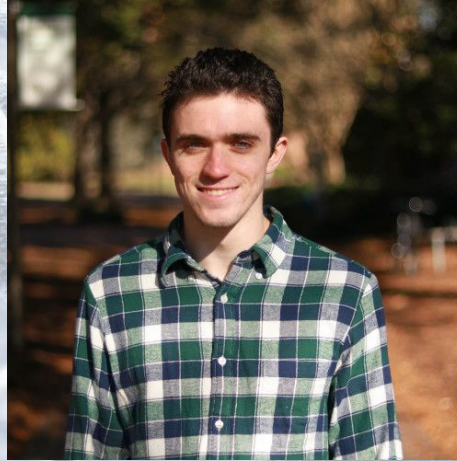# Classifying Zoonotic Viruses
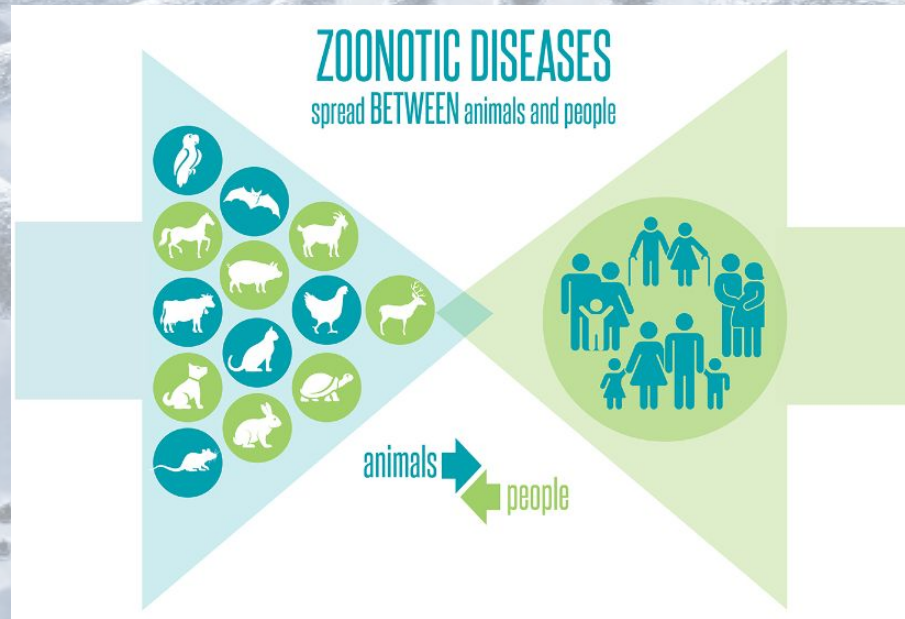
8/24/23

Connect with me

James McLaughlin
LinkedIn
GitHub

What are zoonotic viruses?

- **Caused by germs that spread between animals and people**

- **Can spread through:**
  - **Direct contact**
  - **Indirect contact**
  - **Vectors**
  - **Food**
  - **Water**

ZOONOTIC DISEASES
spread BETWEEN animals and people

animals ➤ people

Bottom line

- **Model can correctly identify associated animal about 95% of the time**

- **Quick way to make predictions for a large number of genetic sequences**

- **Useful supplement to BLAST**

Why it matters

- Identifying the animal origins of COVID is an ongoing search

- 3 out of 4 emerging infectious diseases come from animals

- Key for vaccine development and preventing future outbreaks

Data Overview

- **Data comes from ZOVER**

- **~50,000 unique viruses associated with bats, rodents, mosquitos, and ticks**

- **DNA and protein sequences**

Data Limitations

- **Incomplete sequences**

- **No numeric data**

- **Limited to 4 animals**

Model

- **Voting classifier model that combines XGBoost and Neural Networks**

- **XGBoost has higher accuracy, less overfitting**

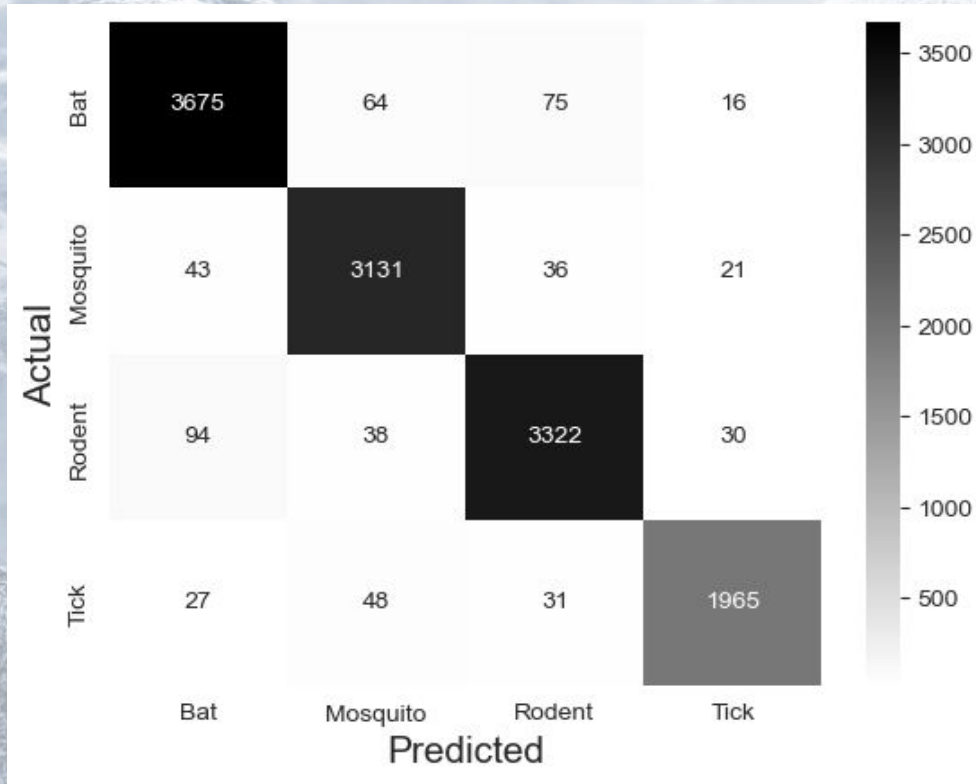- **Neural Networks identify hidden features in data**

Model metrics

Accuracy

95.4%

ROC-AUC Score

0.97

Model metrics

Sources

https://www.cdc.gov/onehealth/basics/zoonotic-diseases.html

https://www.theatlantic.com/science/archive/2023/03/covid-origins-research-raccoon-dogs-wuhan-market-lab-leak/673390/