Machine Learning for Survival Analysis: Benefits and Drawbacks

Hieu “Hugh” Nguyen

PhD Candidate, Biomedical Engineering @JHU

PIs: Bharath Ambale-Venkatesh, Joao Lima, Eliseo Guallar

1/24/2020

Outline

- ML Introduction & Fun Facts
- ML for classification: overview & drawbacks for survival analysis
- ML methods for survival: overview & benefits
- Limitations of current ML survival methods

ML Evolution Over the Years

1950
Alan Turing created a test to check if a machine could fool a human being into believing it was talking to a machine.

1957
First neural network for computers (the perceptron) was invented by Frank Rosenblatt, which simulated the thought processes of the human brain.

1979
Students of Stanford University, California, invented the Stanford Cart which could navigate and avoid obstacles on its own.

2002
A software library for Machine Learning, named Torch is first released.

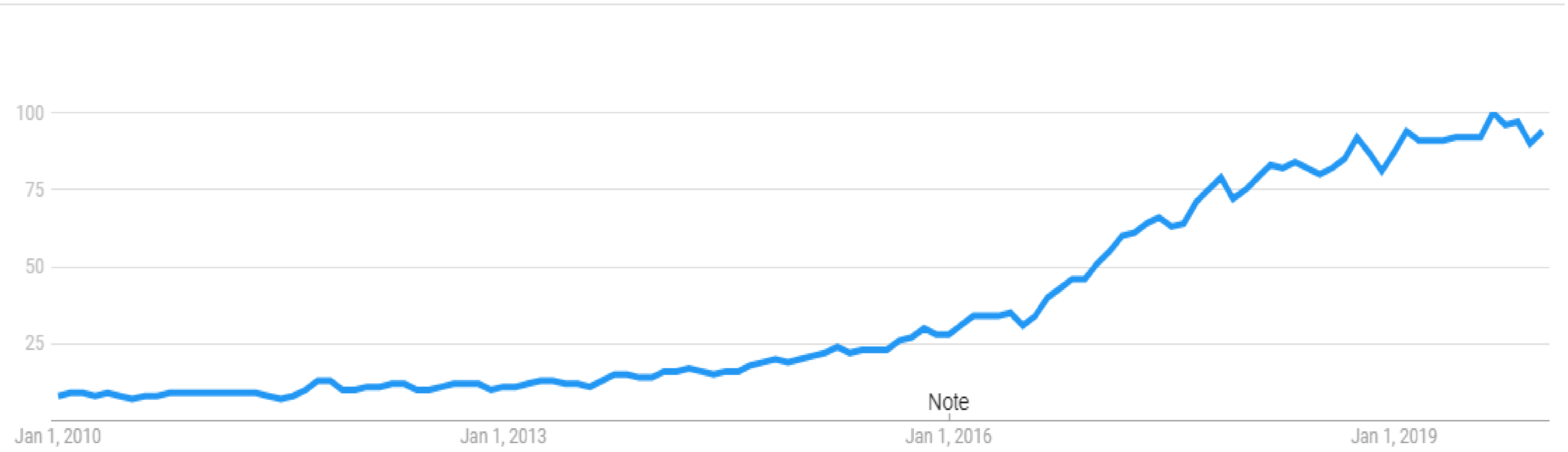
1952
The first computer learning program, a game of checkers, was written by Arthur Samuel.

1967
The Nearest Neighbor Algorithm was written.

1997
IBM's Deep Blue beats the world champion at Chess.

2016
AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Chinese Board Game Go competition.

ML Interest Over Time Since 2010 (Google Trends)



ML-related papers on arXiv from 1990-2018:

- 16,625 papers in total up to 2018

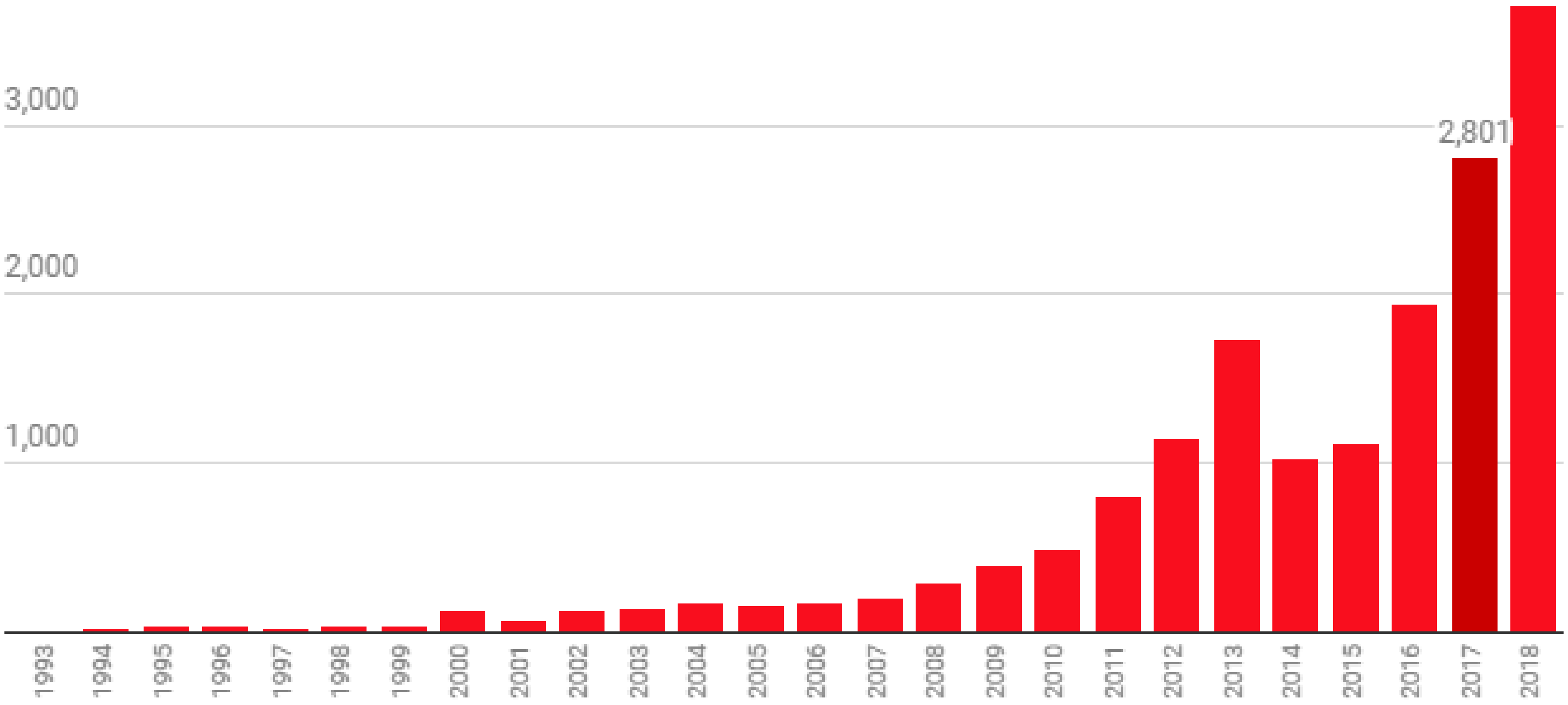
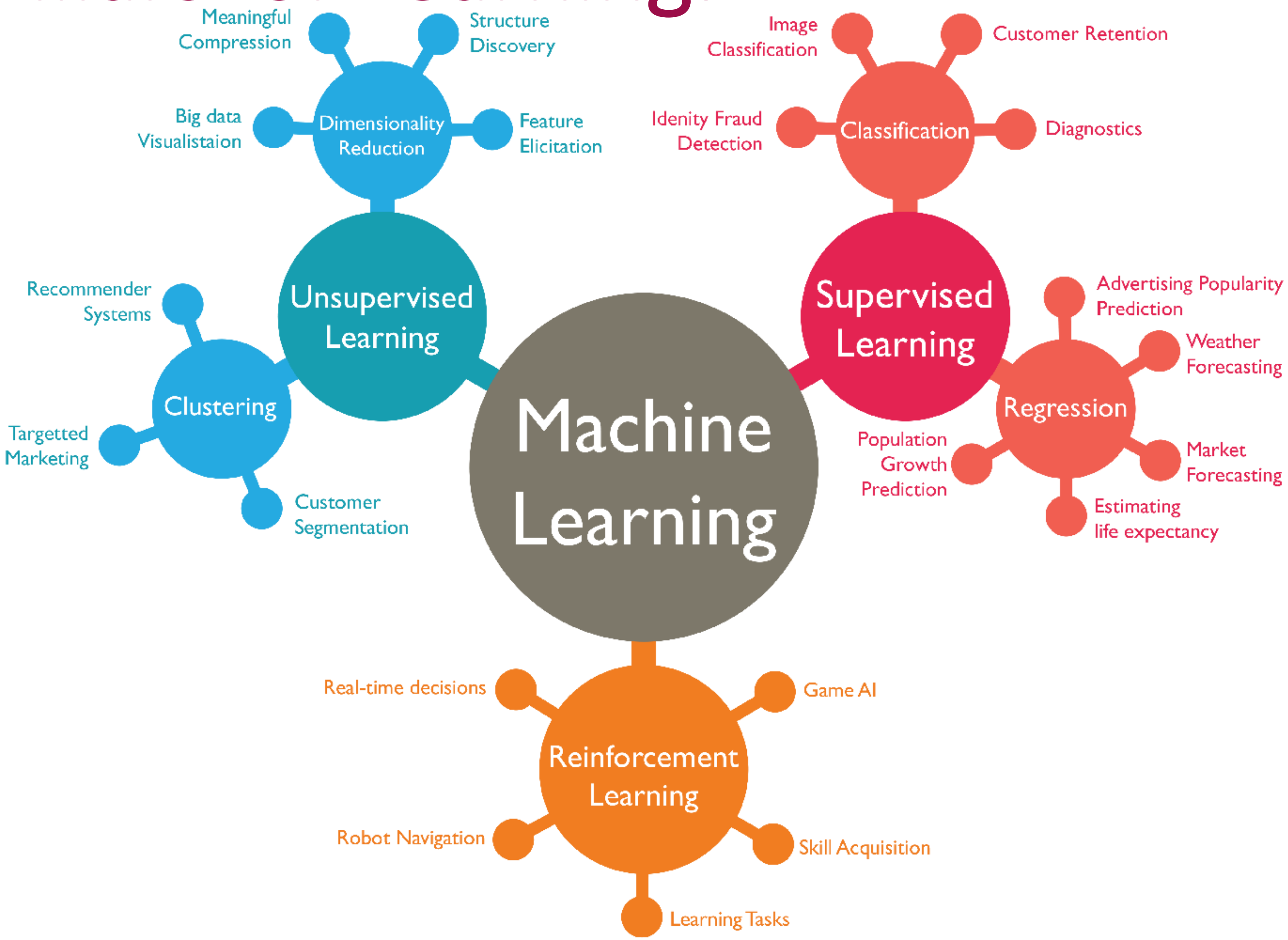


