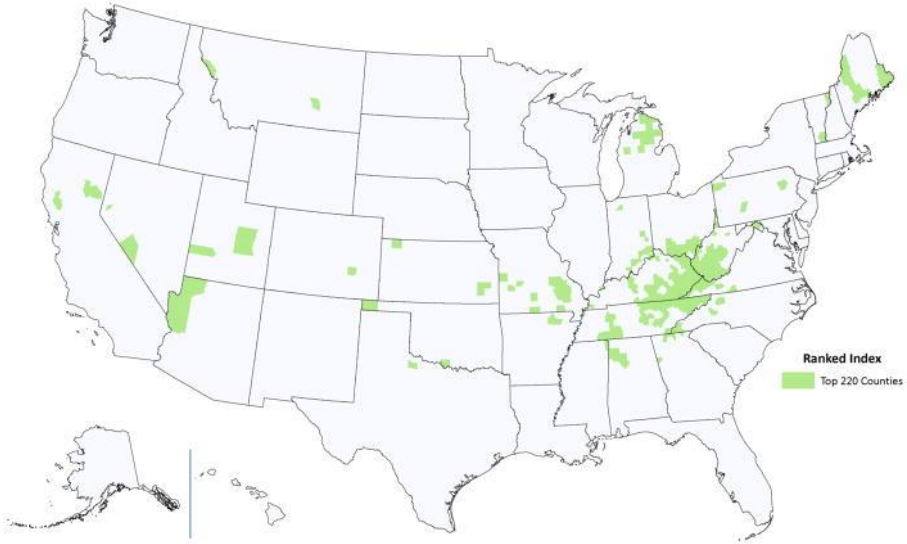
THE OHIO STATE UNIVERSITY

Insights for HIV Interventions among Millennials

Yun Ye, MPH
College of Public Health
Division of Epidemiology

HIV among Millennials

- Several recent HIV outbreaks in millennial-aged young people in US
- CDC has highlighted counties that are at risk for an HIV outbreak



Counties for which estimated vulnerability scores or their upper 90% confidence interval exceeded the 95th percentile. (CDC)

Research Question

What are the demographical, social economical and behavioral features related to *HIV testing* and *risk behavior* among millennials?

Data Source

BRFSS™

The Behavioral Risk Factor Surveillance System (2019)

- Health-related risk behaviors, chronic health conditions, and use of preventive services
- A powerful tool for targeting and building health promotion activities

Sample size

- 418,268 in total, we are focusing on millennials

Primary Outcome

HIV risk behaviors (Y/N)

I am going to read you a list. When I am done, please tell me if any of the situations apply to you. You do not need to tell me which one.

- You have injected any drug other than those prescribed for you in the past year.
- You have been treated for a sexually transmitted disease or STD in the past year.
- You have given or received money or drugs in exchange for sex in the past year.

HIV testing (Y/N)

Have you ever been tested for HIV? Do not count tests you may have had as part of a blood donation. Include testing fluid from your mouth.

Main Variables

Health behaviors	
Drinking	Total number of alcoholic beverages consumed per week
Smoking	Four-level smoker status (Everyday smoker, Someday smoker, Former smoker, Non-smoker)
Diet (fruit)	Total fruits consumed per day
Diet (Vegetable)	Total vegetables consumed per day
Physical exercise	Adults who reported doing physical activity or exercise during the past 30 days other than their regular job
	Minutes of total Physical Activity per week
	Meet the Aerobic and Strengthening standard or not (2-level)

Main Variables

Chronic conditions	
Asthma	Adults who have ever been told they have asthma
Difficulty Concentrating or Remembering	
Diabetes	(Ever told) you had diabetes
Heart attack	Ever Diagnosed with Heart Attack
Stroke	Ever Diagnosed with a Stroke
Healthcare access	
Do you have one person you think of as your personal doctor or health care provider?	
Could Not See Doctor Because of Cost (Y/N)	

Main Variables

Demographics	
Age (in 5-yr categories)	Level of education completed
BMI (Four-categories)	Marital status
Sex	Own or rent your home
Employment	Urban/Rural Status
Imputed race/ethnicity value	Veteran status
Income categories	

Data Cleaning

- Subset to millennials
- Keep the variables we are interested in
- Imputation for missing values in independent variables
- Recode categorical variables
- One hot encoding for the categorical variables with multiple levels
- Log-transform the continuous variables to reduce their range

Statistical Analysis

- Deep neural network (DNN)
- Random forest

DNN Tuning Process

- Adaptive learning rate (power scheduling)
- Early stopping based on validation accuracy (callbacks)
- Adjust for imbalance data (class_weight) if needed
- Careful initialization of weights (initial_weights)

- Batch size
- Hidden layers
- Nodes
- Dropout

HIV Risk Behavior

- Sample size: 81,942
  - 52442 train examples
  - 13111 validation examples
  - 16389 test examples

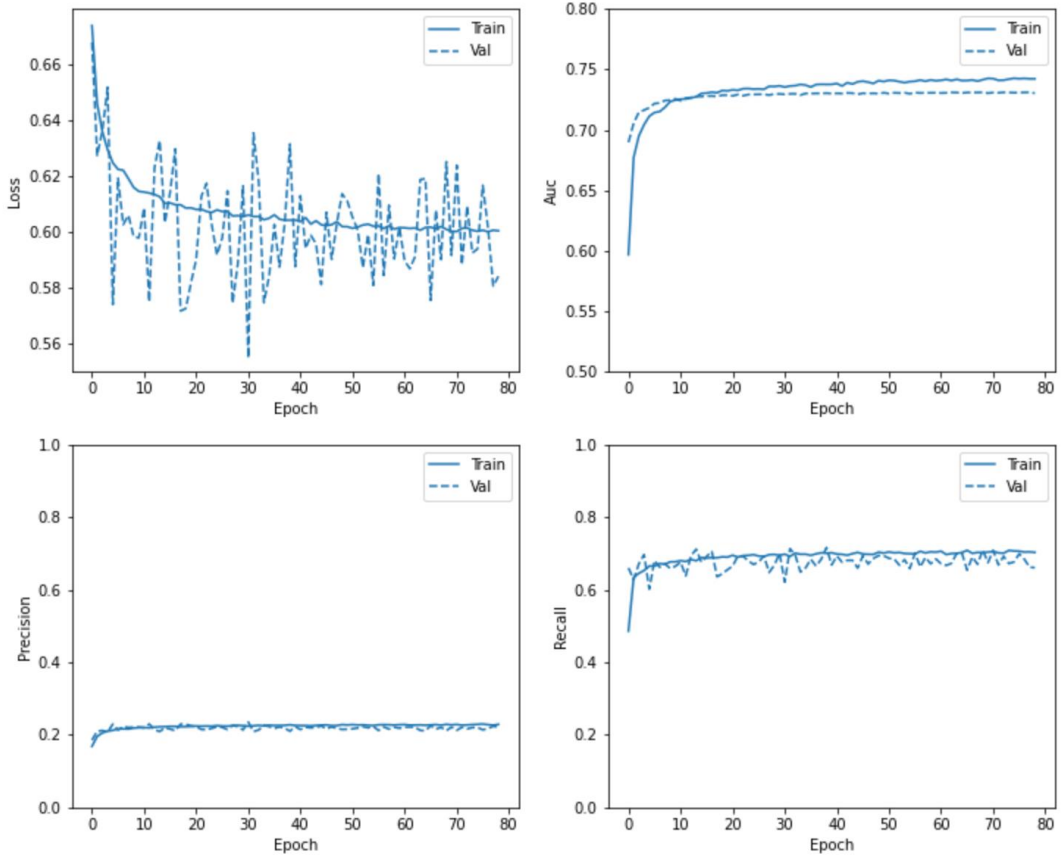
- Imbalance results (12.41% positive)

