Microorganism Classification Project

Kai Uechi

Summary

- Create model to classify microorganisms

- Use data from microscopic images

- Final model reached 98.9% accuracy

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

- DNA sequencing accurate, but slow

- Machine learning could classify by type very quickly

- Speed is important for patients in critical condition

Data

- Continuous data extracted from microscopic images

- 10 total categories of microorganism

- Some harmful, some not

Methods

Data

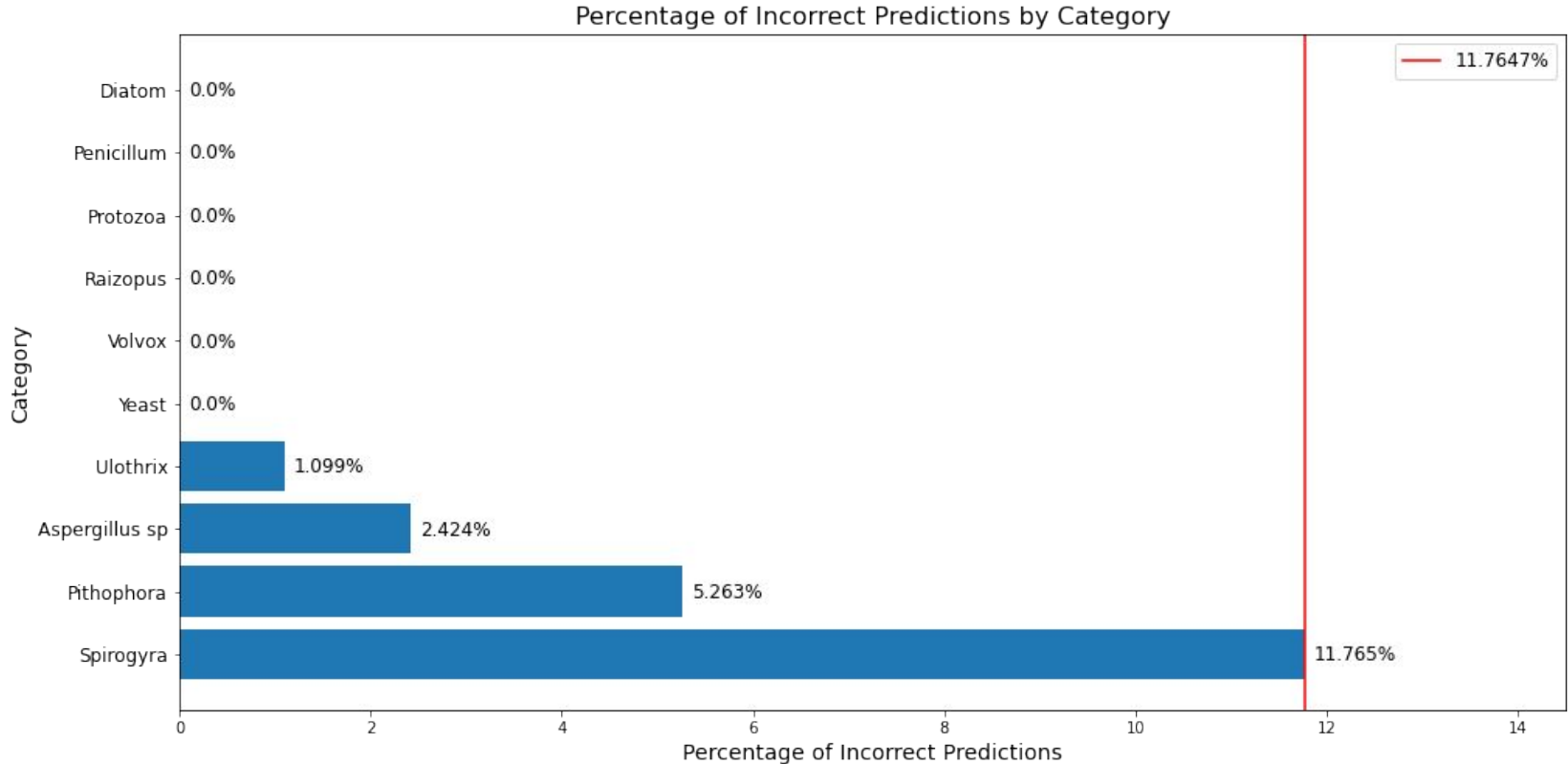- Scale data values
- SMOTE oversampling

Modeling

- Iterative modeling
  - Decision Tree
  - K-Nearest Neighbors
  - Random Forest
  - XGBoost