Chart: MIT Technology Review • Source: arXiv.org • Created with Datawrapper

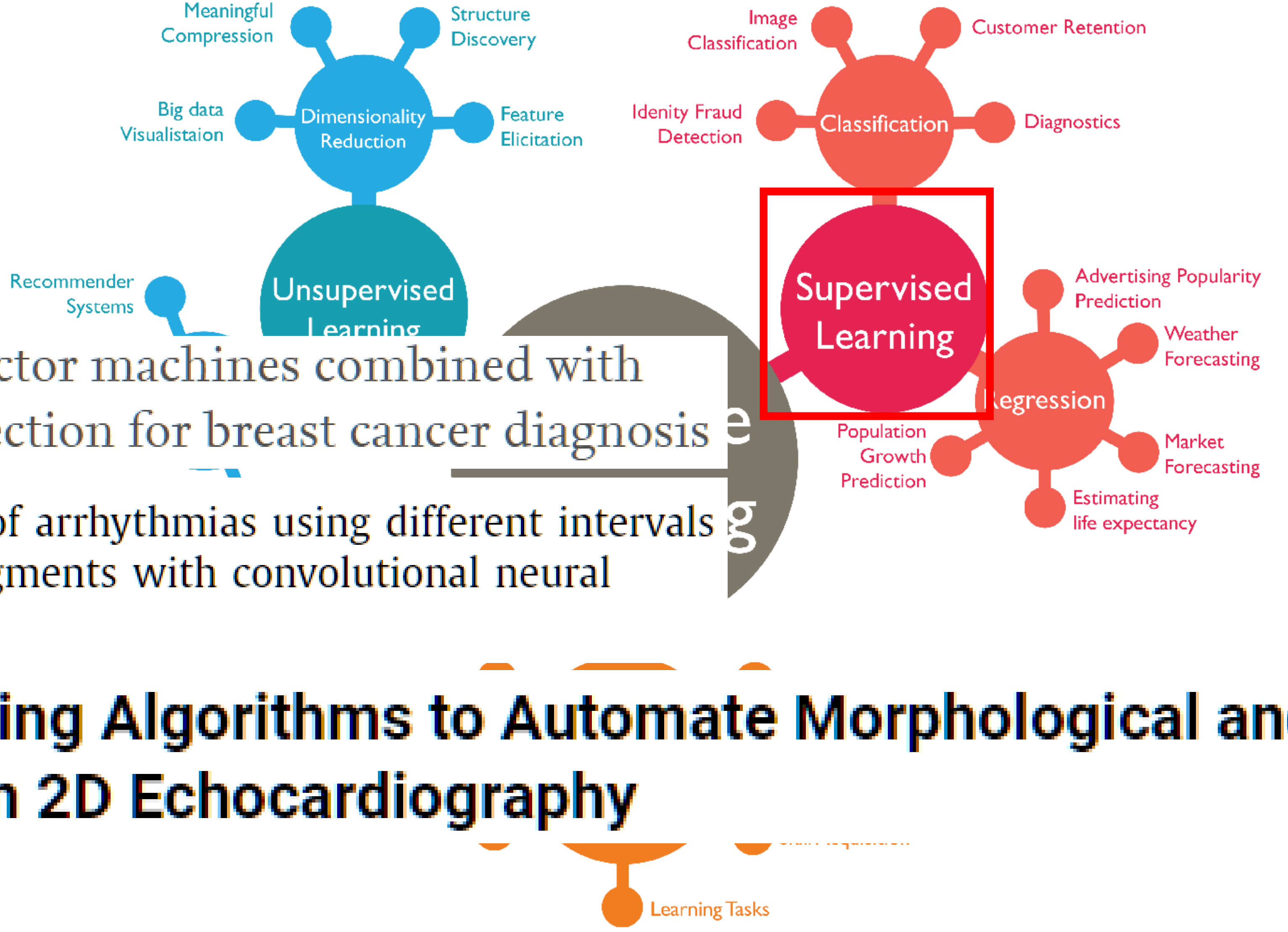
ML is now everywhere…

Three ‘Pillars’ of Learning:



https://www.slideshare.net/awahid/big-data-and-machine-learning-for-businesses

Supervised:

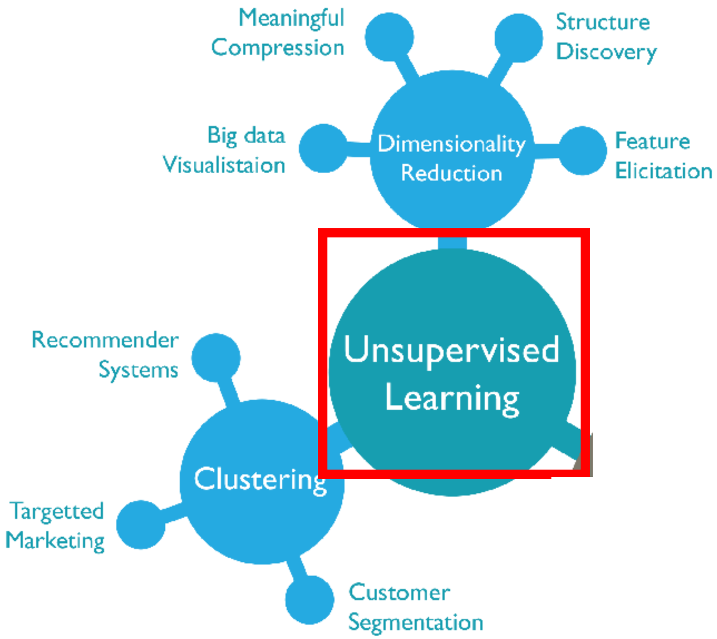


Support vector machines combined with feature selection for breast cancer diagnosis

Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network

Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography

Unsupervised:



SCIENTIFIC REPORTS

Article | Open Access | Published: 17 May 2016

Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis

THE LANCET
Respiratory Medicine

ARTICLES | VOLUME 5, ISSUE 10, P816-826, OCTOBER 01, 2017

Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study

Reinforcement:

ARTICLES
https://doi.org/10.1038/s41591-018-0213-5
nature medicine

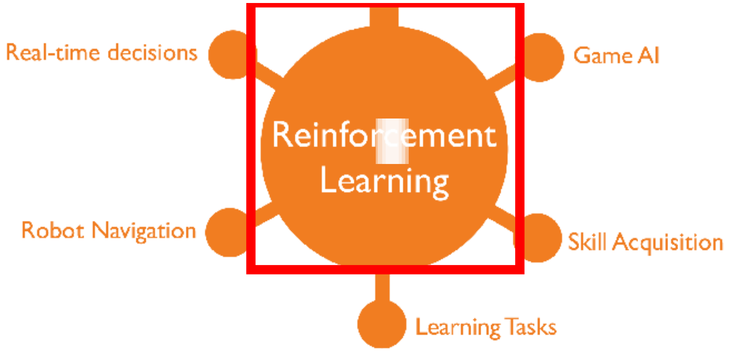
The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care

A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units

Supervised Reinforcement Learning with Recurrent Neural Network for Dynamic Treatment Recommendation

Big-data Clinical Trial Column

Reinforcement learning in clinical medicine: a method to optimize dynamic treatment regime over time



Outline

- ML Introduction & Fun Facts
- ML for classification: overview & limitations for survival analysis
- ML methods for survival: overview & their benefits
- Limitations of current ML survival methods

Supervised Learning Classifiers Used as Risk Prediction/Stratification Models

- <u>At a certain time point</u> (i.e: ICU discharge, 10-year follow-up), map y= f(x)
- Report AUC, sens, spec, PPV, NPV
- Exclude samples without a ‘label’

Classifiers as Risk Prediction?

