



# Cicada Zombies



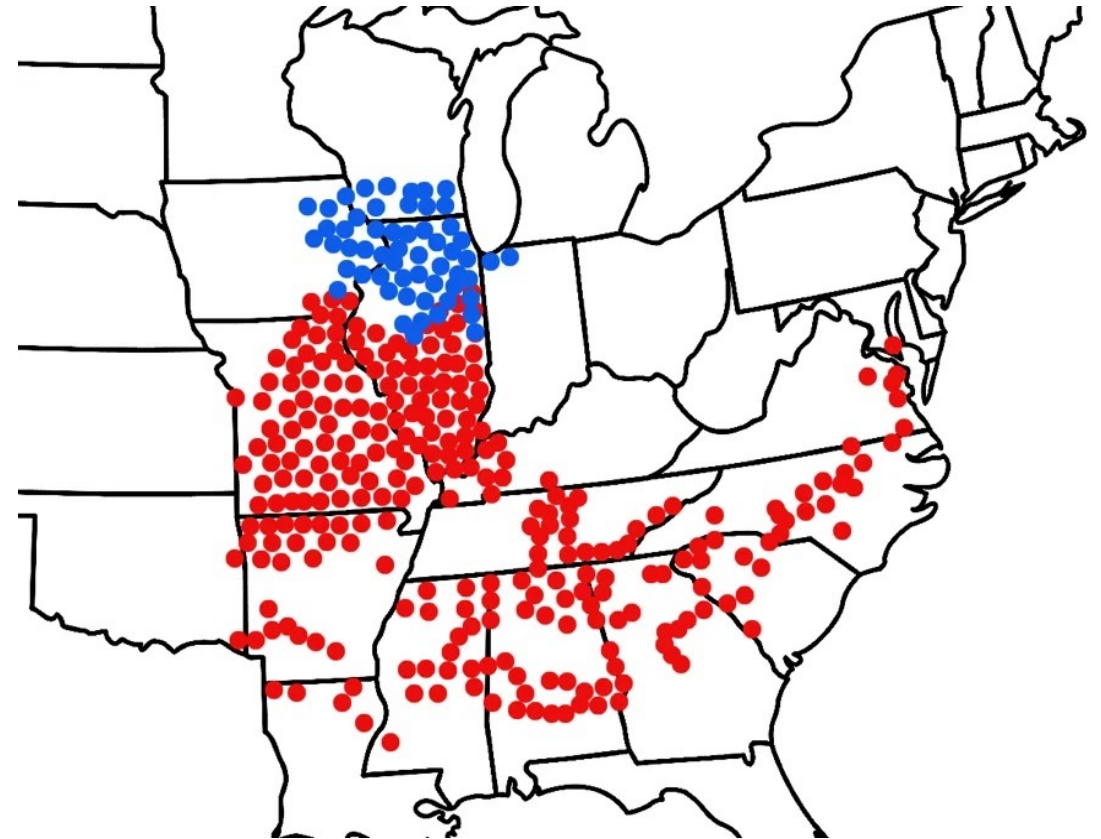
- Douglas Stauffer
- Prayagdeep Parija
- Wojciech Tralle
- Henry Tucker



# Periodical Cicadas



- Massive broods of juvenile cicadas from the genus *Magicicada* spend 13 or 17 years underground before emerging en masse for an adult stage of a few weeks.



- This year, **17-year Brood XIII** (the Northern Illinois Brood) and **13-year Brood XIX** (the Great Southern Brood) are emerging together for the first time since 1803.

# Zombie Fungus Phenomenon



- The fungus *Massospora cicadina* infects only the periodical cicadas.
- During infection, the fungus grows concealed in the abdomen of the cicada.
- Eventually the rear part of the abdomen falls off to reveal a "plug" of spore-producing fungus.
- The cicadas continue living as "zombies" while spreading the spores.

