

## INTRODUCTION

The COVID-19 pandemic emerged a need for a centralized health data repository for epidemiological data. As a result, the National Institute of Health initiated the NCATS National COVID Cohort Collaborative (N3C) to create one of the largest centralized clinical and omics data repositories. It is unclear, however, if the aggregated data can be used to study and apply machine learning models for subpopulations ranging from pregnant women to those living with cardiovascular disease. In this study, we attempt to develop machine learning models to explore the convoluted relationship between COVID-19 and various epidemiological features. Specifically, our goal is to use a combination of pregnancy specific data to analyze and create predictive models for Covid-19 classification.

## Objective

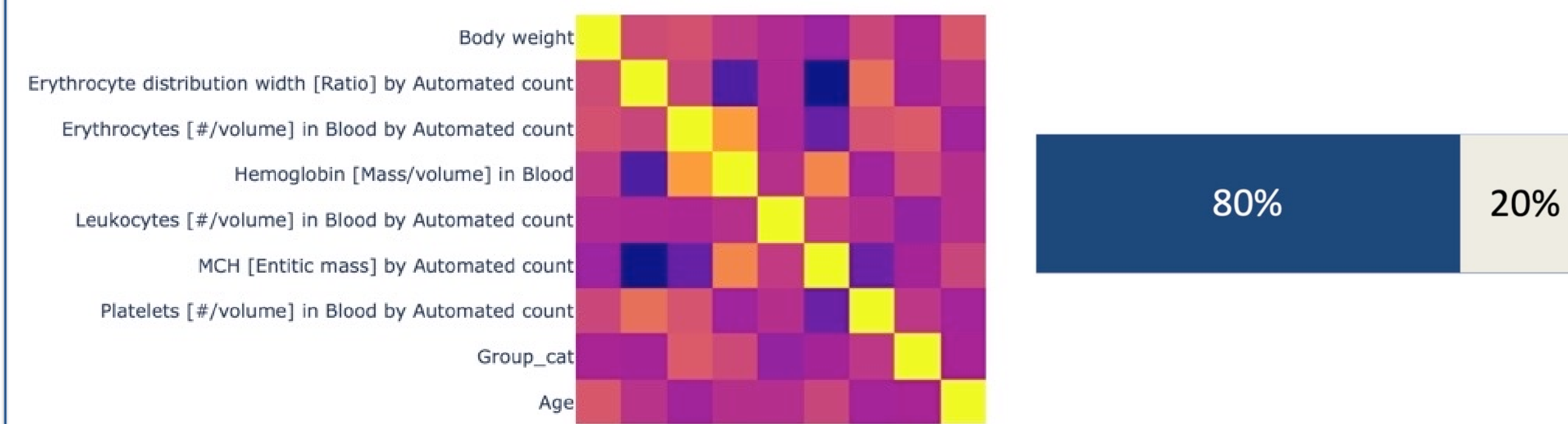
Obtain pregnancy specific data to analyze and create predictive models for Covid-19 classification

