

Individual-Level Fatality Prediction of COVID-19 Patients

Xia ZHENG, Alex LIU, Haoyu SHI

March 14, 2021

Introduction

Introduction

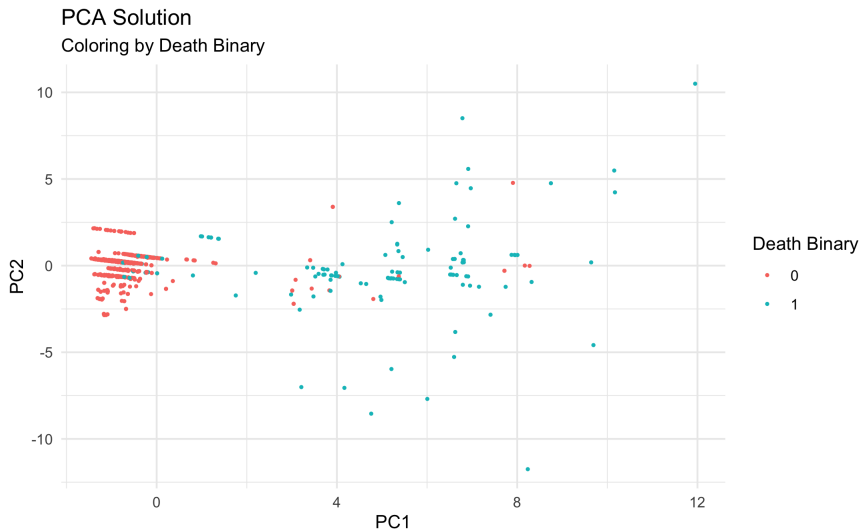
- ▶ COVID-19 has impacted everyone but not equally.
- ▶ Are there any association between different symptoms and pre-existing health conditions?
- ▶ What are the predominating symptoms and chronic diseases that may lead one to die from COVID-19?
- ▶ Provide better healthcare based on outcome prediction of a patient

Data

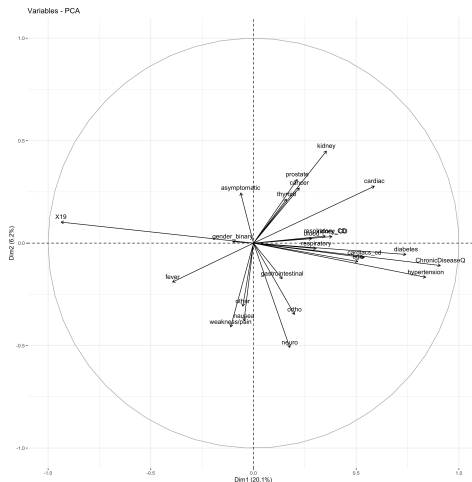
- ▶ Wolfram, "Patient Medical Data for Novel Coronavirus COVID-19" Dec, 2019 - April, 2020
- ▶ contains several variables for age, sex, symptoms, chronic disease, and binary code for death
- ▶ select patients with all the information recorded

Can onset of some symptoms imply the onset of others?

PCA solution - clear separation



Correlation among symptoms



Correlation among symptoms

Positive correlated symptoms:

- ▶ nausea, weakness/pain, other symptoms
- ▶ multiple chronic diseases, cardiac symptoms
- ▶ gastrointestinal symptoms, orthopedic chronic disease

Surprising discover:

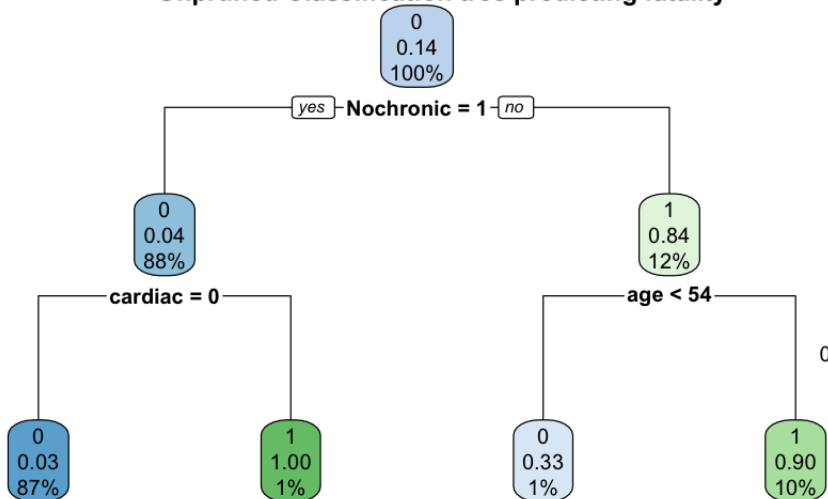
- ▶ 64.76% symptomatic patients experience fever.