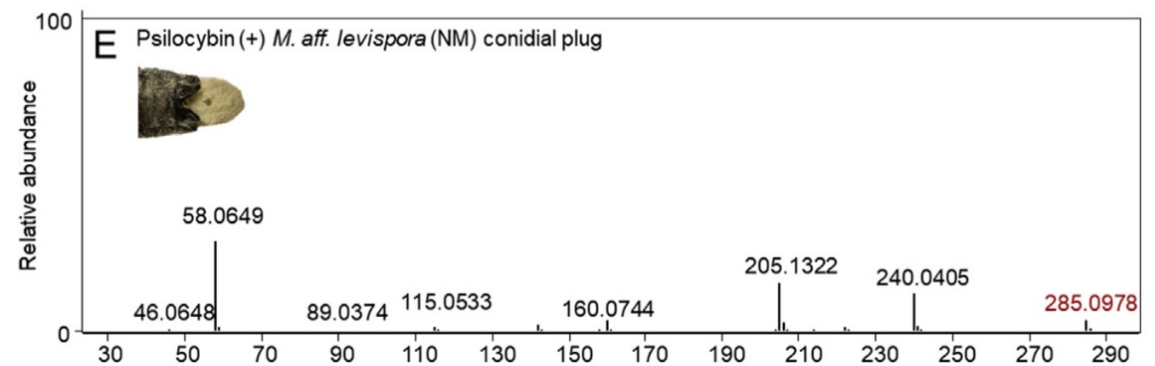
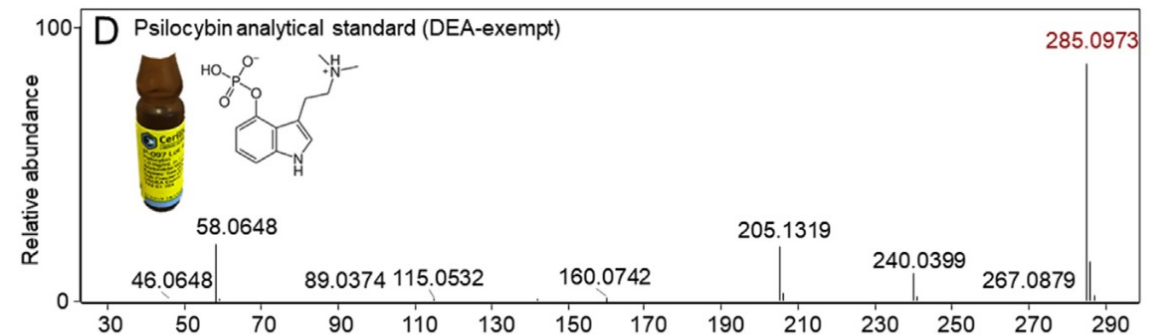


# Research Interest & Goal



## Determine infection status of cicadas via photo recognition.

- Drug research: psilocybin, the active ingredient in hallucinogenic mushrooms, is found in the bodies of infected cicadas.
- Climate change: periodic cicadas evolved millions of years ago, but the eastern US landscape is much younger due to actions of glaciers. Thus, the cicadas have already experienced major climate change.
- The emergence of the cicadas is controlled by soil temperature, a critical ecological indicator.
- General interest to entomologists and mycologists.



# Data Collection



- First data collection: about 600 infected vs 1300 healthy

- Second data collection: about 1100 infected vs 600 healthy

- Third data collection: about 600 infected vs 600 healthy



```
import requests
import os
# from website project link:
# Create a directory to store images
os.mkdir('all_images')
os.chdir('all_images')

# List of image URLs
image_urls = [
    'https://inaturalist-open-data.s3.amazonaws.com/photos/3818838/medium.jpg',
    'https://static.inaturalist.org/photos/3887251/medium.jpeg',
    'https://inaturalist-open-data.s3.amazonaws.com/photos/3901209/medium.jpeg',
    'https://inaturalist-open-data.s3.amazonaws.com/photos/3910554/medium.jpg',
    'https://static.inaturalist.org/photos/3935952/medium.jpeg',
    'https://static.inaturalist.org/photos/3961070/medium.jpeg',
    'https://inaturalist-open-data.s3.amazonaws.com/photos/3968917/medium.jpeg',
    'https://inaturalist-open-data.s3.amazonaws.com/photos/3969617/medium.JPG',
    'https://static.inaturalist.org/photos/3996546/medium.JPG',
    'https://static.inaturalist.org/photos/4081842/medium.jpeg',
    'https://inaturalist-open-data.s3.amazonaws.com/photos/6054684/medium.jpg',
    'https://inaturalist-open-data.s3.amazonaws.com/photos/8139163/medium.jpeg',
    'https://static.inaturalist.org/photos/39570000/medium.jpeg'
```