## N3C Data Enclave

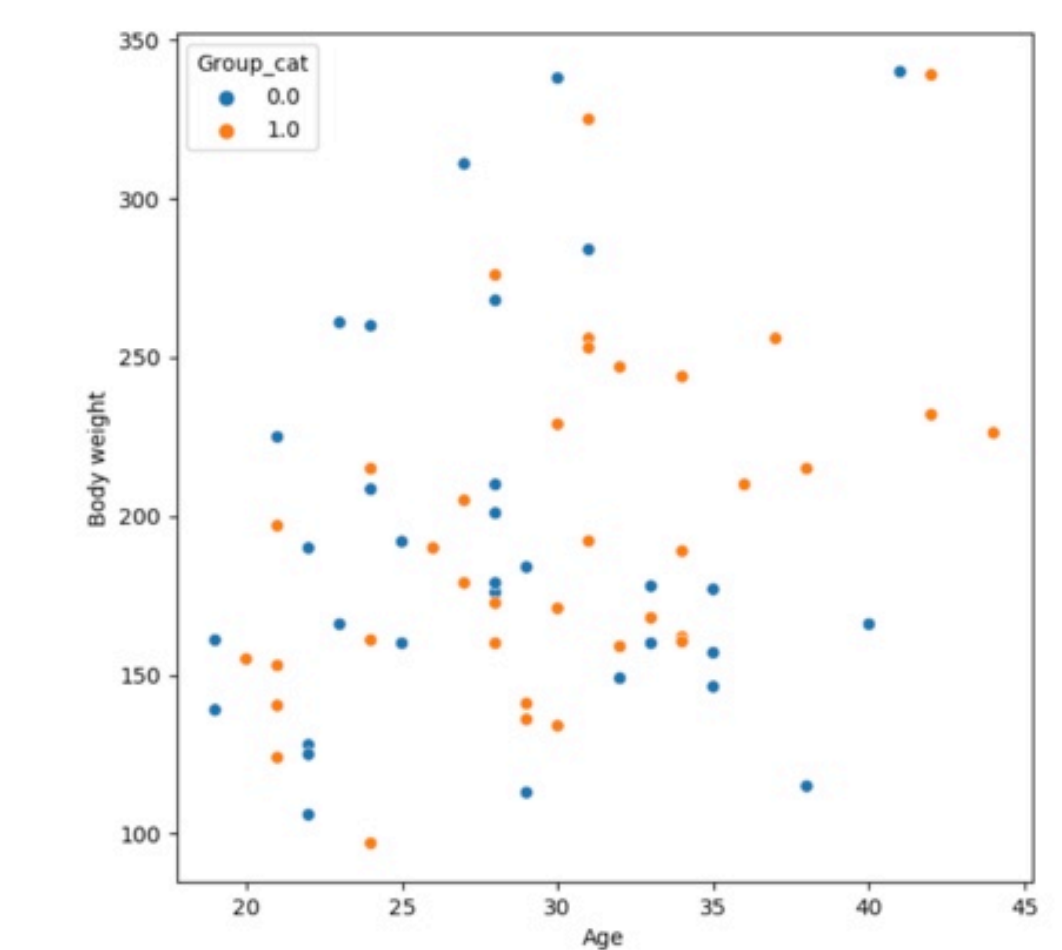
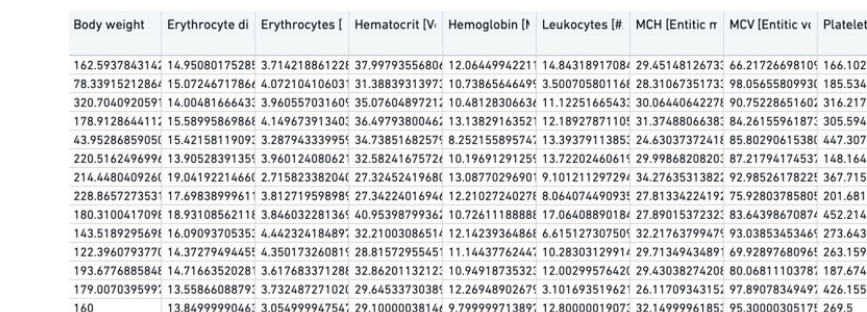


