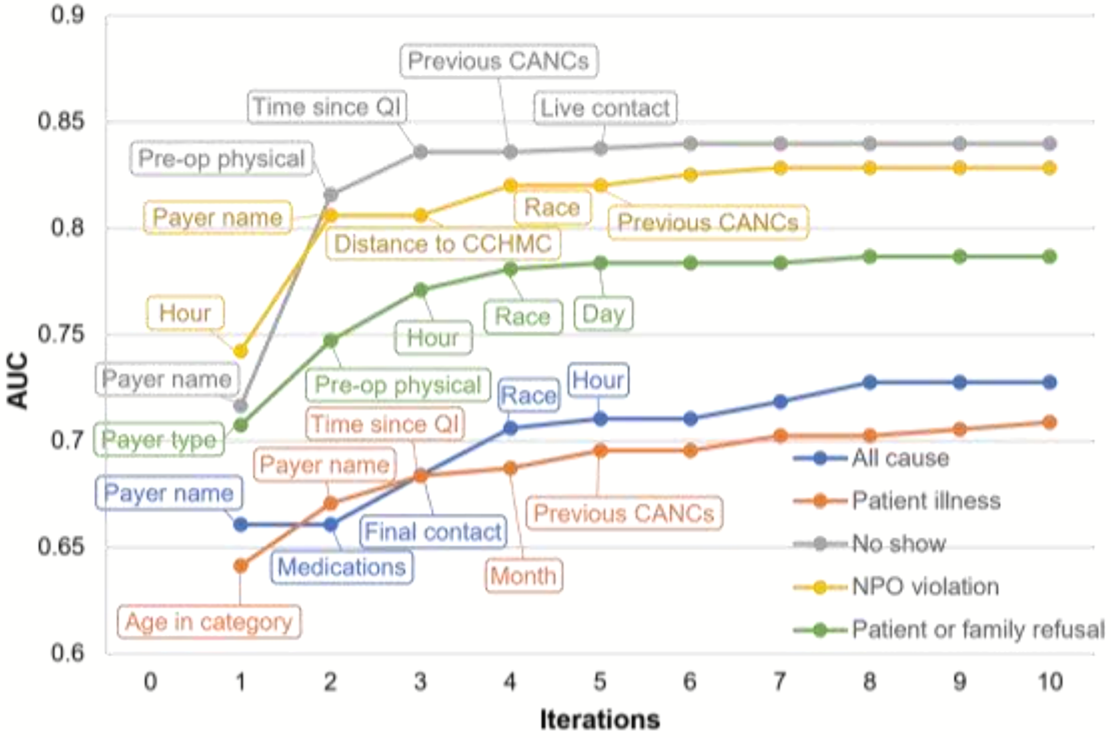
PERFORMANCE OF ALL-CAUSE SURGERY CANCELLATION MODELS

Burnet campus dataset



Gradient-boosted logistic regression (GBL) performs best – SVM and LR+L2 similar

INFERENCE: MOST IMPORTANT FEATURES FOR PREDICTING SURGERY CANCELLATION



Cross-validation performance of iterative step-forward feature selection (L1-normalized logistic regression)

The 'top'* five variables yielded more than 95% of performance gain in feature selection

** = adding most predictive power to model*

CARDIOLOGY 2024

Children's Hospital of Philadelphia Cardiac Center

EXAMPLE: ACUTE KIDNEY INJURY FOLLOWING PEDIATRIC CARDIAC SURGERY

- Common problem - 27% of all PICU admissions in AWARE study
  - Challenge of standardized definition e.g. KDIGO
  - Challenge of setting threshold – any AKI? severe AKI?
- Associated with poor outcome
  - Odds ratio of 1.8 for death in AWARE (2.7 for cardiovascular)
- Predictable *in advance* - in retrospective cohorts ROC AUC > 0.95
  - Continuously updating prediction model also reported
- Existence of credible interventions
  - Avoidance of nephrotoxic drugs, targeting renal perfusion pressure, 'prophylactic' post-op peritoneal dialysis
- *Appropriate area for study of ML application*

CONCLUSIONS

- Data science approaches will likely shed light on complex problems in perioperative care of pediatric cardiac patients
  - Post-surgical AKI is likely area of early clinical application
- Machine learning is 'just' statistics on a massive scale
- Good data is fundamental and can be taken from many sources
- Important to understand pitfalls in ML and ways to avoid
  - *"Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)" statement 2015*