# Data Challenges!



- Inappropriate pictures for study
- Imbalanced data set
- Badly positioned cicadas
- A lot of noise in the data sets



# Data Challenge Accepted!



- Three different data sets as input: raw, cleaned and cleaned + balanced,
- We cleaned the data set by removing inadequate images
- Stratified the data set into a 20% Test, 10% Validation, 70% Train



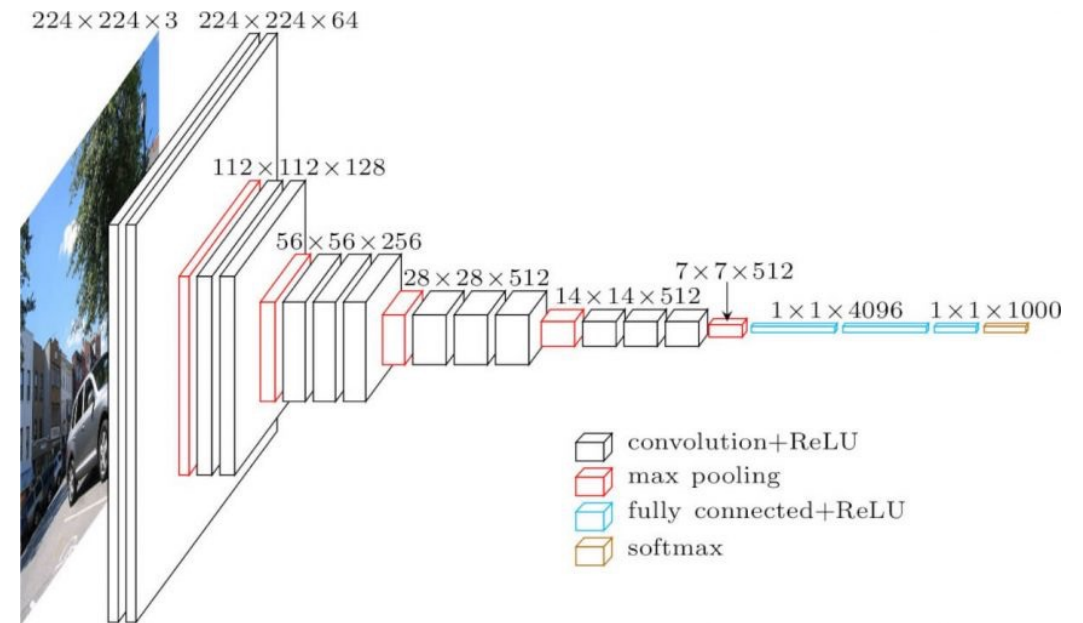
# Model Details



- We found that CNN models are the best for image classification. We trained three of them:

- VGG-16
- VGG-19
- Inception- V2

## Models-I





# Metrics Used



We don't want a diseased cicada to get misclassified. So our metrics in order of importance are:



F1 > precision > recall > accuracy

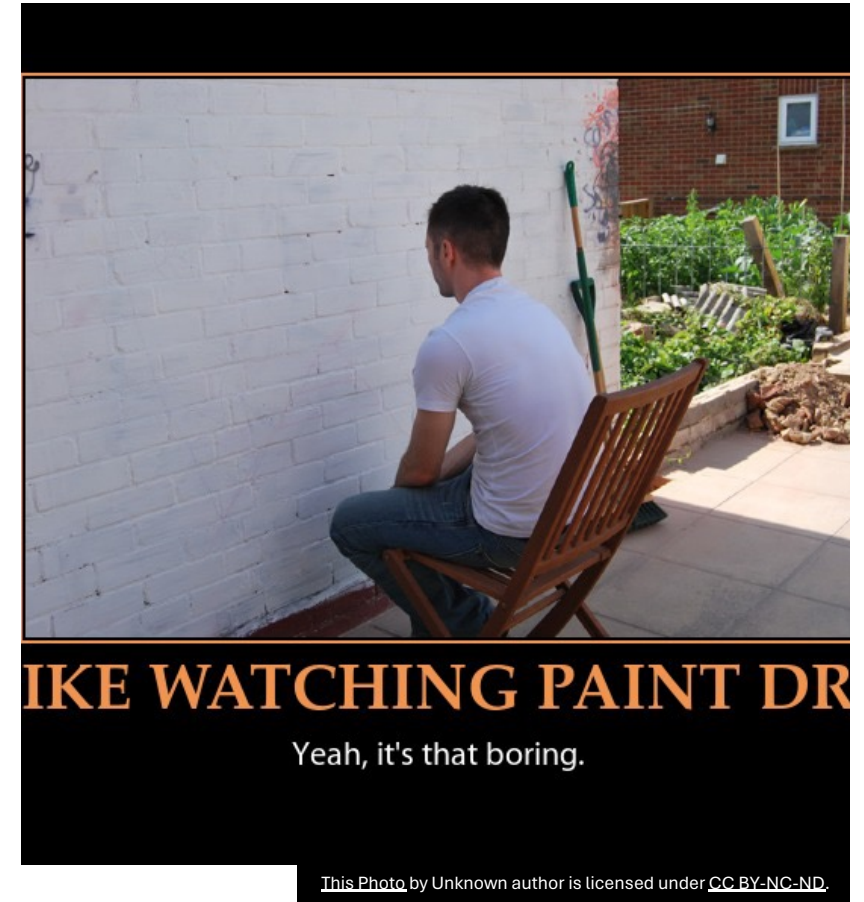


Also the runtime of these models !

# Code Debacle



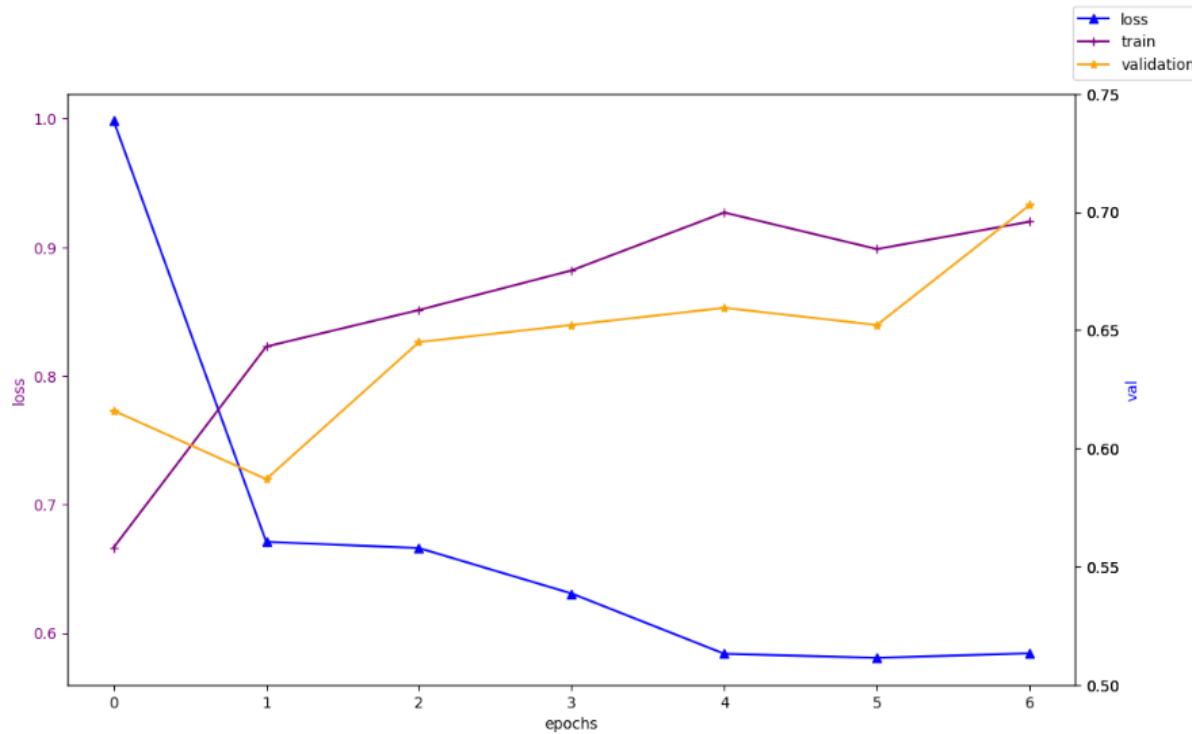
- Test results were poor compared to Training and Validation
- Training a CNN models takes a long time. Waiting for the code to run was like "watching paint dry"