Age	
18-24	22,154
25-29	18,089
30-34	20,075
35-39	21,624
Birth Sex	
Female	40,260
Male	41,682

Model 1:

- 1 layer
- 128 nodes
- 0.5 dropout

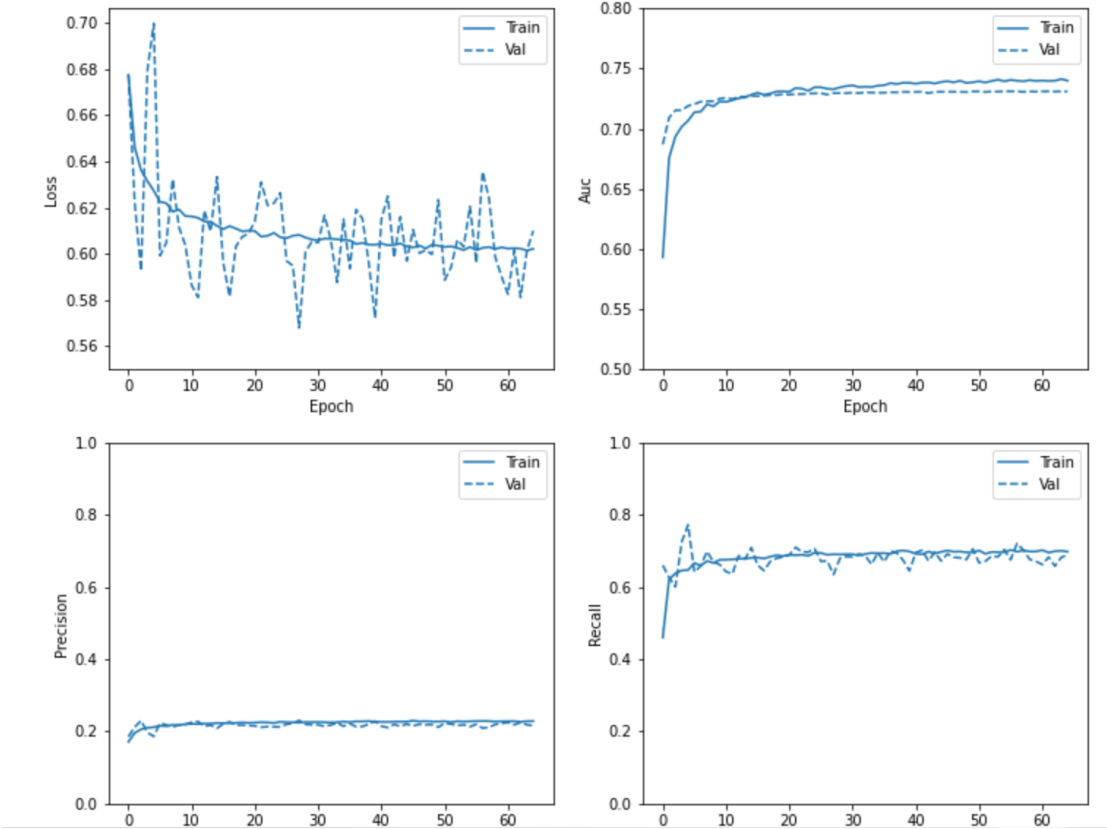
- PRC-AUC: 0.29



Model 2: More nodes

- 1 layer
- 2048 nodes
- 0.5 dropout

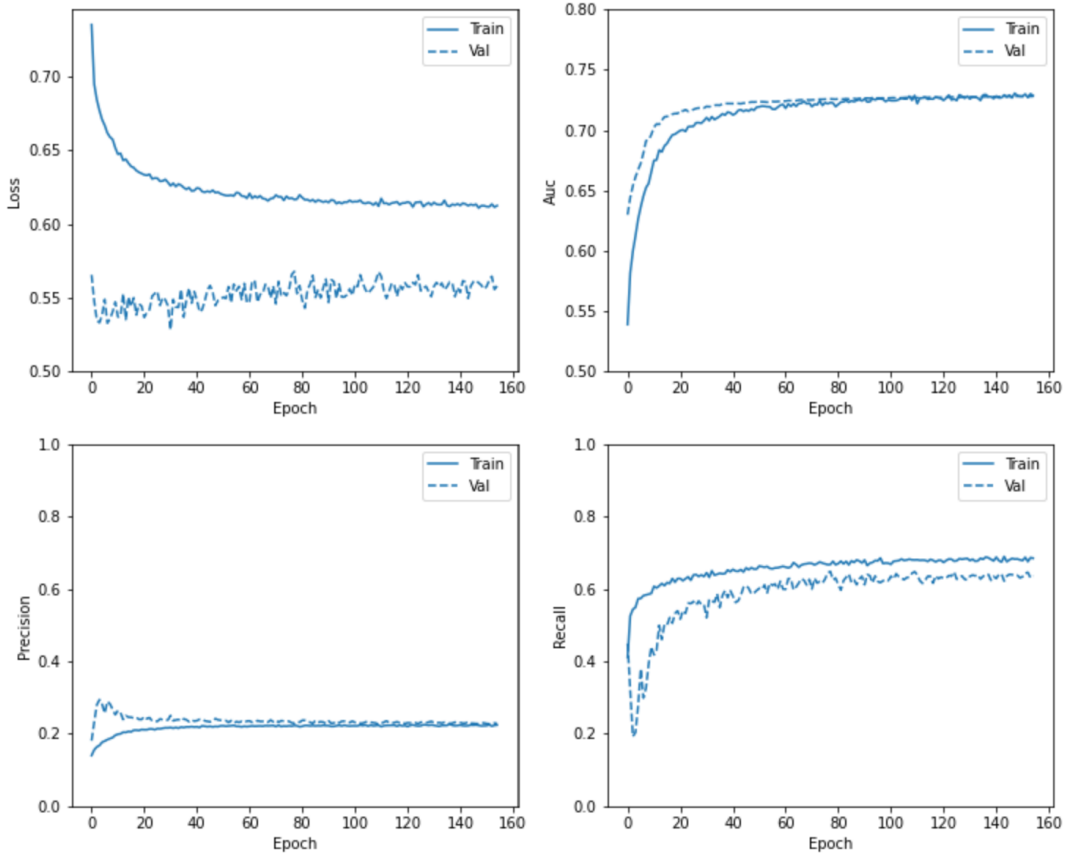
- PRC-AUC: 0.28



Model 3: More layers

- 2 layers
  - 128 nodes
  - 0.5 dropout
  - 64 nodes
  - 0.2 dropout