- A lot of papers use classifiers for dynamic risk prediction (aka 'early warning model')
- Time-window approach
- Myriad packages and online resources for 'plug and play' ML classification and regression methods (i.e: caret, mlr, scikit-learn)

ORIGINAL

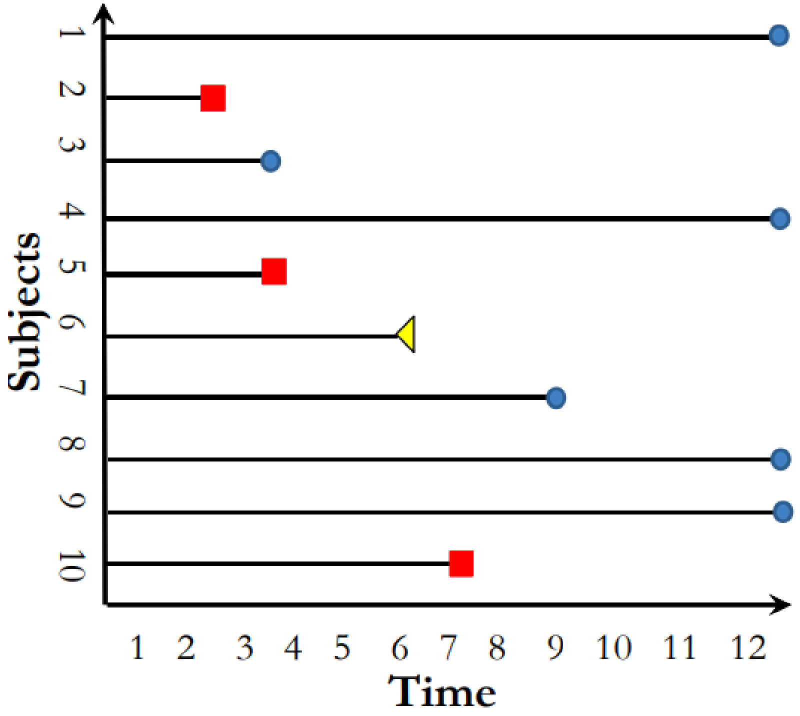
Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms

Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU

Research article | Open Access | Open Peer Review | Published: 29 December 2018

Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk

Are ML Classifiers or Regressors Suitable for Survival Analysis?



Ping Wang, Yan Li, Chandan, K. Reddy, "**Machine Learning for Survival Analysis: A Survey**". ACM Computing Surveys (under revision), 2017.

Classification Problem:

3 +ve and 7 -ve

Cannot predict the time of event

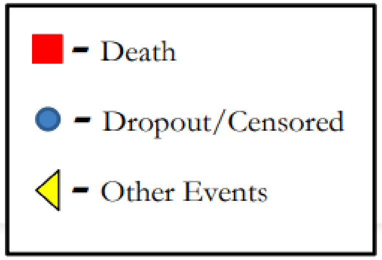
Need to re-train for each time

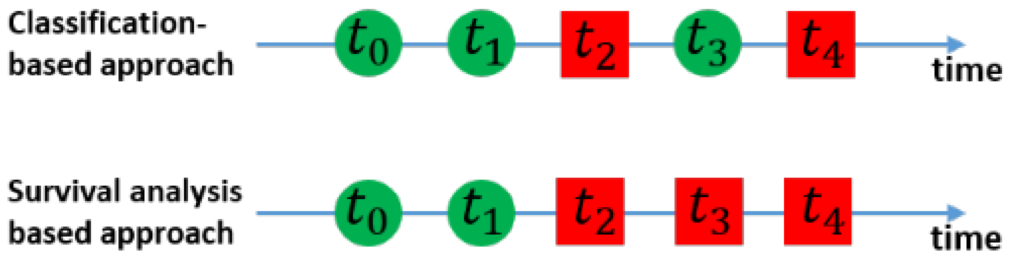
Regression Problem:

Can predict the time of event

Only 3 samples (not 10)

– loss of data





P. Zheng, S. Yuan, and X. Wu, "SAFE: A Neural Survival Analysis Model for Fraud Early Detection," 2018.

Outline

- ML Introduction & Fun Facts
- ML for classification: overview & limitations for survival analysis
- ML methods for survival: benefits & overview of each method
- Limitations of current ML survival methods

Motivations for the birth of ML survival:

- ML Classifiers are not good enough
- Traditional statistical survival methods are also not good enough with their assumptions:

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots + \beta_n x_{in})$$

1. Proportional hazards: constant hazard ratio over time
2. Each covariate contribute linearly to the model
3. Each covariate is independent of each other
4. Do not converge for high number of covariates

ML survival methods could overcome Cox's limitations:

1. Proportional hazards: constant hazard ratio over time → flexible
2. Each covariate contribute linearly to the model → allow for complex, non-linear relationships among covariates
3. Each covariate is independent of each other → less susceptible to correlated covariates
4. Do not converge for high number of covariates → can handle hundreds, thousands of covariates. Could be used as feature selection method or dimensionality reduction method

Overview of ML Survival Methods:

Most are adaptations of ML Classifiers

- Tree-based
- Boosting
- Support Vector Machine
- Bayesian-based
- Deep Learning/Neural Networks

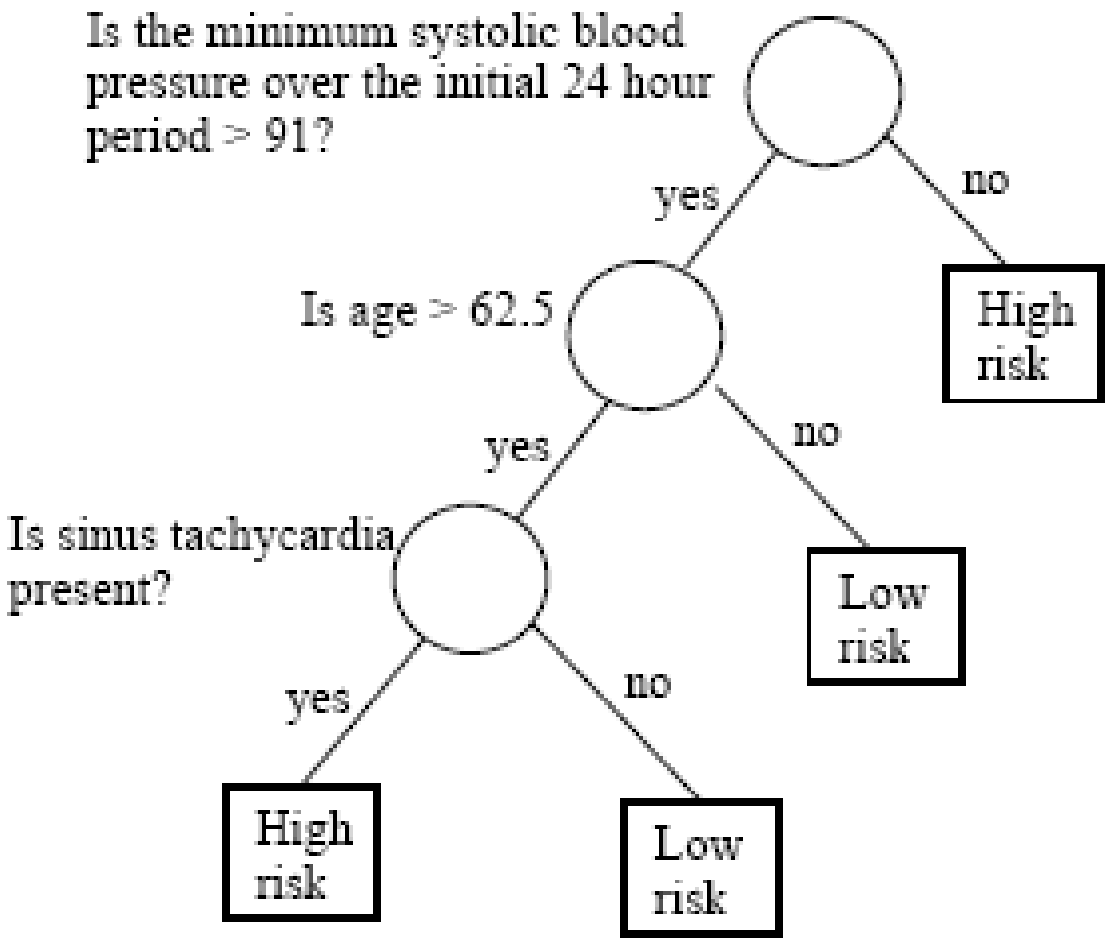
Overview of ML Survival Methods:

Most are adaptations of ML Classifiers

- Tree-based
- Boosting
- Support Vector Machine
- Bayesian-based
- Deep Learning/Neural Networks

Tree-based Survival Methods:

- Survival Tree (rpart)