Results: Deploy Model For Use

- Model accuracy is 98.9%

- Would allow for quick results if deployed

- Less effective at identifying Spirogyra

Results: Deploy Model For Use

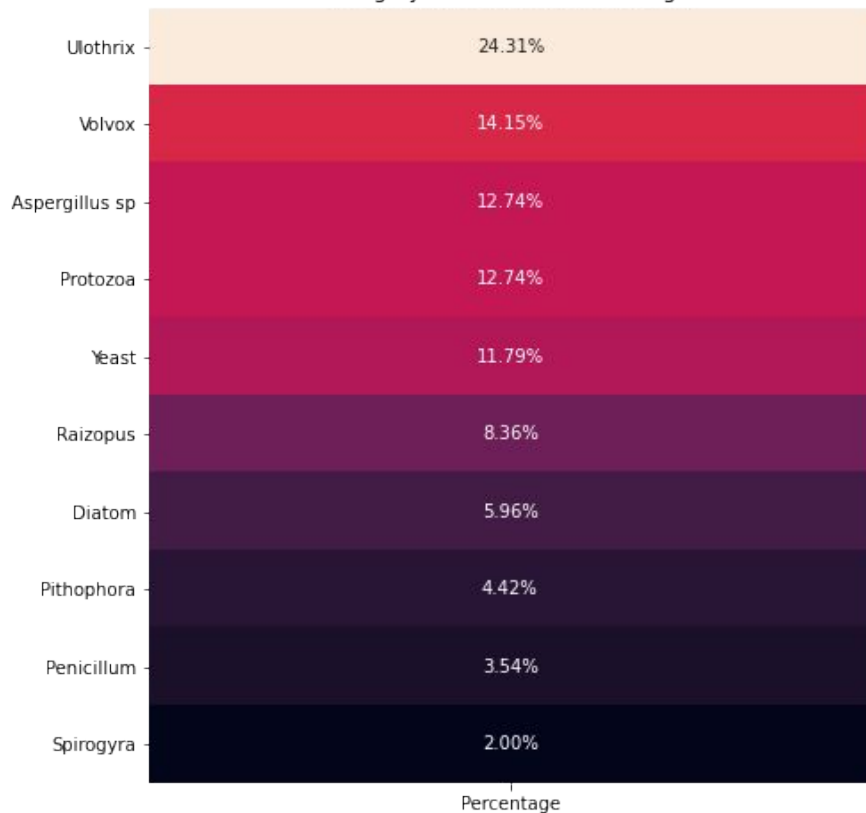Percentage of Incorrect Predictions by Category

Results: Collect More Spirogyra Data

- Spirogyra inaccuracy could be improved

- Spirogyra worst performing and has least data

- More Spirogyra data should improve performance

# **Results:** Collect More Spirogyra Data

Category Distribution (Percentage)	
Ulothrix	24.31%
Volvox	14.15%
Aspergillus sp	12.74%
Protozoa	12.74%
Yeast	11.79%
Raizopus	8.36%
Diatom	5.96%
Pithophora	4.42%
Penicillum	3.54%
Spirogyra	2.00%

Percentage

Category F1 Scores (Final Model)	
Ulothrix	0.988
Volvox	0.998
Aspergillus sp	0.985
Protozoa	0.993
Yeast	0.993
Raizopus	1.000
Diatom	0.990
Pithophora	0.966
Penicillum	1.000
Spirogyra	0.918

F1

Results: Collect Key Features With Increased Accuracy

- Model relies on some features more than others

- Most important are 'ConvexArea' and 'raddi'