## Dataset Preparation

9 Features    95 Samples    36 Covid(+)    59 Covid(-)

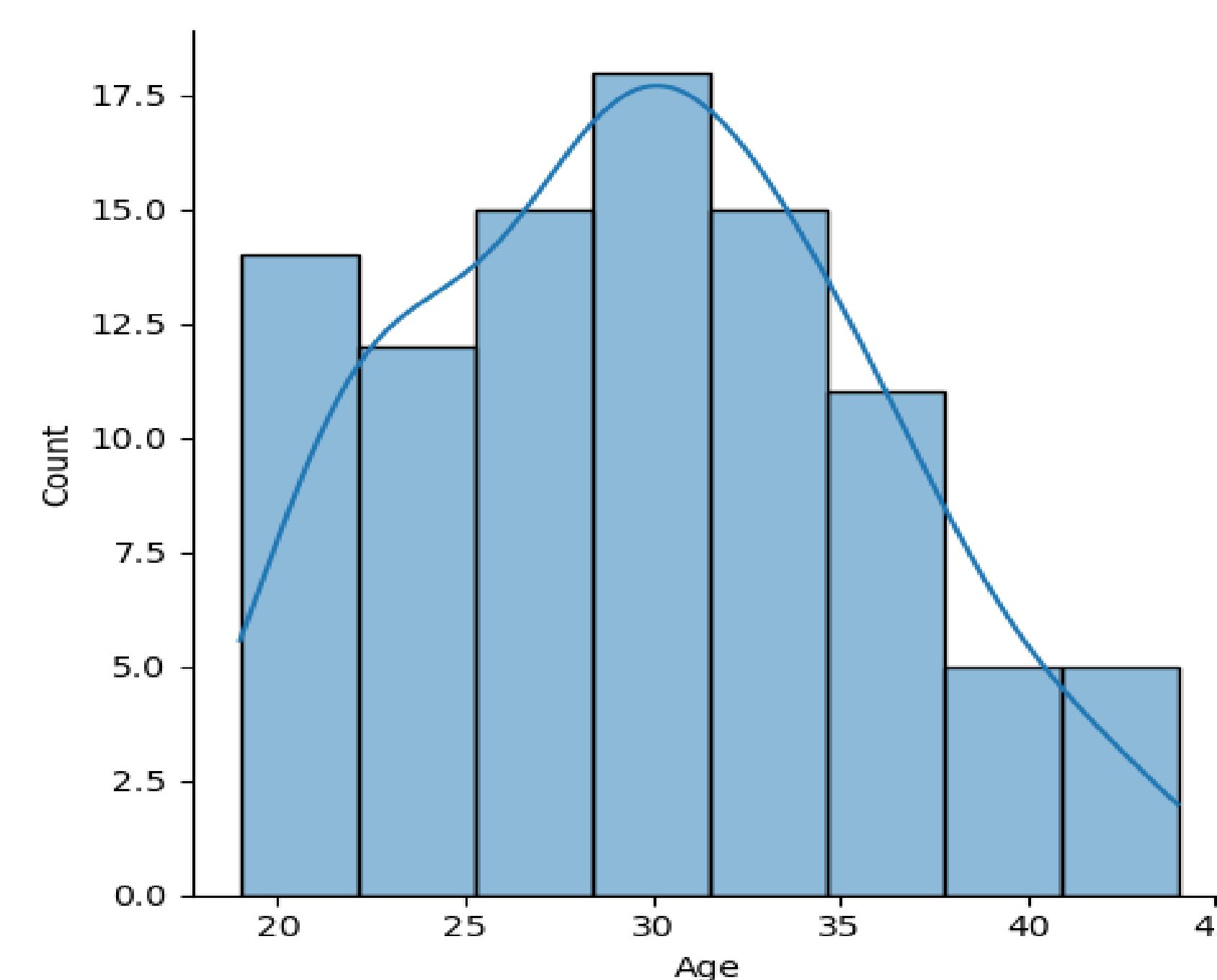


## Random Forest and KNN

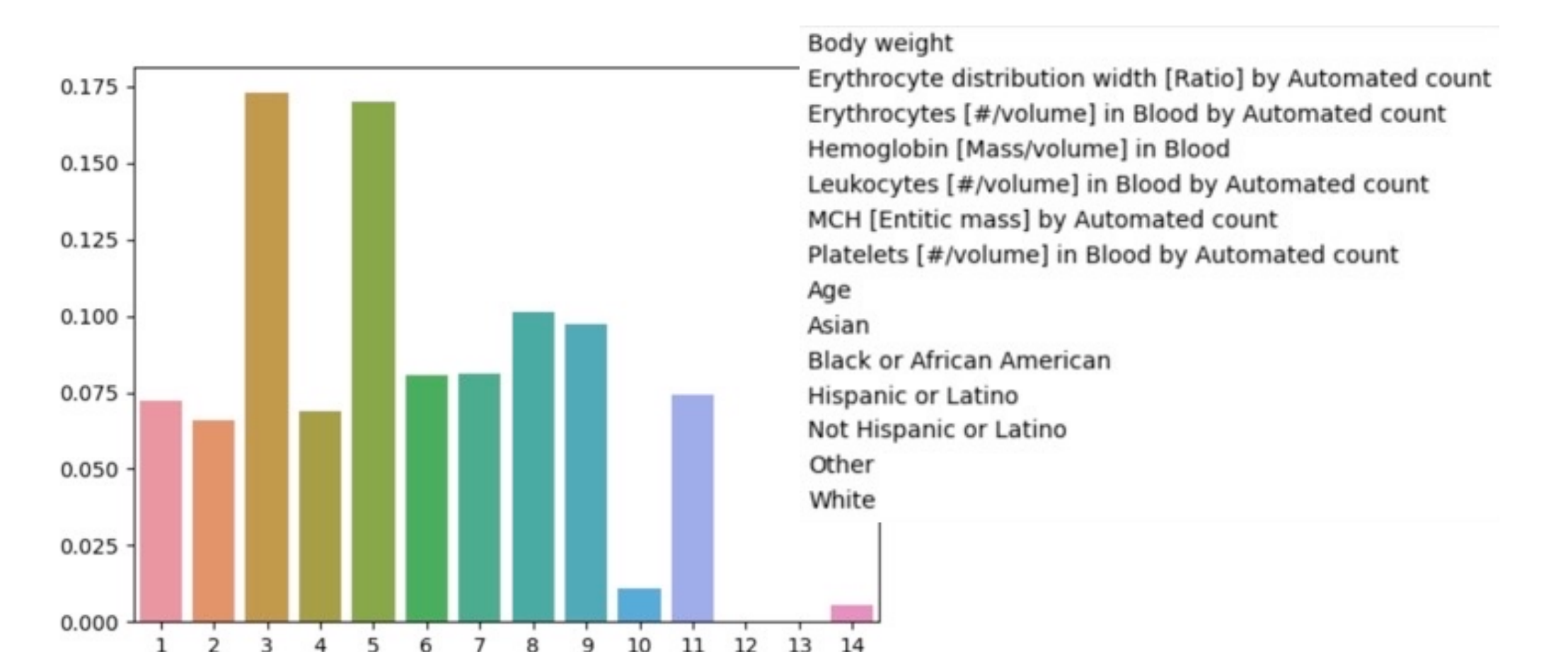


[0,1]