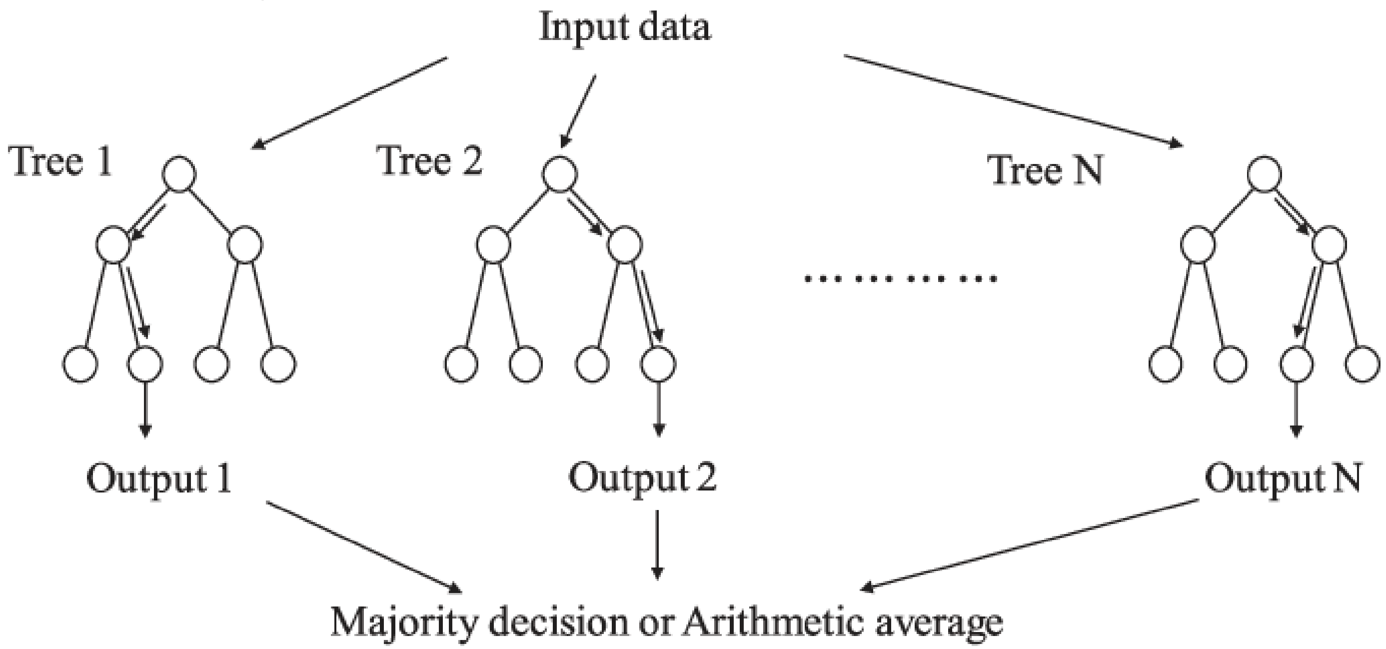


M. LeBlanc and J. Crowley, “Survival trees by goodness of split,” *J. Am. Stat. Assoc.*, 1993.

Tree-based Survival Methods:

Many trees together (majority vote – Yay to democracy!):

- Random Survival Forest (randomForestSRC)
- Conditional-Inference Forest
 (cForest)
- RF-SLAM

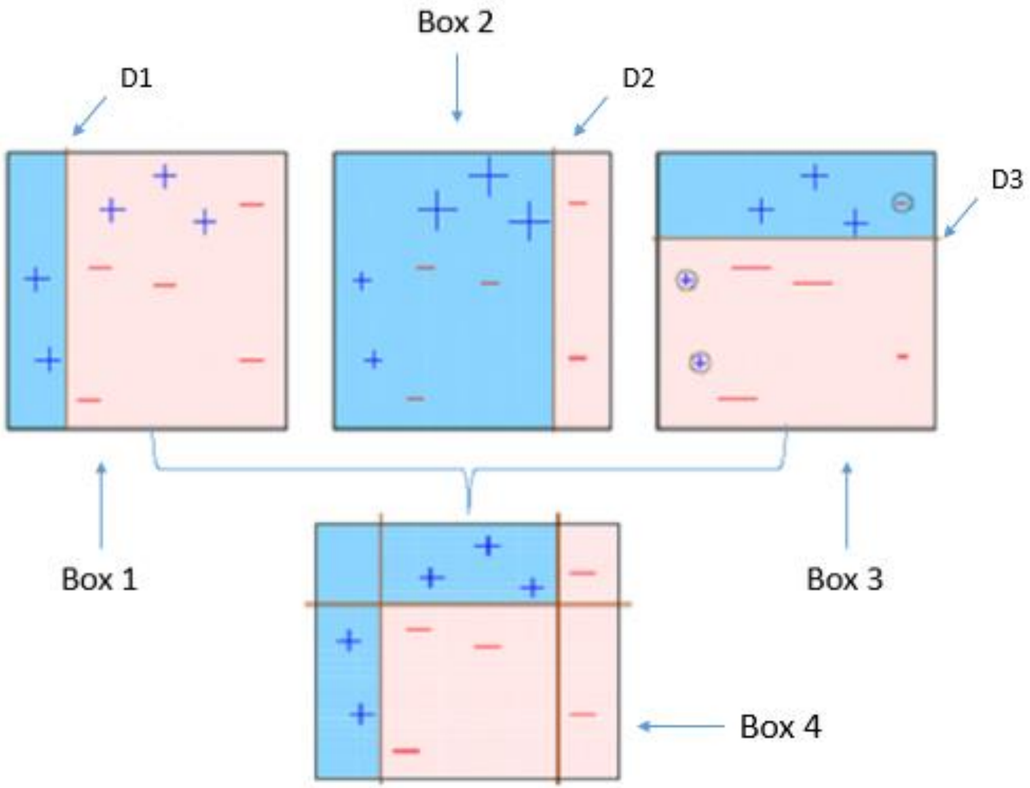


H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, “Random survival forests,” *Ann. Appl. Stat.*, 2008.
T. Hothorn, K. Hornik, and A. Zeileis, “Unbiased recursive partitioning: A conditional inference framework,” *J. Comput. Graph. Stat.*, 2006.
S. Wongvibulsin, K. C. Wu, and S. L. Zeger, “Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis,” *BMC Med. Res. Methodol.*, vol. 20, no. 1, pp. 1–14, 2019.

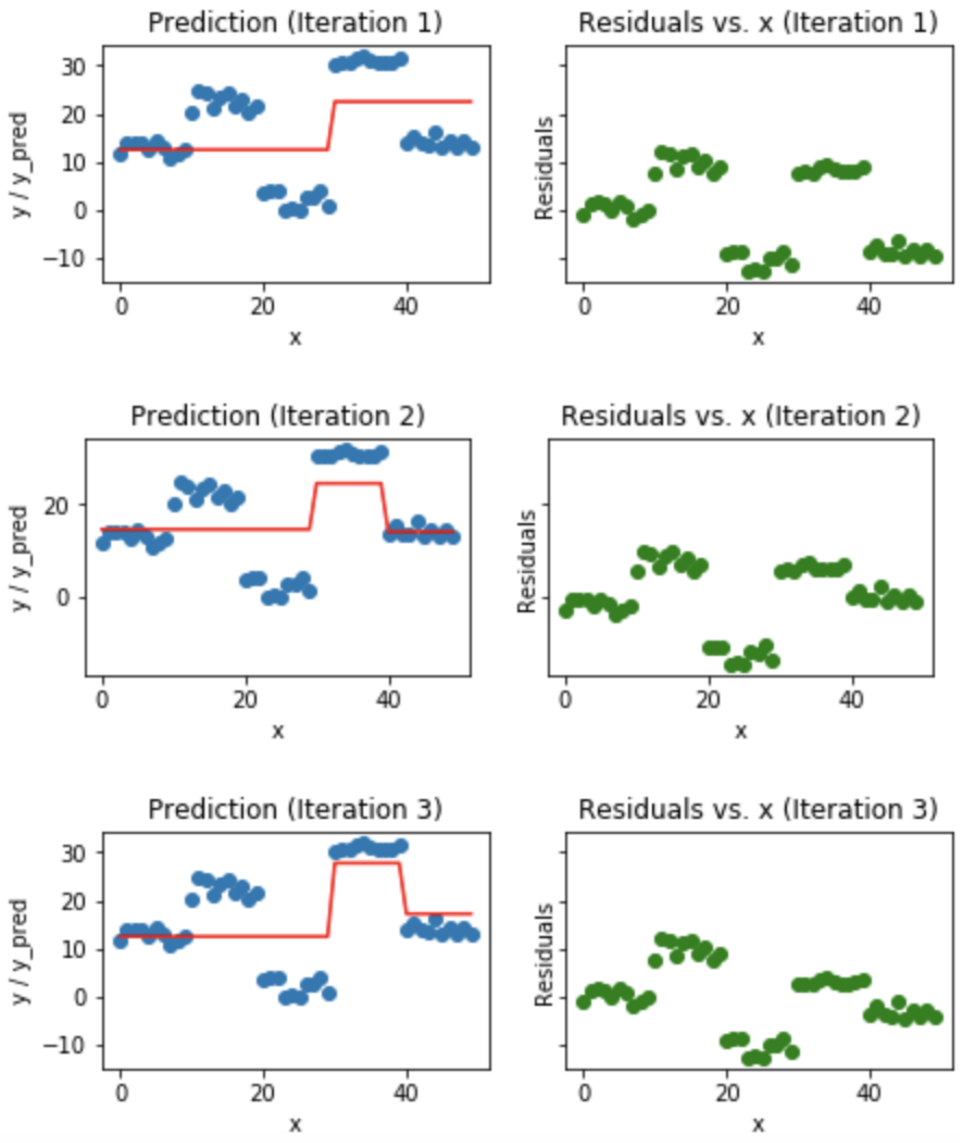
Overview of ML Survival Methods:

- Tree-based
- Boosting
- Support Vector Machine
- Bayesian-based
- Deep Learning/Neural Networks

Boosting:



https://medium.com/greyatom/a-quick-guide-to-boosting-in-ml-acf7c1585cb5



Boosting:

Gradient Boosting Machine (gbm)

Boosting Concordance Index (Github GBMCI, BoostCI)

Generalized Linear Model Boosting (GlmBoost)

Multivariate boosting for longitudinal data

First Hitting Time Model For XGBoost (Github HitBoost)

G. Ridgeway, “The state of boosting,” *Comput. Sci. Stat.*, 1999.

Y. Chen, Z. Jia, D. Mercola, and X. Xie, “A Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index,” *Comput. Math. Methods Med.*, vol. 2013, pp. 1–8, 2013.