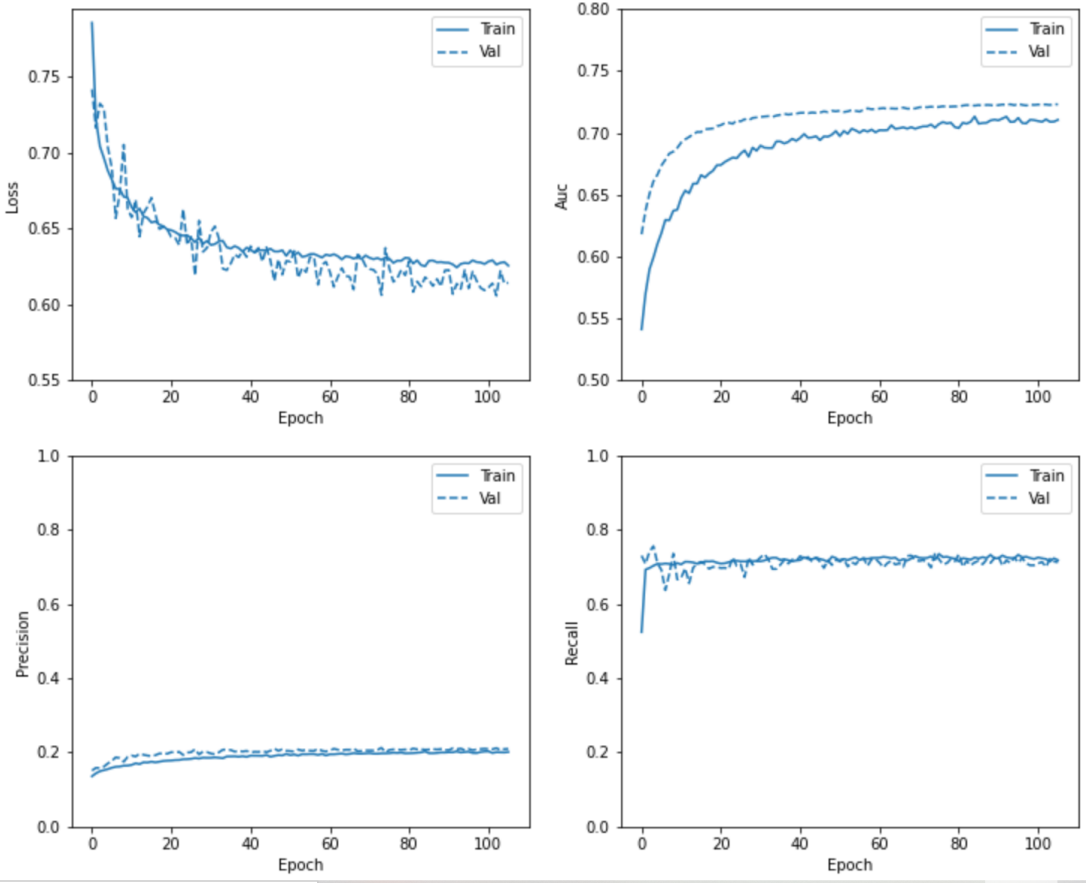
- PRC-AUC: 0.28



Model 4: Less nodes

- 1 layer
- 32 nodes
- 0.5 dropout

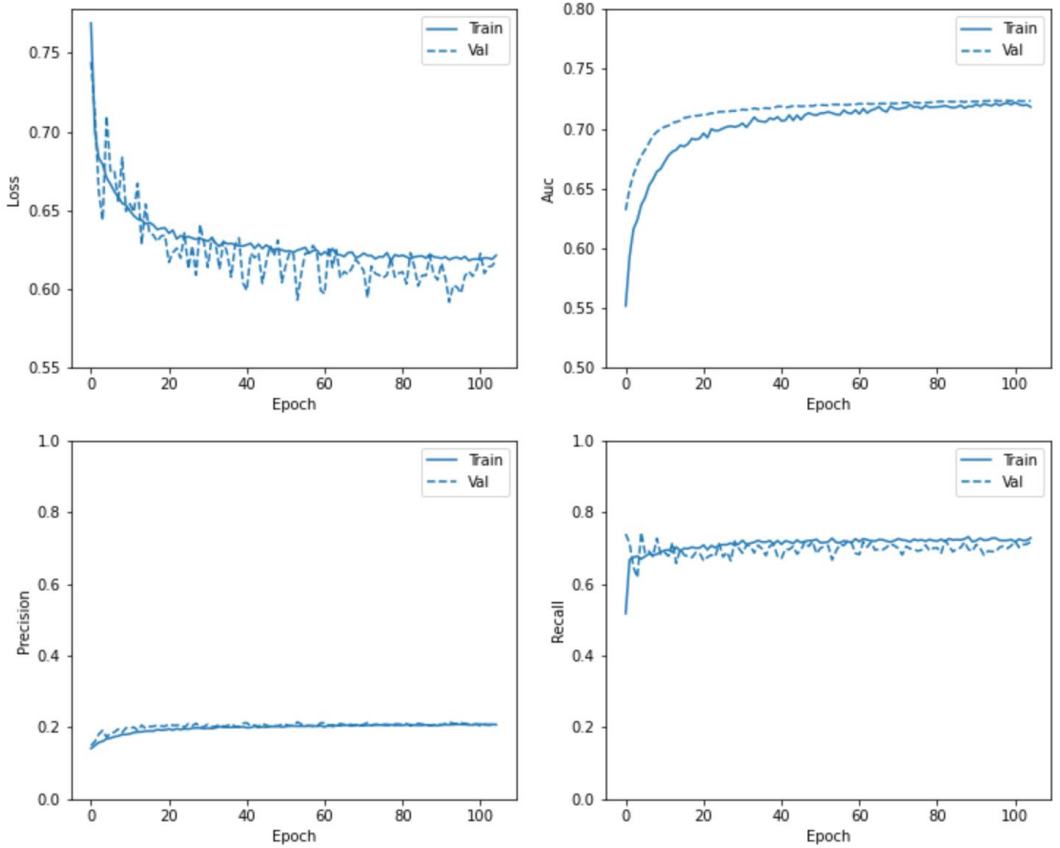
- PRC-AUC: 0.26



Model 5: Increase the number of nodes

- 1 layer
- 64 nodes
- 0.5 dropout

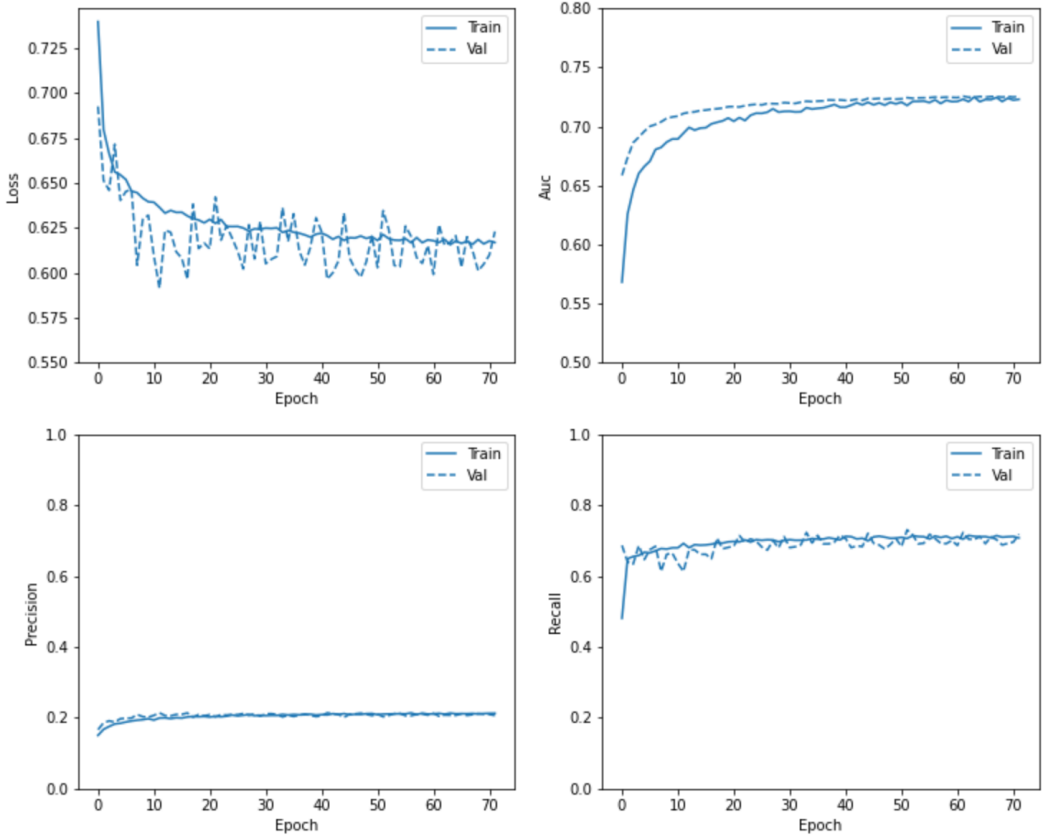
- PRC-AUC: 0.27



Model 6: Increase the number of nodes

- 1 layer
- 96 nodes
- 0.5 dropout

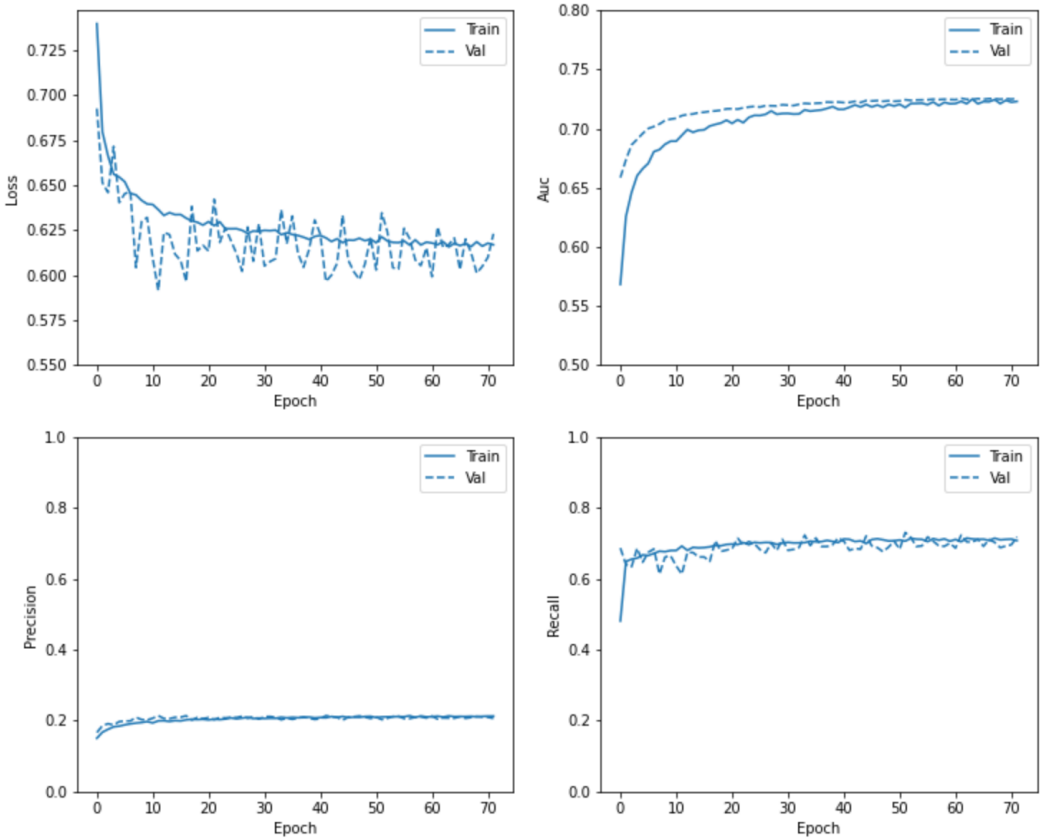
- PRC-AUC: 0.28



Random Forest

- Adjusted for imbalance data

- Mean Accuracy: 0.873
- Mean Precision: 0.330



DNN results

HIV Testing

https://www.avert.org/infographics/what-happens-when-you-go-hiv-test

HIV Testing

- Sample size: 80,166
  - 51305 train examples
  - 12827 validation examples
  - 16034 test examples

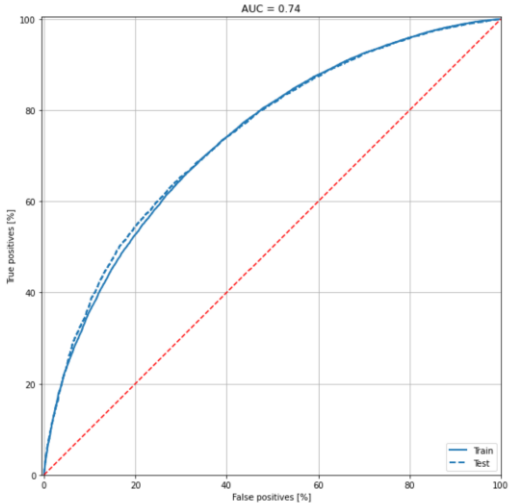
- Balance results (49.30% positive)