# Model Train & Test Results Conflicted



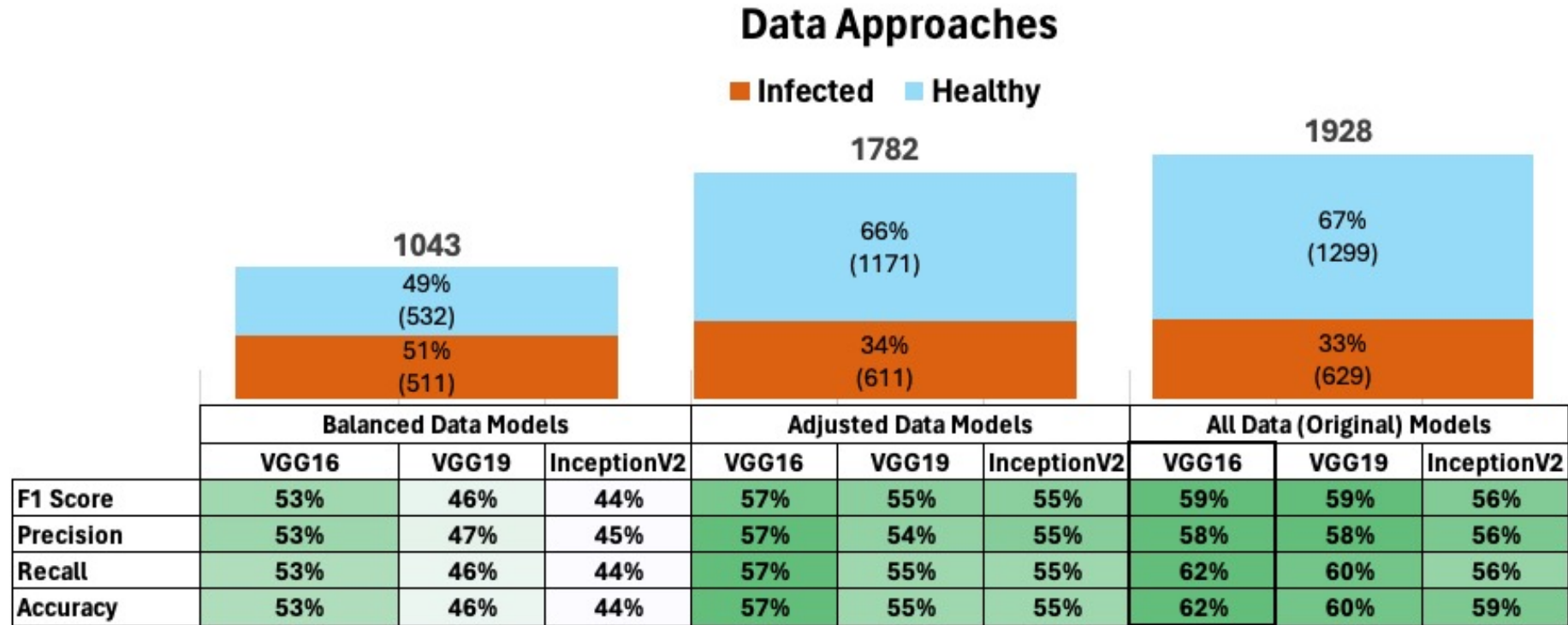
- Test results were poor compared to Training and Validation
- Accuracy was not the best measure on the best performing models
- The image data sets all had favorable accuracy results, but testing did not.



Model: VGG16 with original data set

		Actual Values	
		Positive	Negative
Predicted Values	Positive	7%	25%
	Negative	13%	54%

# Model Results and Selection



\*Additional Models were researched: ReNet50, ResNet20, and EfficientNet

- F1 Score was the most important metric for the final model selection
- Accuracy is important, but classifying a Cicada incorrectly, is very important in disease predictions
- Each data approach had relatively the same

# Insights and Implications



Find more images,  
correctly labelled,



Found 15,000 healthy  
images and  
augment/transform the  
549 diseased images and  
see what happens,



Use random cropping to  
ensure that cicadas are  
correctly positioned,



Determine  
how bioclimatic  
variables influence  
infection rates in  
cicadas,



Examine how climate  
change and warmer  
soil temperatures impact  
the timing of emergence  
of cicadas,



Development of a final  
product, possibly a  
mobile app.

**Real life implications:** If the model performed better, it could be useful for pharmaceutical companies researching spread of fungi in humans.