A. Mayr, H. Binder, O. Gefeller, and M. Schmid, “The Evolution of Boosting Algorithms - From Machine Learning to Statistical Modelling,” pp. 1–32, 2014.

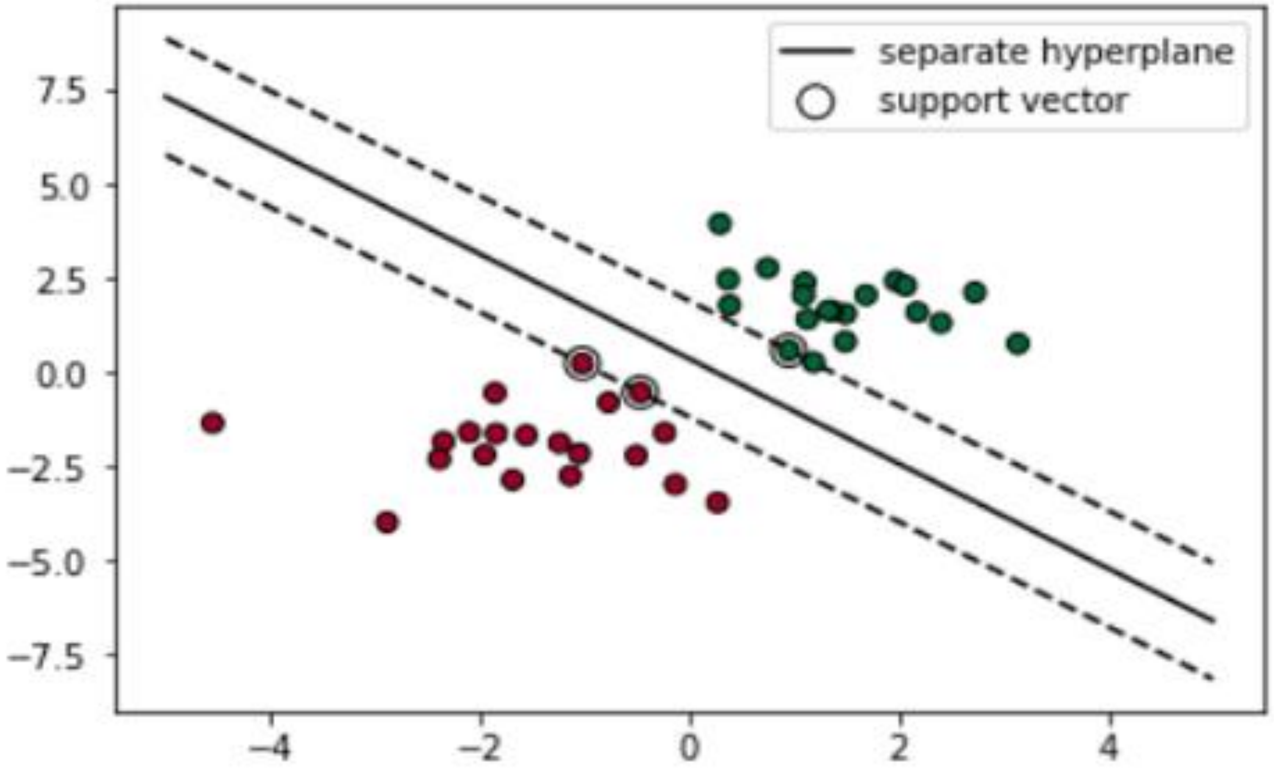
P. Bühlmann, “Boosting for high-dimensional linear models,” *Ann. Stat.*, 2006.

P. Liu, B. Fu, and S. X. Yang, “HitBoost: Survival Analysis via a Multi-Output Gradient Boosting Decision Tree Method,” *IEEE Access*, vol. 7, pp. 56785–56795, 2019

Overview of ML Survival Methods:

- Tree-based
- Boosting
- Support Vector Machine
- Bayesian-based
- Deep Learning/Neural Networks

SVM:



SV Regresion for censored data

SVM Combining Regression & Ranking

Relevance Vector Machine Survival

F. M. Khan and V. Bayer-Zubek, “Support vector regression for censored data (SVRc): A novel tool for survival analysis,” in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2008.

V. Van Belle, K. Pelckmans, S. Van Huffel, and J. A. K. Suykens, “Support vector methods for survival analysis: A comparison between ranking and regression approaches,” *Artif. Intell. Med.*, 2011.

C. J. K. Fouodo, I. R. König, C. Weihs, A. Ziegler, and M. N. Wright, “Support vector machines for survival analysis with R,” *R J.*, vol. 10, no. 1, pp. 412–423, 2018.

F. Kiaee, H. Sheikhzadeh, and S. Eftekhari Mahabadi, “Relevance Vector Machine for Survival Analysis,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 27,

Overview of ML Survival Methods:

- Tree-based
- Boosting
- Support Vector Machine
- Bayesian-based
- Deep Learning/Neural Networks

Bayesian Survival Methods:

Naïve-Bayes

Bayesian Network as feature selection

Bayesian + Accelerated Failure Time

Likelihood

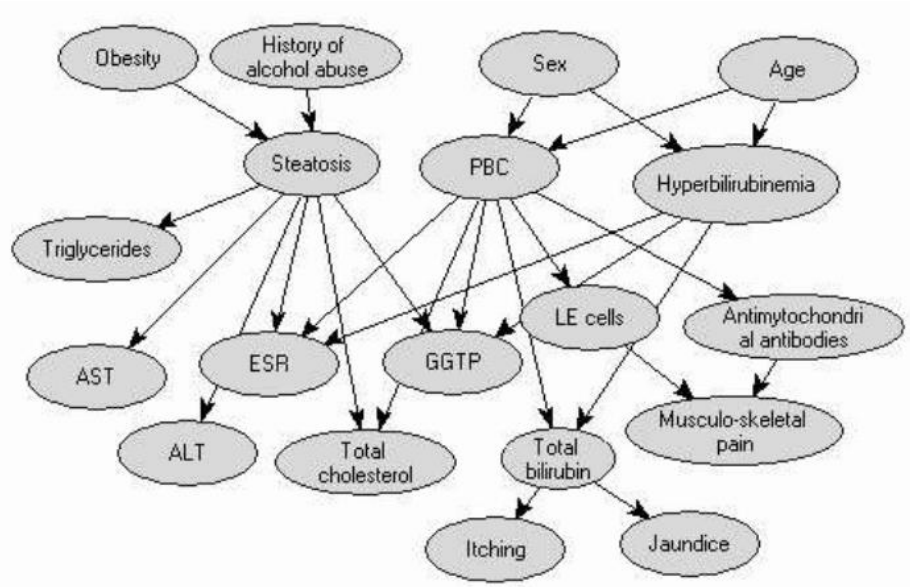
Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$



M. J. Fard, P. Wang, S. Chawla, and C. K. Reddy, "A Bayesian Perspective on Early Stage E *Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3126–3139, 2016.

C. T. Volinsky, D. Madigan, A. E. Raftery, and R. A. Kronmal, "Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke," *J. R. Stat. Soc. Ser. C Appl. Stat.*, 1997.

P. J. G. Lisboa, H. Wong, P. Harris, and R. Swindell, "A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer," *Artif. Intell. Med.*, vol. 28, no. 1, pp. 1–25, 2003.

Overview of ML Survival Methods:

- Tree-based
- Boosting
- Support Vector Machine
- Bayesian-based
- Deep Learning/Neural Networks:
  - For Structured Data
  - For Image Data

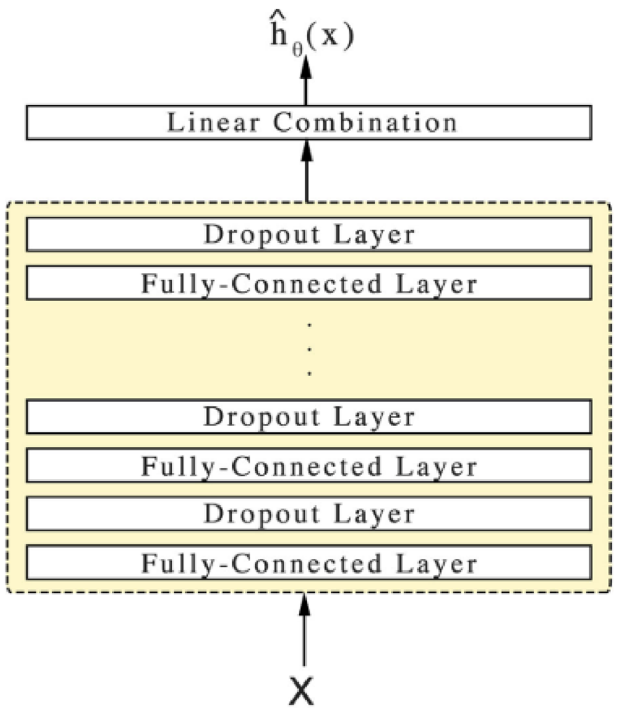
Deep Learning Survival for structured data:

Cox-based:

- DeepSurv
- Cox-nnet

Discrete-time:

- Nnet-survival
- DeepHit
- Dynamic DeepHit



J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Med. Res. Methodol.*, vol. 18, no. 1, pp. 1–12, 2018.