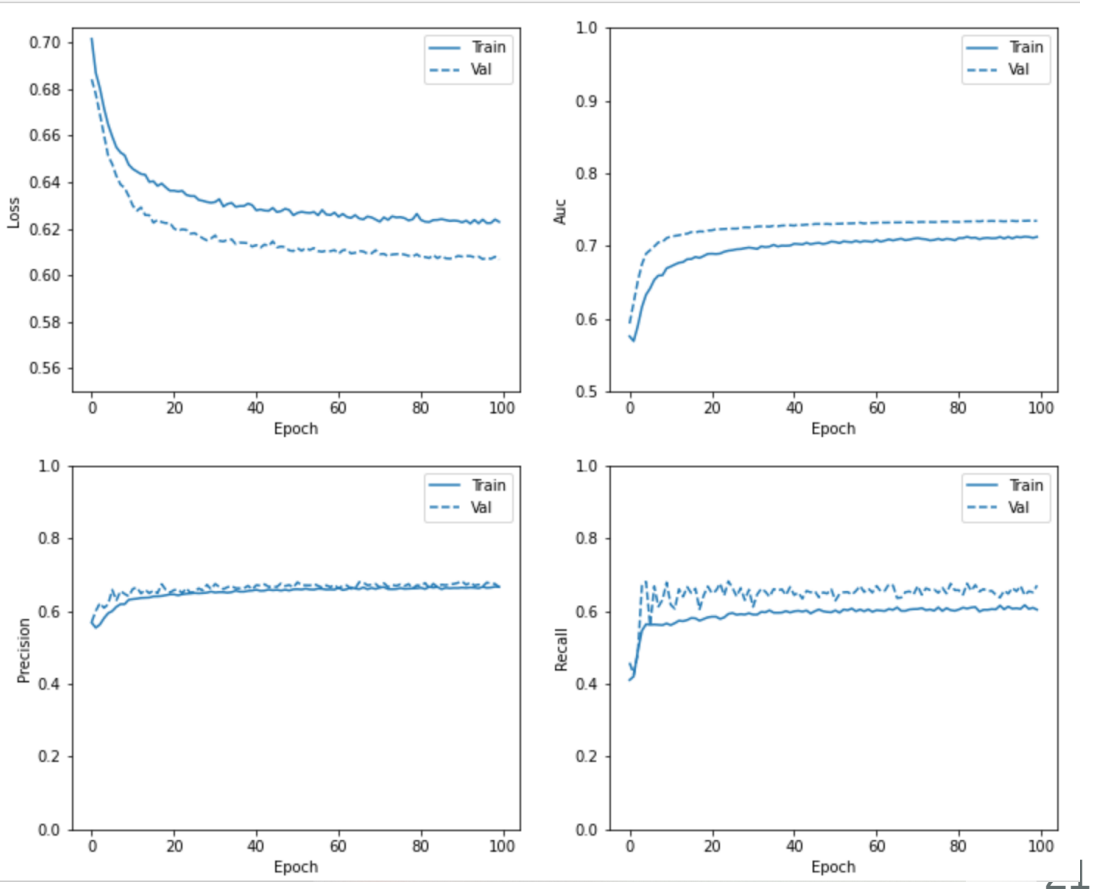
Age	
18-24	21,759
25-29	17,708
30-34	19,589
35-39	21,110
Birth Sex	
Female	39,135
Male	41,031

Model 1:

- 1 layer
- 16 nodes
- 0.5 dropout

- ROC-AUC: 0.74

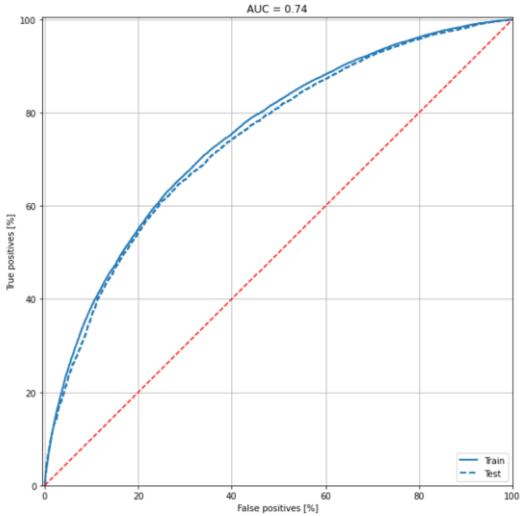


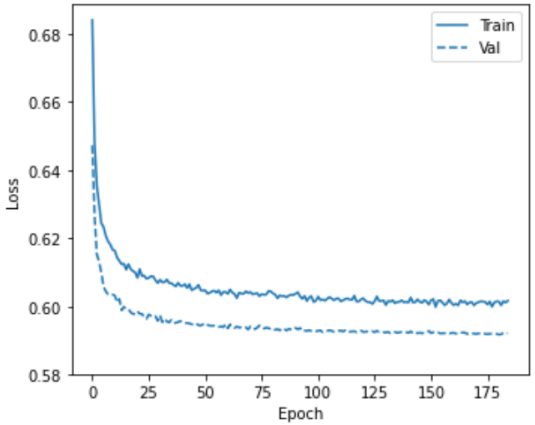


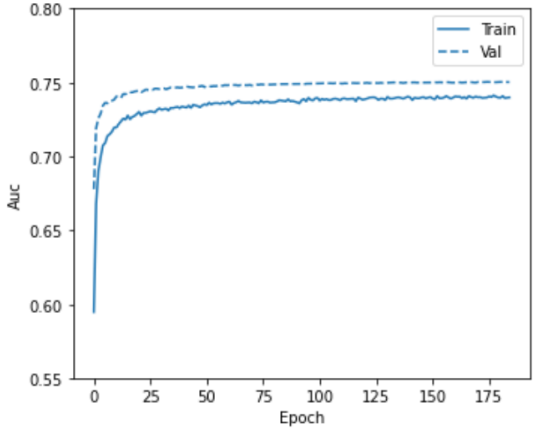
Model 2: More nodes

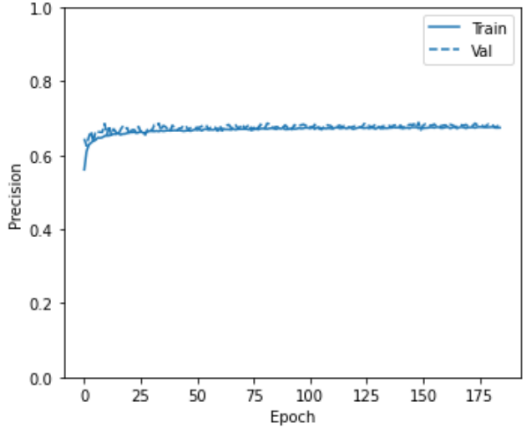
- 1 layer
- 128 nodes
- 0.5 dropout

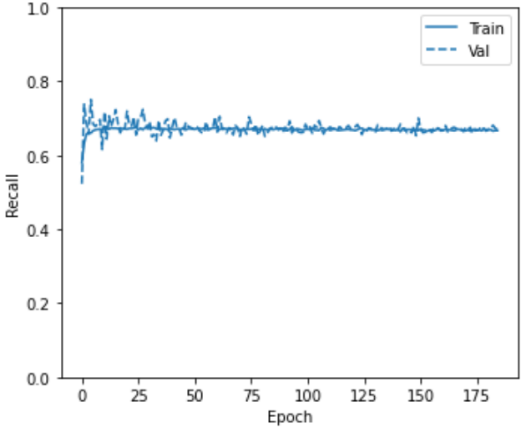
- ROC-AUC: 0.74







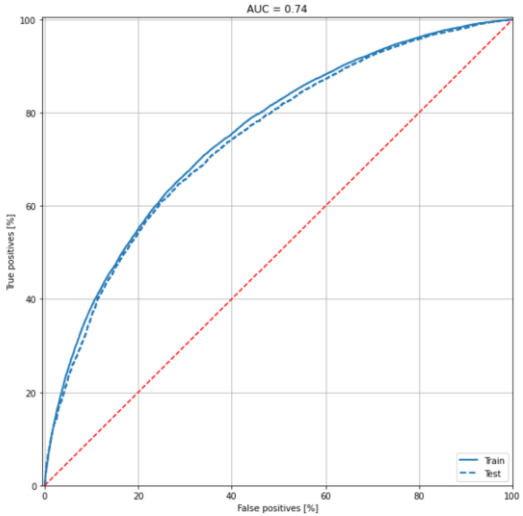


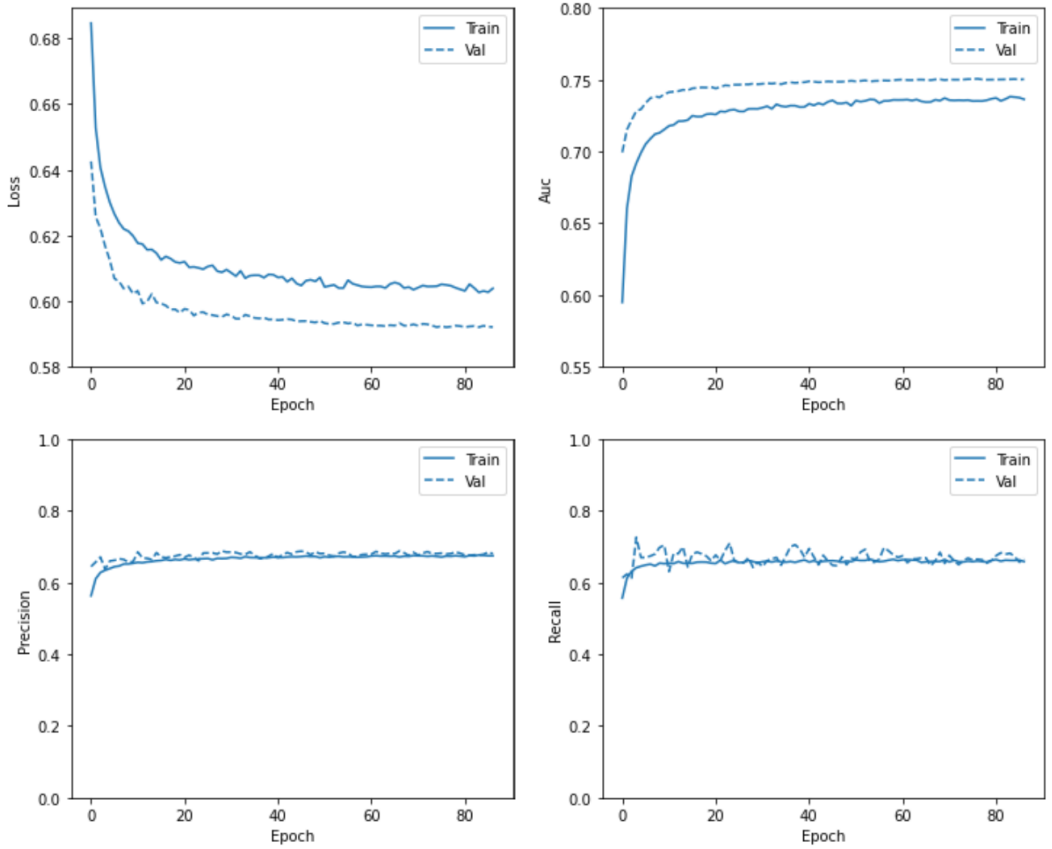


Model 3: More layers

- 2 layers
- 128 nodes (0.5 dropout)
- 64 nodes (0.2 dropout)

- ROC-AUC: 0.74

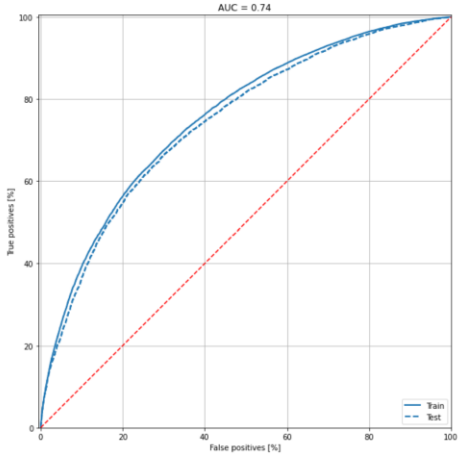


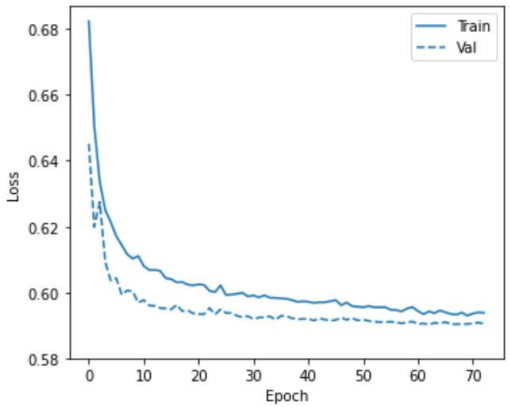


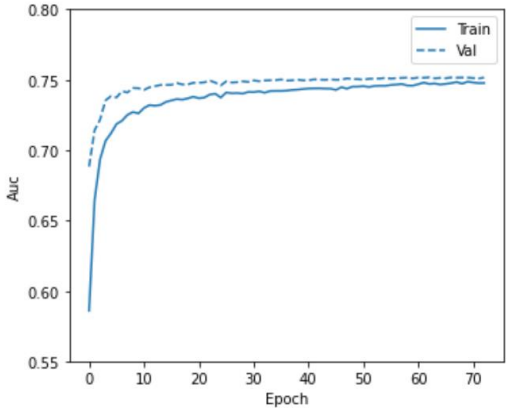
Model 4: More complicated

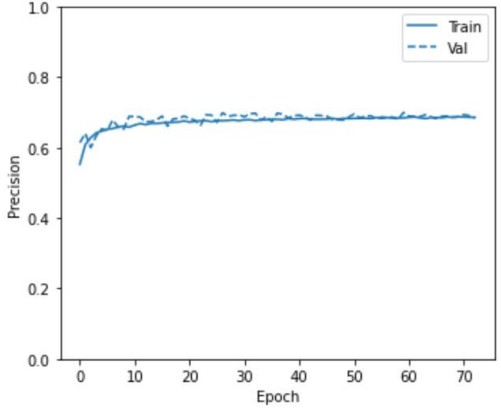
- 3 layers
- 1024 nodes (0.5 dropout)
- 512 nodes (0.2 dropout)
- 128 nodes (0.2 dropout)

