
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
 - Conclusion
-

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

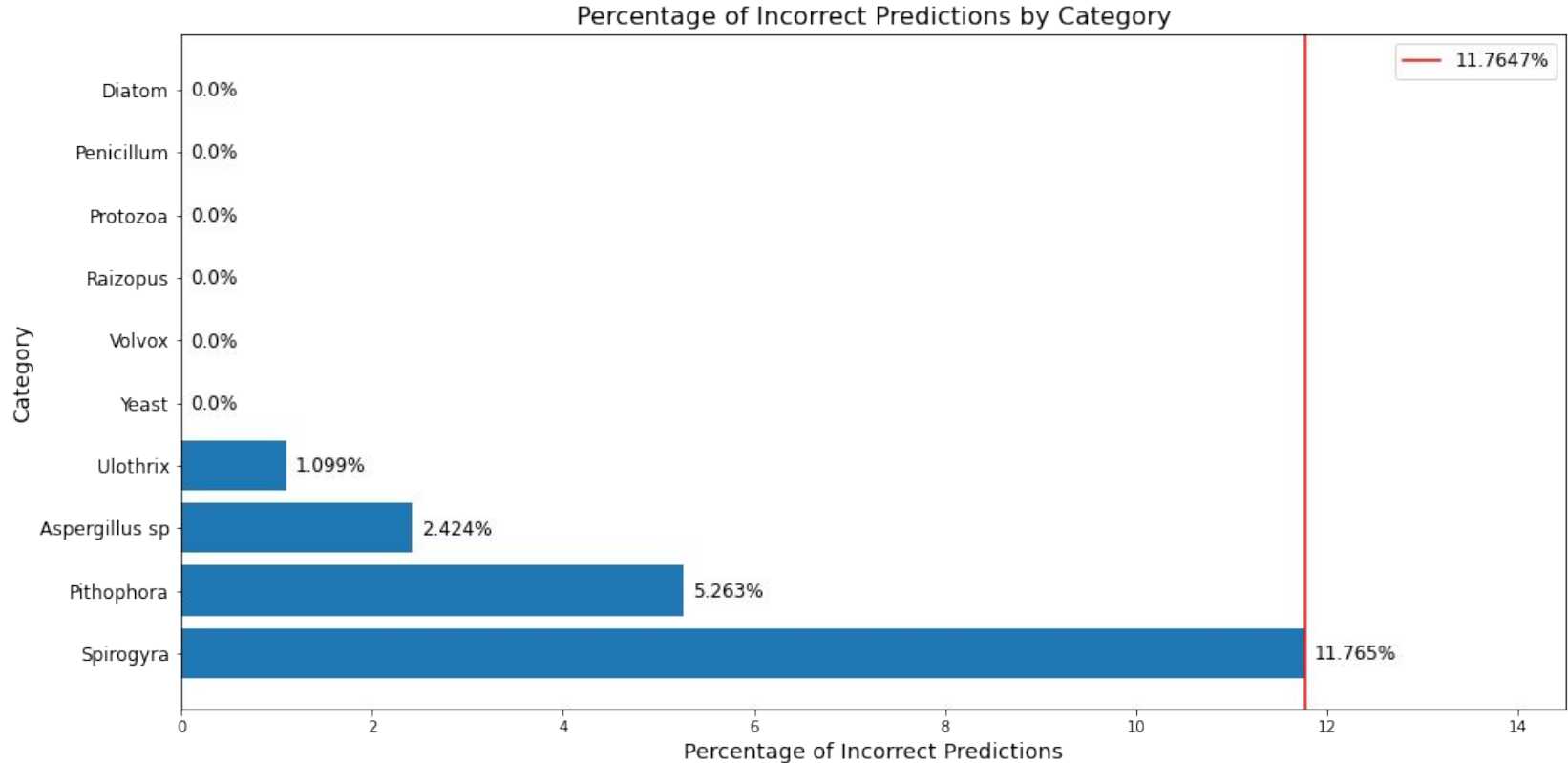
Modeling

- 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

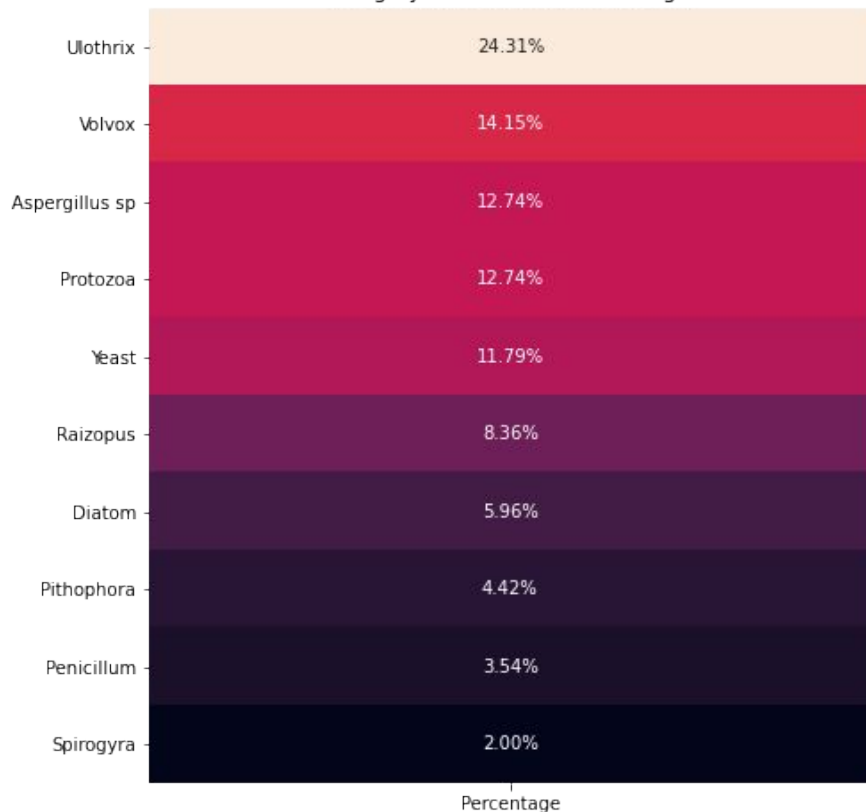


Results: Collect More Spirogyra Data