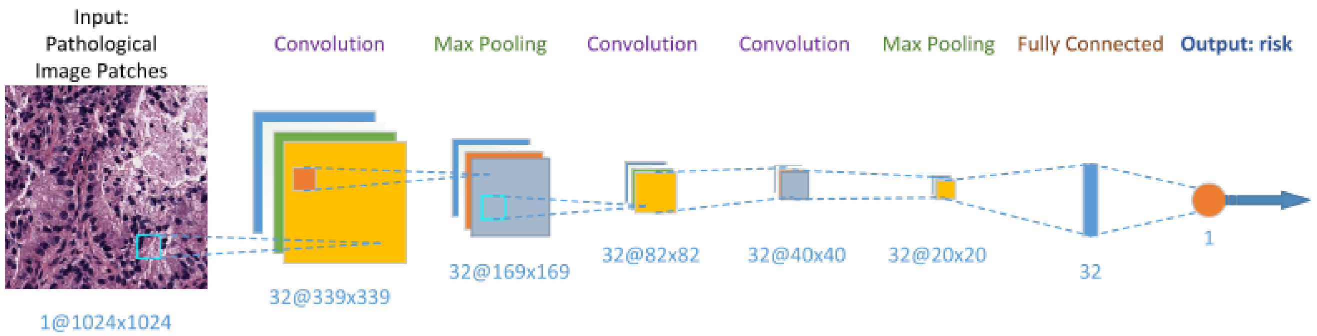
T. Ching, X. Zhu, and L. X. Garmire, "Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data," *PLoS Comput. Biol.*, vol. 14, no. 4, pp. 1–18, 2018.

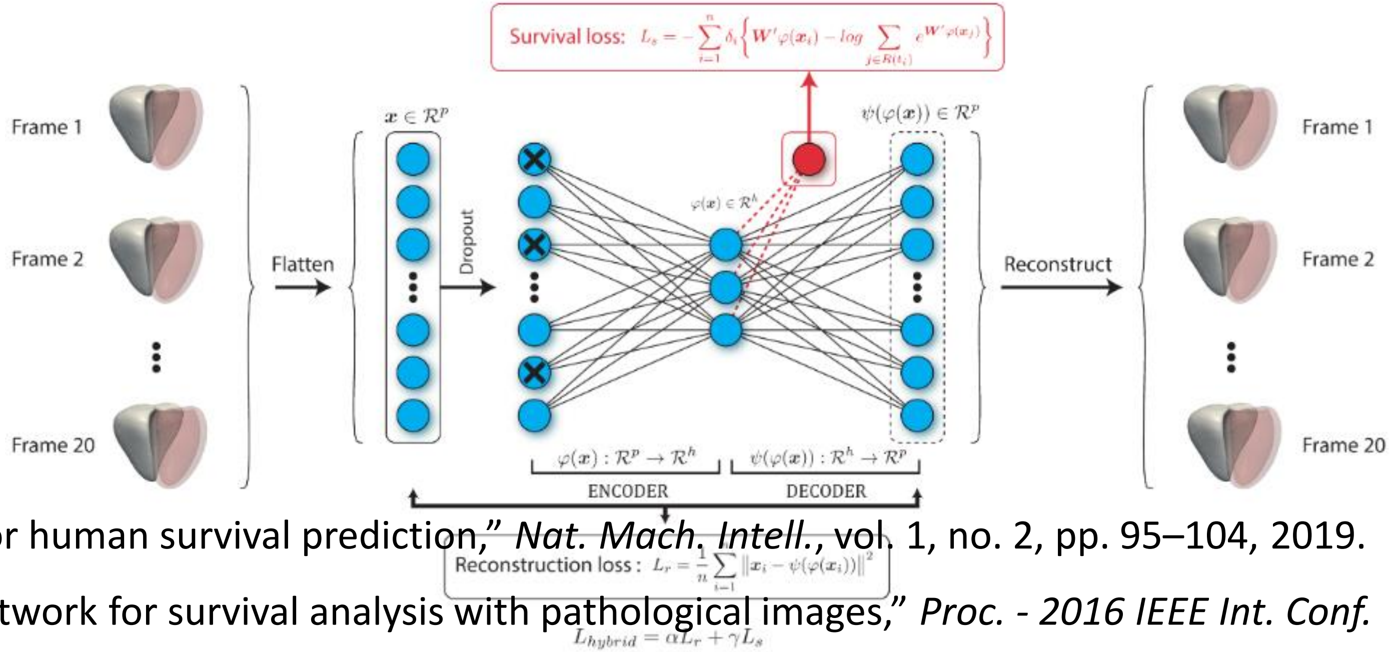
M. F. Gensheimer and B. Narasimhan, "A scalable discrete-time survival model for neural networks," *PeerJ*, vol. 2019, no. 1, pp. 1–19, 2019.

C. Lee, W. R. Zame, J. Yoon, and M. Van Der Schaar, "DeepHit: A deep learning approach to survival analysis with competing risks," *32nd AAAI Conf. Artif. Intell. AAAI 2018*, pp. 2314–2321, 2018.

Deep Learning Survival for Image Data:

- DeepConvSurv
- DeepConvSurv + clustering for high resolution images
- 4-D survival for sequences of MRI images





G. A. Bello *et al.*, “Deep-learning cardiac motion analysis for human survival prediction,” *Nat. Mach. Intell.*, vol. 1, no. 2, pp. 95–104, 2019.

X. Zhu, J. Yao, and J. Huang, “Deep convolutional neural network for survival analysis with pathological images,” *Proc. - 2016 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2016*, no. 1, pp. 544–547, 2017.

C. Zhao, X. Zhou, M. Mao, and P. Tang, “WSISA: Making Survival Prediction From Whole Slide Histopathological Images,” *IEEE Access*, vol. 2, pp. 970–975, 2017.

Outline

- ML Introduction & Fun Facts
- ML for classification: overview & limitations for survival analysis
- ML methods for survival: overview & their benefits
- Limitations of current ML survival methods

Limitations of state-of-the-art ML survival:

- Performance gain
- Interpretability
- Longitudinal data
- Recurrent events
- Competing risks
- Multi-modal data

Lack of comprehensively verified performance gains

- In most papers: the gain in C-index or time-varying AUC only stay within the 1 or 2% increase compared to the traditional methods
- Only compared their proposed method against two or three other methods
- Very limited reports available that gather these methods together and evaluate them on the same dataset(s).

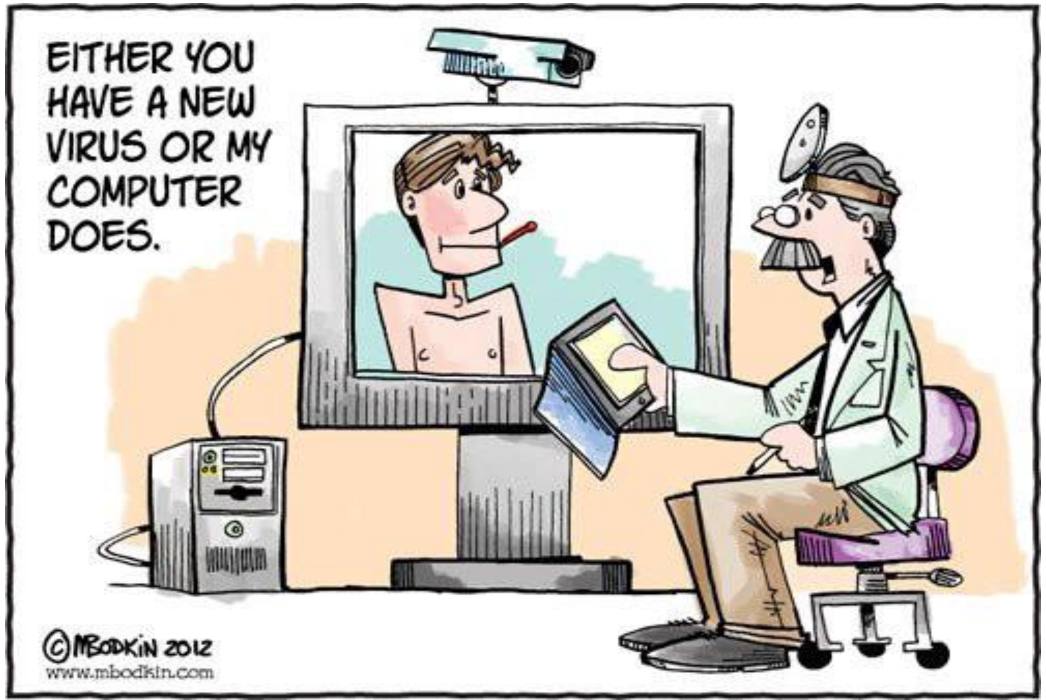
Lack of interpretability

- Limited interpretability: there's heatmap, but anything else?
- The belief that black-boxes are necessary to make good prediction is not backed by evidence

C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.