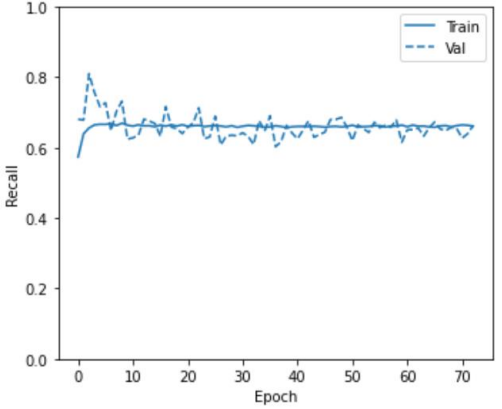
- ROC-AUC: 0.74







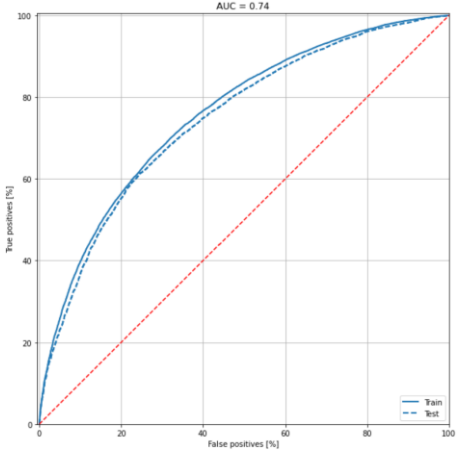


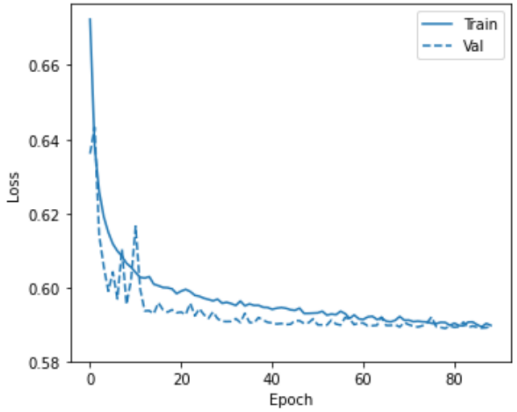


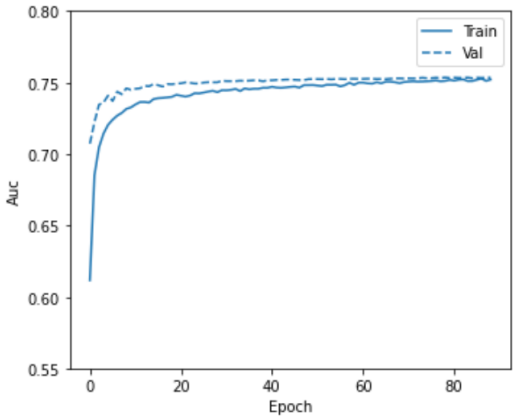
Model 5: More complicated

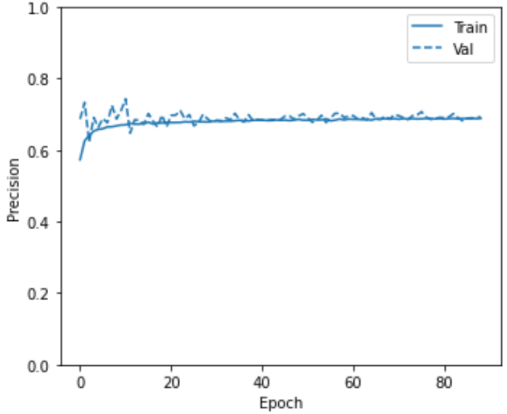
- 3 layers
- 2048 nodes (0.5 dropout)
- 1024 nodes (0.2 dropout)
- 1024 nodes (0.2 dropout)