Fever is a relatively independent symptom but it is strongly negative correlated to cardiac symptoms.

Fatality Prediction using Tree-Based Learners

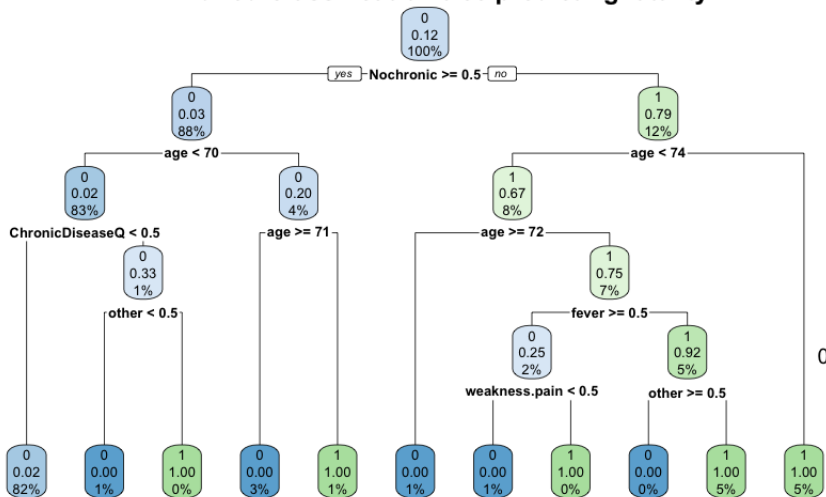
A Simple Classification Tree

Unpruned Classification tree predicting fatality

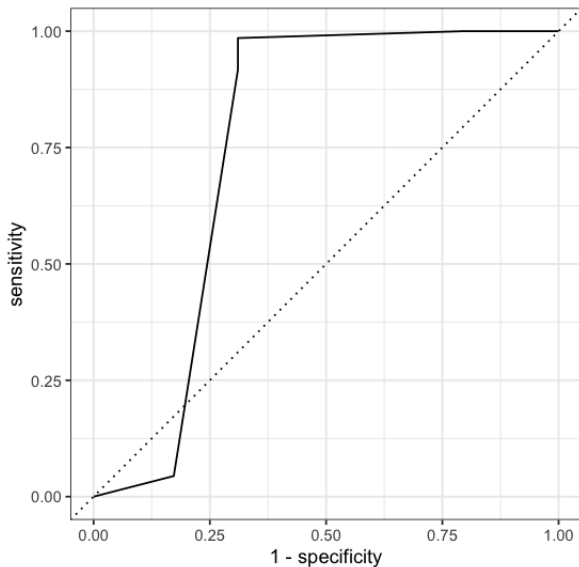


A Pruned Classification Tree

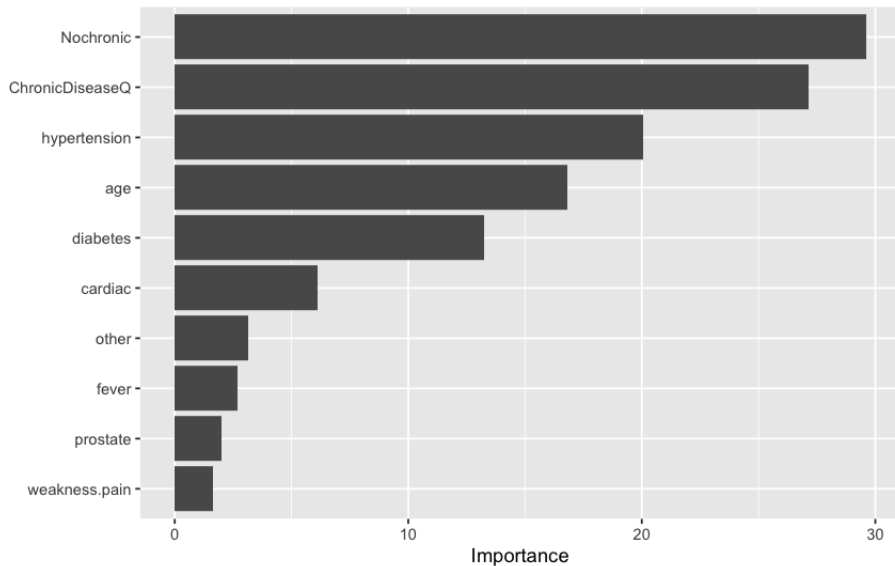
Pruned Classification tree predicting fatality