- Better feature accuracy should improve performance

Results: Collect Key Features With Increased Accuracy

Final Model Feature Importance
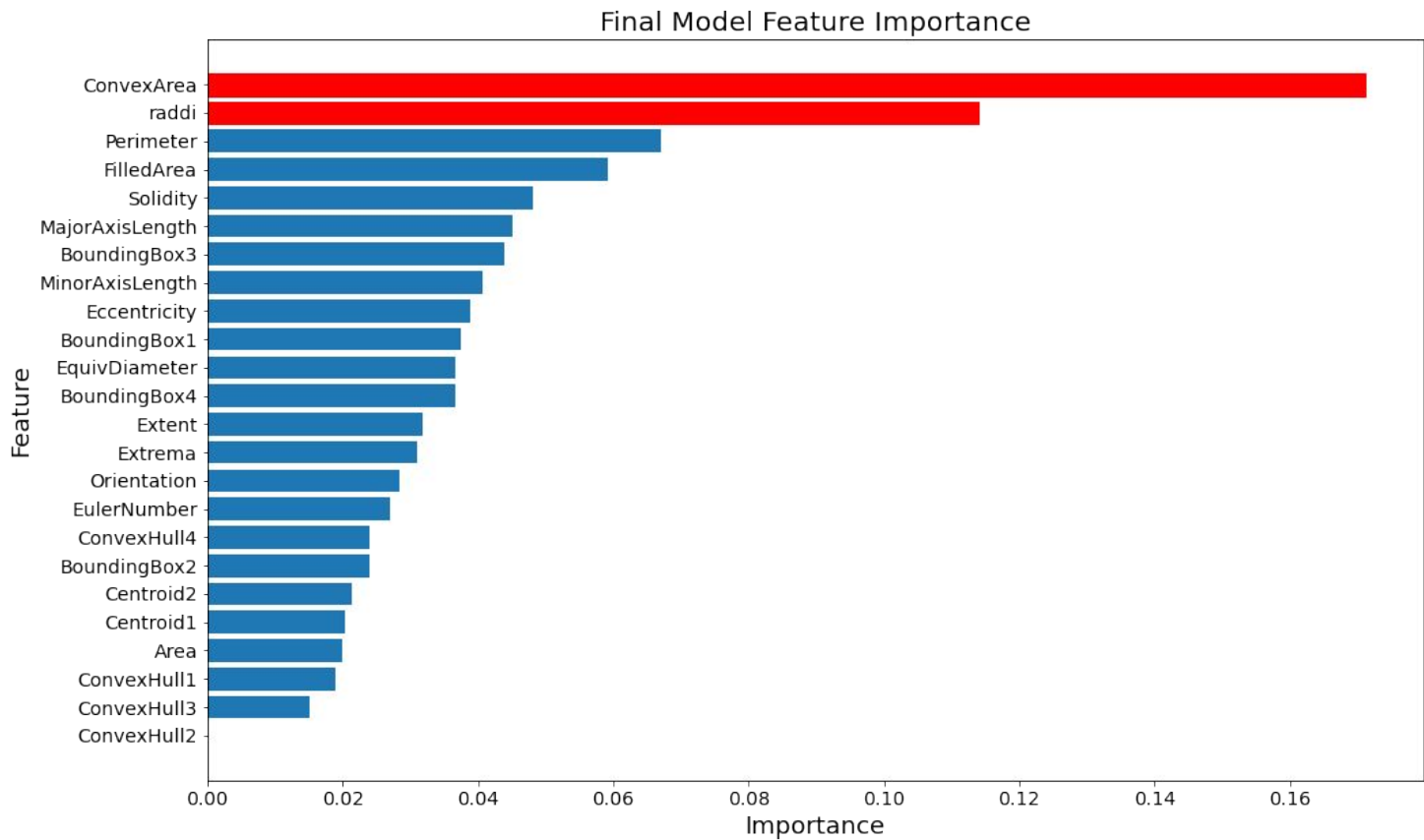
Conclusions

- Deploy model for use

- Collect additional Spirogyra data

- Collect Key Features With Increased Accuracy

Next Steps:

- Expand model categories
  - Additional categories
  - More detailed identification

Thank You!

Email: kaiuechi@gmail.com
GitHub: @kaiuechi
LinkedIn: kai-uechi