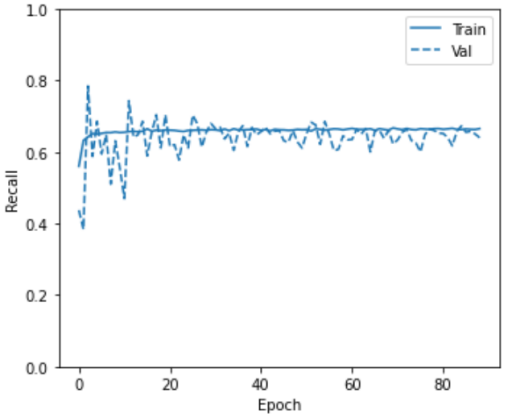
- ROC-AUC: 0.74





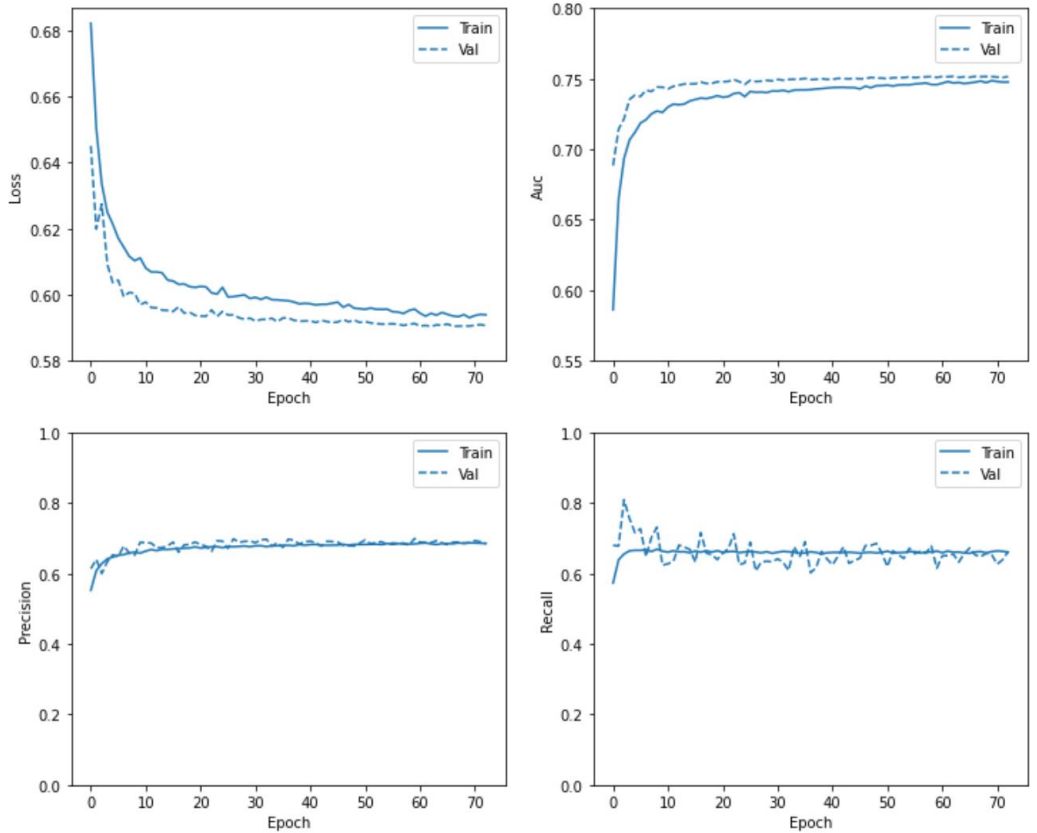






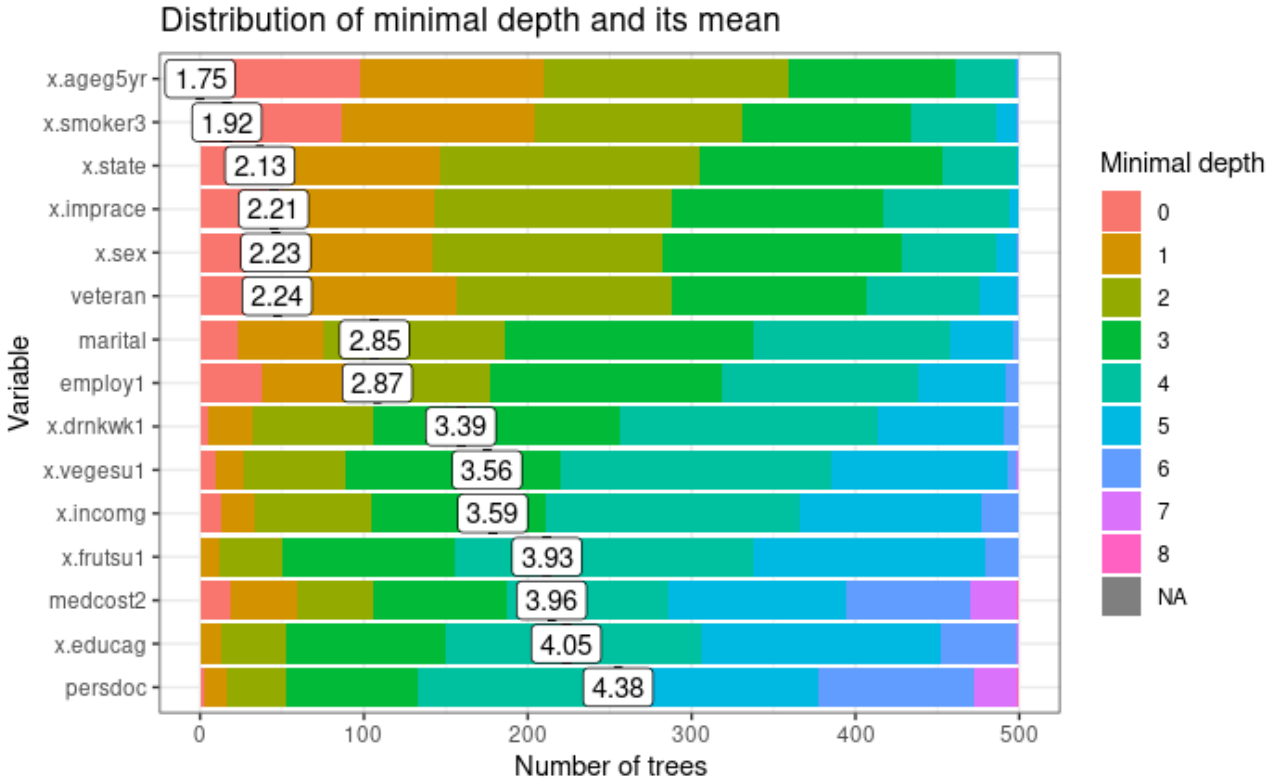
Random Forest

- Mean Accuracy: 0.641
- Mean ROC AUC: 0.694

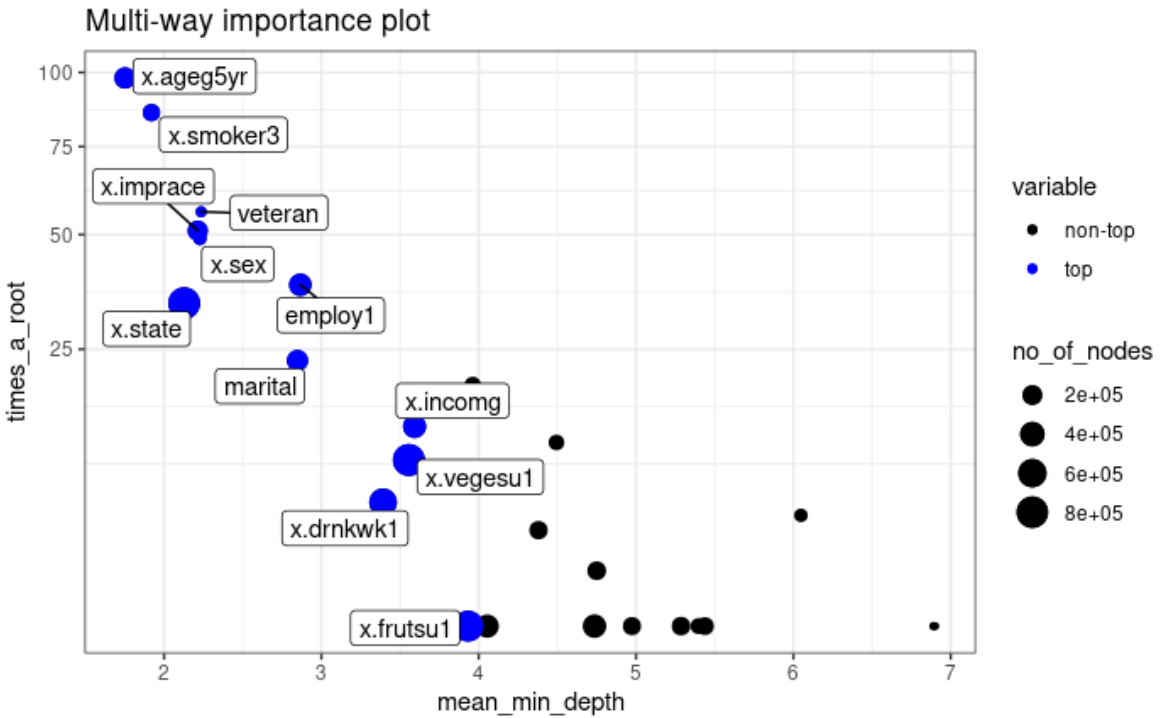


DNN results (ROC-AUC = 0.74)

Random Forest



Random Forest



Discussion and Conclusions

HIV risk behaviors

- Random forest performed slightly better than DNN in accuracy and precision
- Probably due to the imbalance nature of the data

HIV testing

- For this balanced outcome, DNN performed better than random forest

Overall, the predictions for both outcomes were not ideal.

- Little association
- Methods

Acknowledgement

Cankun Wang (TA)

Questions?