- 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)



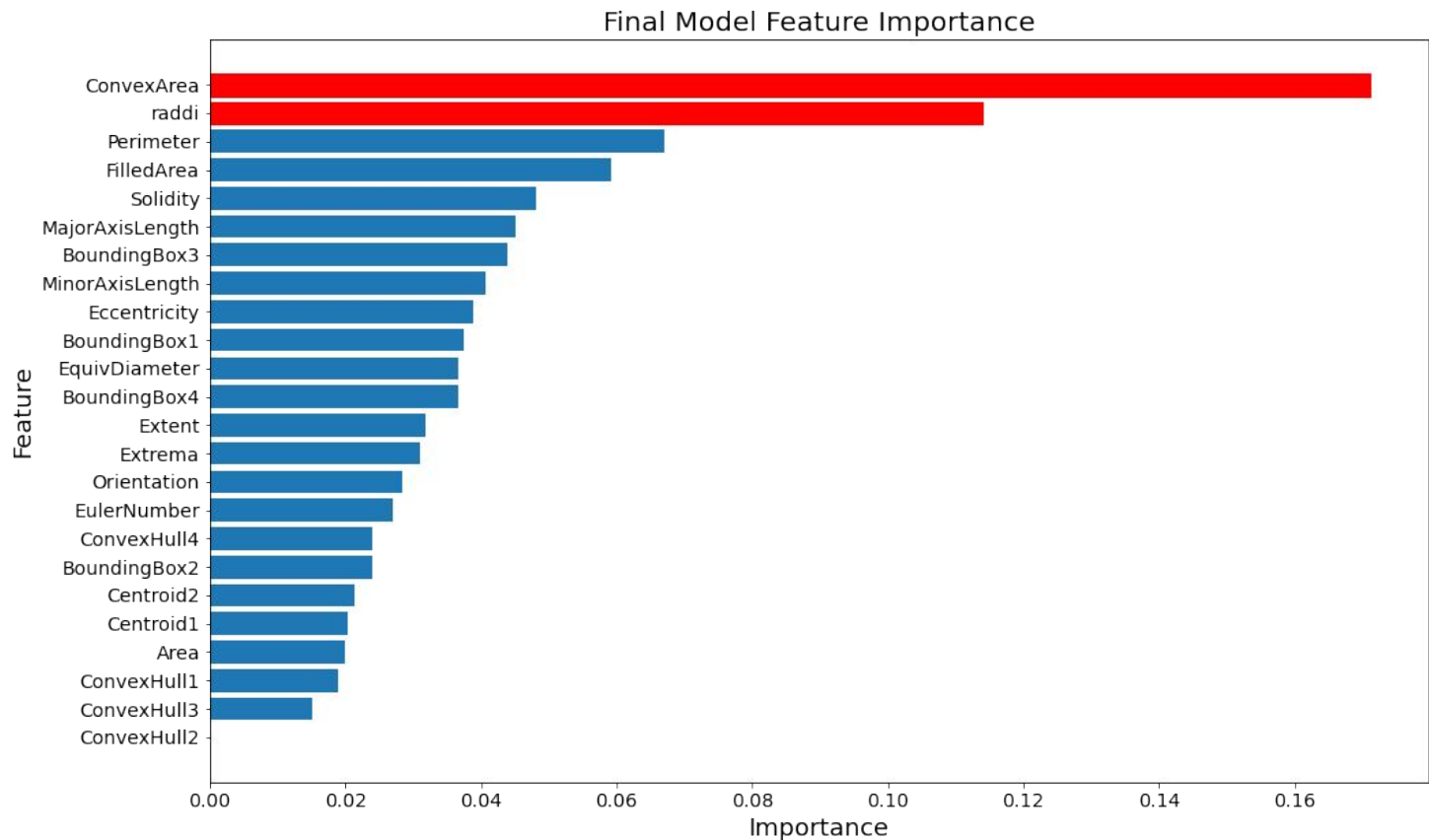
Category F1 Scores (Final Model)



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



Conclusion

- 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](https://github.com/kaiuechi)

LinkedIn: [kai-uechi](https://www.linkedin.com/in/kai-uechi)