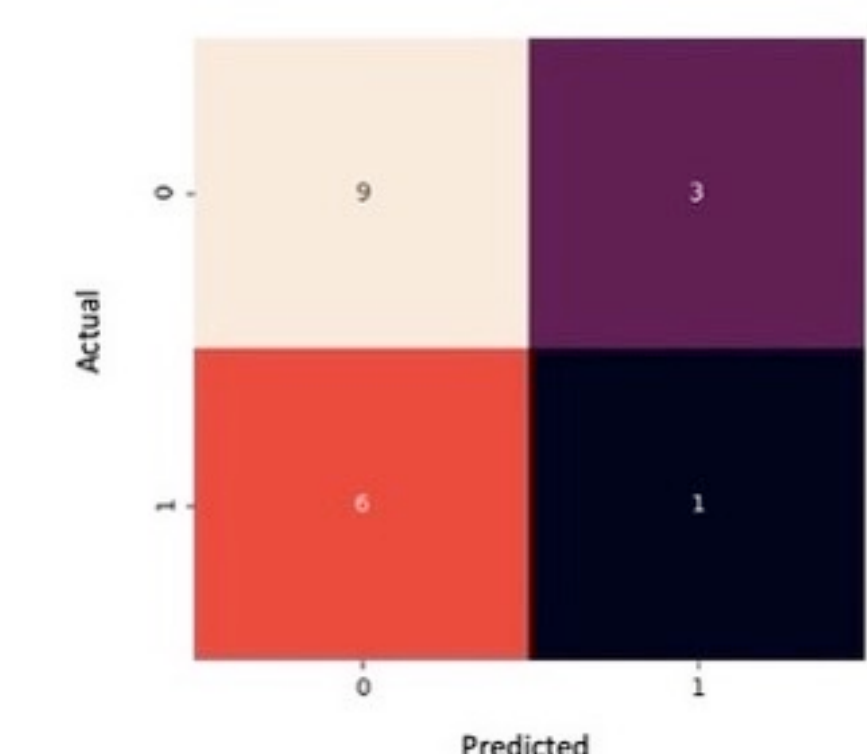
## Pregnancy Dataset Composition



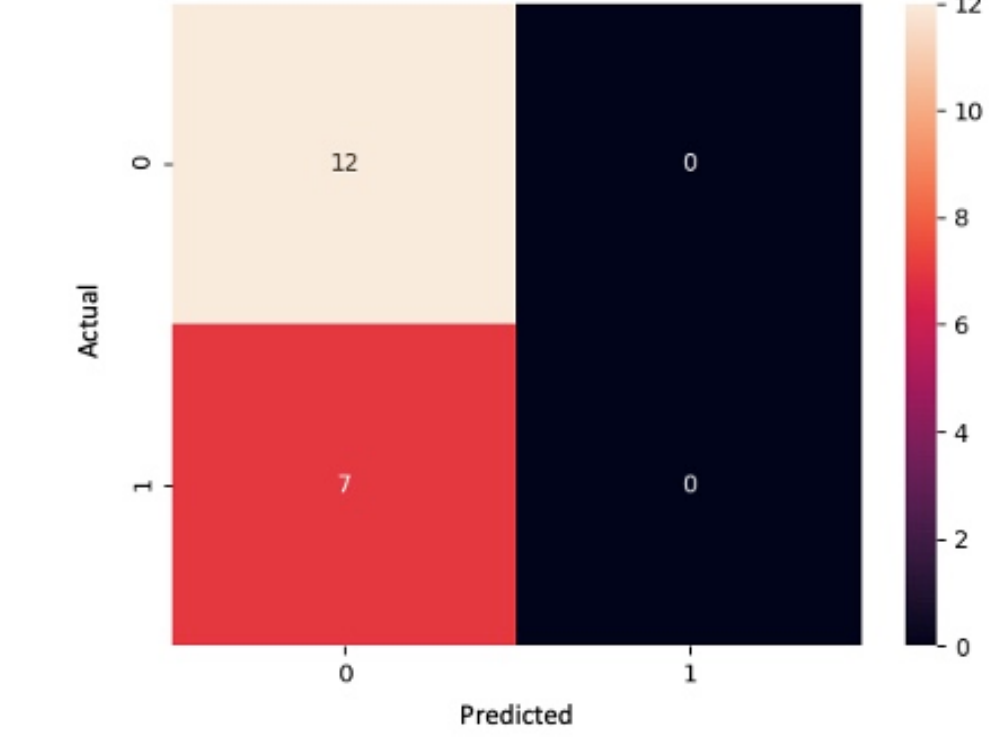
## Model Output



Random Forest Results



KNN Results



## Conclusion

- N3C Platform supports analysis and machine learning projects for study of Covid and subpopulations.
- With more time the full capability of the data can be used to reveal new insights and predictive models for target populations.