ROC Curve

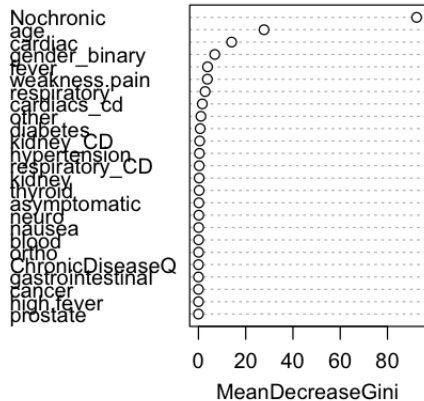
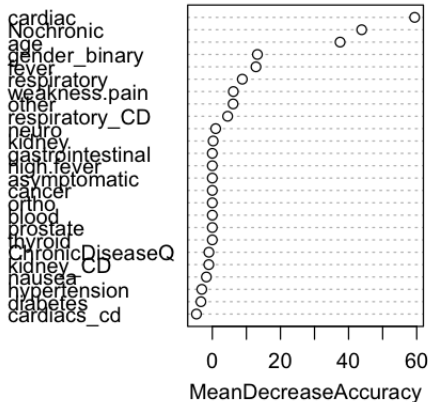


Variable Importance Plot



Bagged Trees

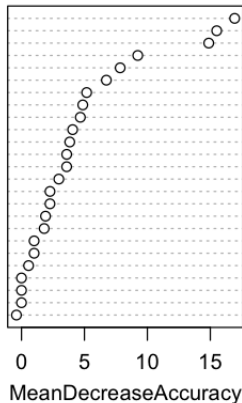
fatal_bag



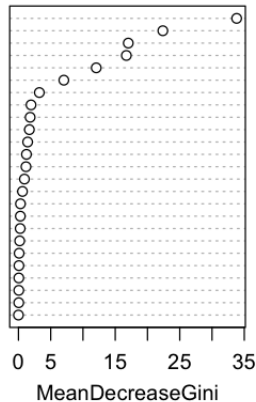
Random Forests

fatal_rf

Nochronic
cardiac
age
ChronicDiseaseQ
hypertension
fever
respiratory
weakness.pain
diabetes
kidney_CD
cardiacs_cd
other
respiratory_CD
gender_binary
asymptomatic
neuro
blood
kidney
prostate
ortho.
thyroid
gastrointestinal
high.fever
cancer
nausea



Nochronic
ChronicDiseaseQ
age
hypertension
cardiac
diabetes
cardiacs_cd
gender_binary
fever
respiratory
weakness.pain
kidney_CD
respiratory_CD
kidney
other
thyroid
blood
neuro
nausea
ortho
cancer
prostate
asymptomatic
gastrointestinal
high.fever



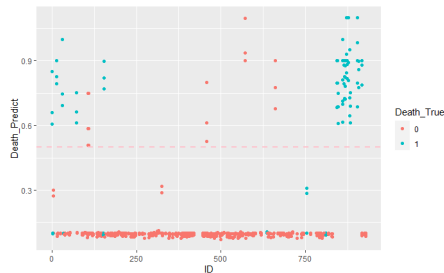
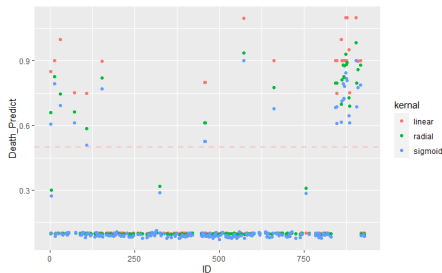
Comparing the Tree-Based Learners

Model	Prediction Accuracy on Test Set
Simple Classification Tree	0.9440
Pruned Tree	0.9483
Bagged Trees	0.9310
Random Forests	0.9526

Fatality Prediction Results from Other Methods

SVM

- ▶ SVM provides non-binary values
- ▶ 0.5 cut-off
- ▶ hard to interpret than tree



Comparing KNN (K=6) and SVM

Statistics	KNN (K=6)	SVM
Sensitivity	0.7931	0.7931
Specificity	0.9753	0.9704
Precision	0.8214	0.7931
Accuracy	0.9526	0.9483

Conclusion

Result and Discussion

- ▶ Our models are very consistent in terms of prediction and interpretation
- ▶ Very high specificity but lower sensitivity. Need to be cautious to merely rely on our machine learning models for evaluation.
- ▶ Chronic diseases, such as hypertension, diabetes, and cardiac, are predominant risk factors associated with the death for COVID-19 cases.
- ▶ Provide better prevention and healthcare for old people with chronic diseases.

Limitation and Improvement

- ▶ Not very sure about ChronicDiseaseQ and NonChronic in our data due to lack of sufficient documentation. However, dropping these two variables makes the sensitivities in our models lower. Trade-off between accuracy and interpretation.
- ▶ Only use patients with all records may have bias.
- ▶ In the early stage, many death with COVID-19 might not be recorded. Most early cases are clustered in certain regions, which may not be generalized to all populations.
- ▶ May contact to healthcare system/medical institution for more reliable and up-to-date records