- Performance Gain vs. Complexity

Trade-off

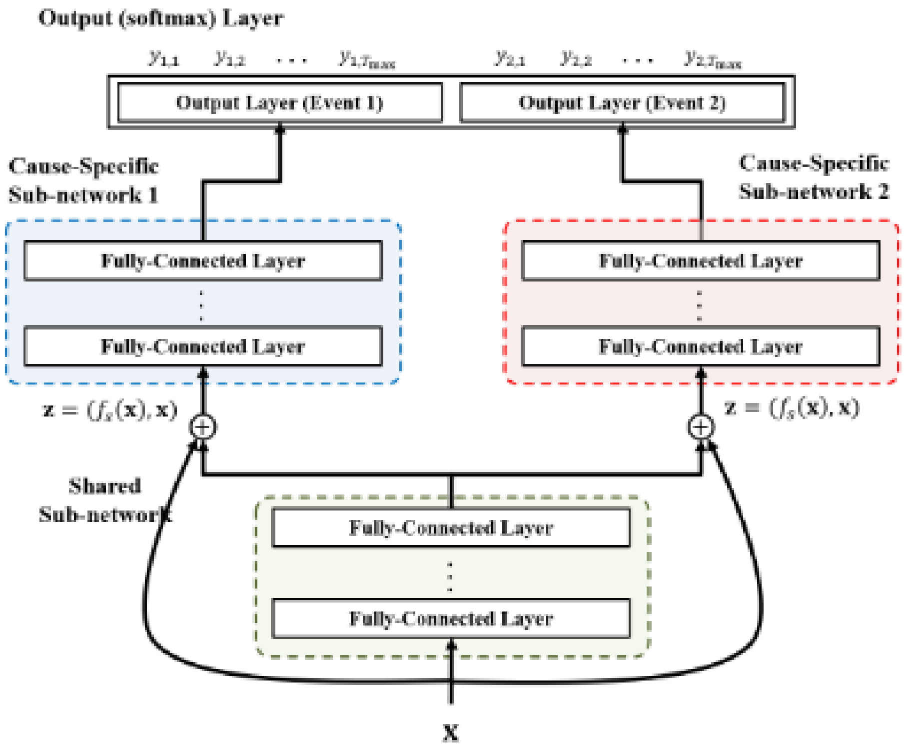


Lack of capability to handle longitudinal data

- Most covariates, such as biomarkers and risk factors, are repeatedly collected over time
- Except for a few methods (RF-SLAM, Dynamic-DeepHit, RNN-SURV), most ML methods only use the last available measurement, discarding valuable information that could enhance accuracy, such as evolution of covariates throughout the year

Lack of capability to deal with competing risks and recurrent events

- Only DeepHit and RSF can deal with competing risks
- Only RF-SLAM addresses recurrent events, however it treats past events quite simple



Lack of capability to incorporate image data

- Performances of DeepConSurv and 4-D survival are still low
- Use Cox's loss function → inherit Cox's proportional hazards assumption

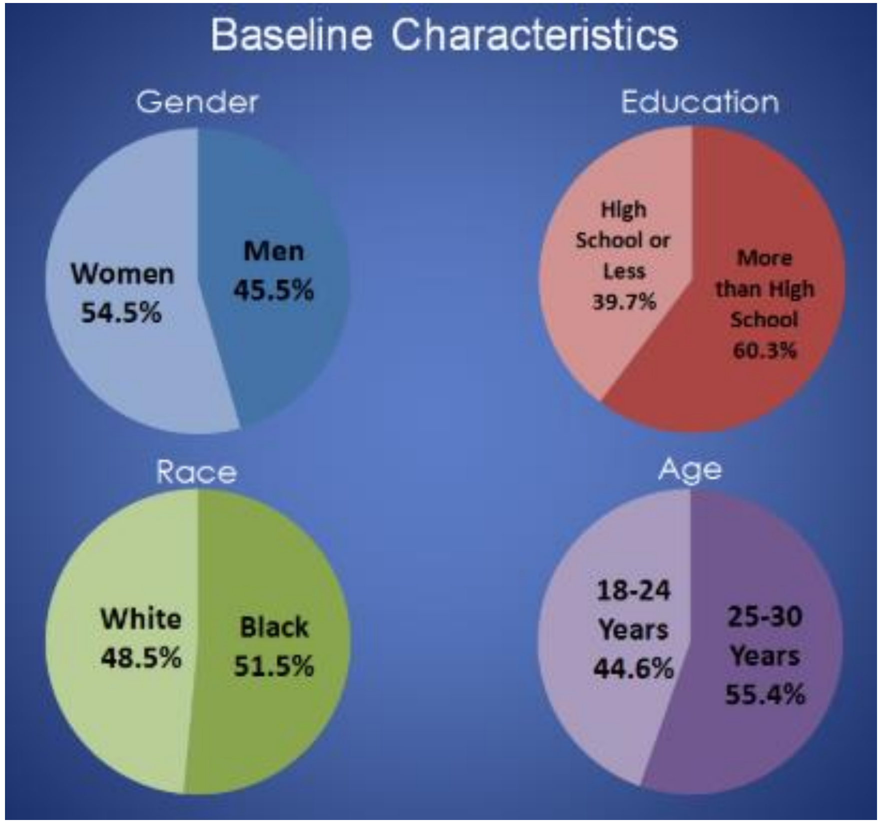
- No method to-date that has been extended to integrate latent image variables with structured data features such as demographics and labs

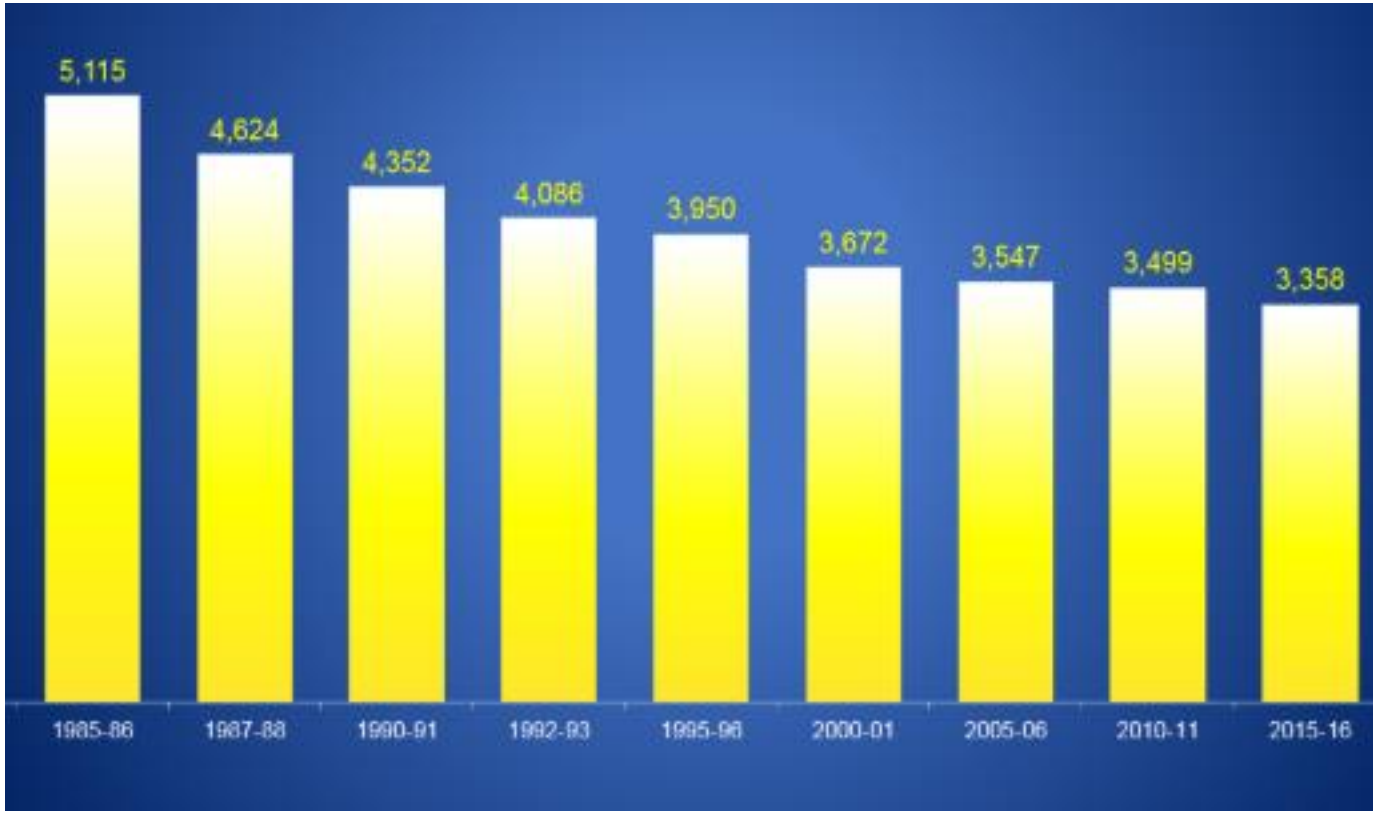
Computational Time and Resources

More hyperparameters → exponential increase in runtime for hyperparameter tuning, especially for advanced tuning methods (i.e. Bayesian Optimization, Particle Swarm)

- RSF: ~5 hyperparam
- Boosting: ~7 hyperparam
- Neural Network Architecture: >12 hyperparam

Motivating example (part I of my thesis): Evaluating performance of ML survival methods vs. Cox





https://www.cardia.dopm.uab.edu/

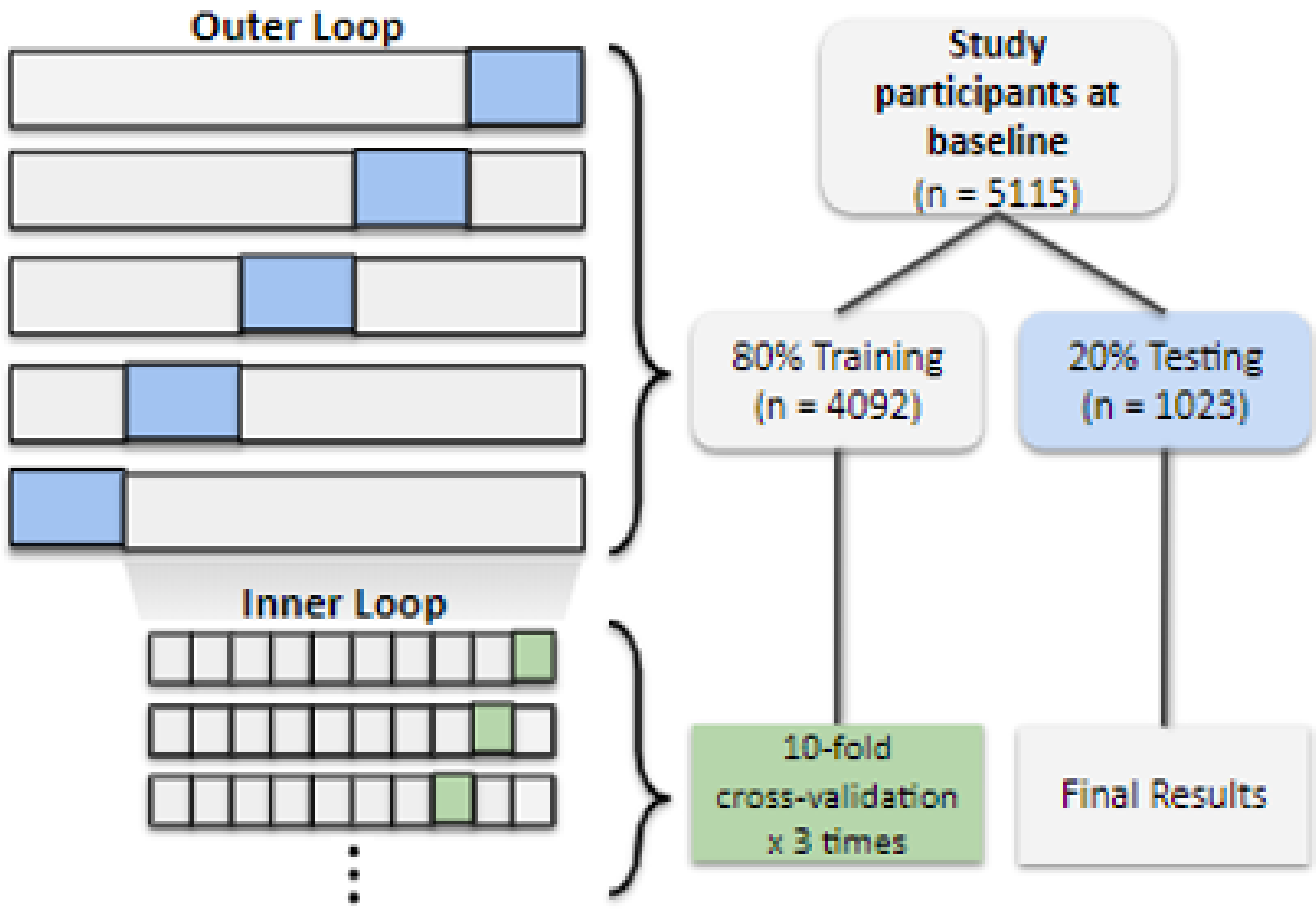
Data and Outcome

- Year 5 (Exam 3) data was selected to predict disease onset (for now)

(Cross-sectional variables)

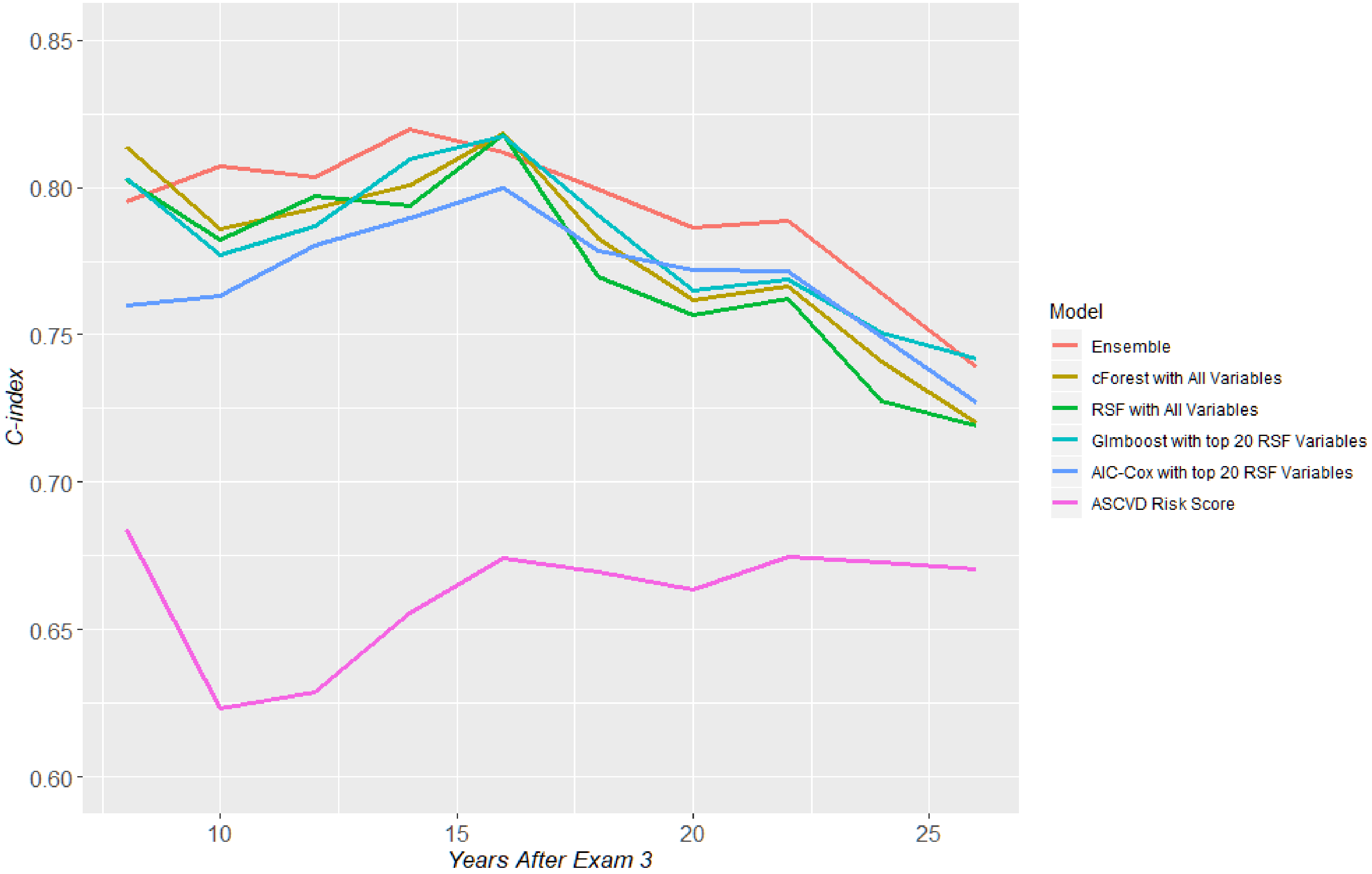
- 520 variables (including derived variables) (will see later)
- 245 CVD events by 2018

Workflow


Outer Loop
Inner Loop
Study participants at baseline
(n = 5115)
80% Training
(n = 4092)
20% Testing
(n = 1023)
10-fold cross-validation x 3 times
Final Results

Name	Description	Package or Github name (in R, asterisk if in Python)
Coxph	Cox proportional hazard[13]	Survival
AIC-Cox	Akaike Information Criterion for Cox regression[14]	MASS
LASSO-Cox	Least absolute shrinkage and selection operator (L1) for Cox regression[15]	Glmnet
RSF	Random survival forest[11]	Rsfrc
Cforest	Conditional inference survival forest[16]	Party
CoxBoost	Component-wise likelihood-based boosting for Cox[17]	CoxBoost
Gbm	Gradient boosting machine[18]	Gbm
Glmboost	Gradient boosting with component-wise linear model[19]	Mboost
Cox-nnet	Cox proportional hazard adaptation to neural network[20]	*Cox-nnet[a]
DeepSurv	Cox proportional hazard adaptation to feed-forward neural network[21]	*DeepSurv[b]
Nnet-Survival	Discrete-time deep learning approach for survival analysis[22]	*Nnet-survival[c]
DeepHit	Discrete-time deep learning approach for survival analysis wih competing risks[23]	*DeepHit[d]

Preliminary Results



Summaries

- There are growing interests in ML
- Many ML methods have been adapted for survival analysis that could overcome limitations of the traditional statistical methods
- Still, the current ML methods have a number of limitations. Further research studies and external validation /verification studies are imperative.

Thank you for listening!

hnguye78@jh.edu

linkedin.com/in/hieuhughnguyen

I always welcome collaborations, feedbacks, opinions, and ideas