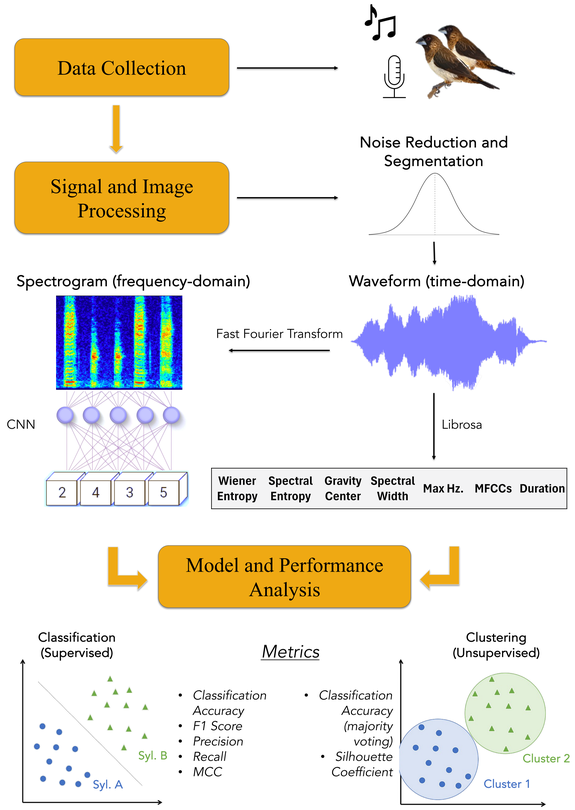
A Comparative Analysis of Machine Learning Algorithms for Birdsong Classification

Felipe Oliveira[1], Kofi Vordzorgbe[2], Sam Sober[2]
[1]Georgia Institute of Technology, [2]Emory University

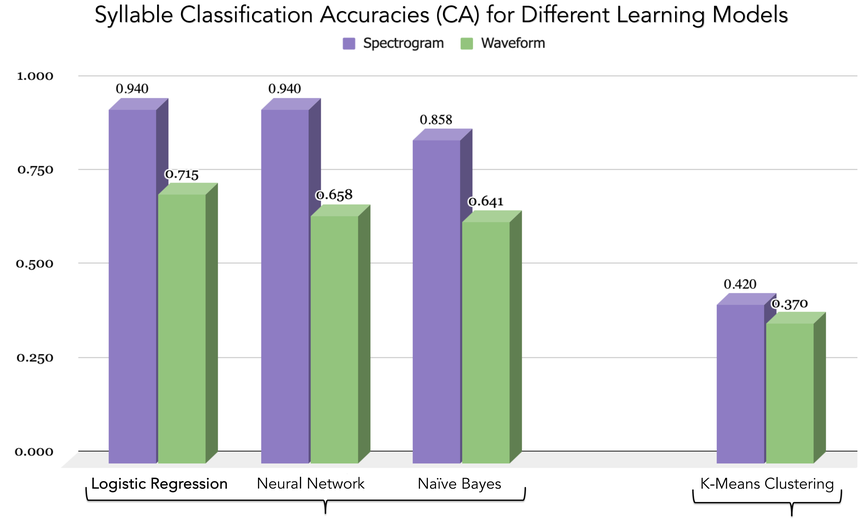
Introduction

Bengalese finches are songbirds used as a model organism for human language learning. They sing distinct syllables that we can use to measure the acquisition of vocal motor skills. However, a significant challenge lies in manually labeling the syllables in the birds' songs. This project aims to test various machine learning classification models to optimize syllable identification and reduce human bias inherent to manual labeling.

Methods

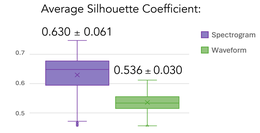


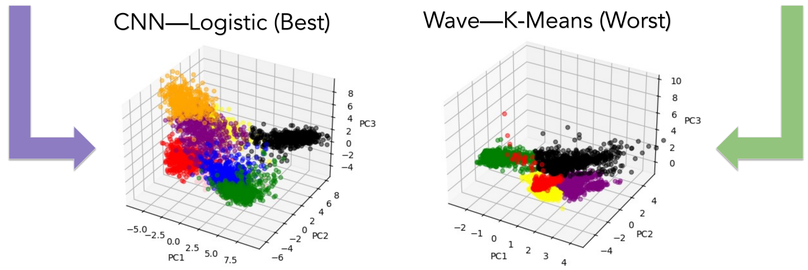
Results



Spectrogram	CA	F1	Precision	Recall	MCC
Logistic Regression	**0.940**	**0.939**	**0.939**	**0.940**	**0.924**
Neural Network	0.940	0.934	0.928	0.940	0.924
Naïve Bayes	0.858	0.882	0.923	0.858	0.829

Waveform	CA	F1	Precision	Recall	MCC
Logistic Regression	**0.715**	**0.680**	**0.680**	**0.715**	**0.595**
Neural Network	0.658	0.578	0.70	0.658	0.518
Naïve Bayes	0.641	0.648	0.662	0.641	0.513





References

1. Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. Animal Behaviour, 59(6), 1167–1176.
2. Yarden Cohen, David Aaron Nicholson, Alexa Sanchionim, Emily K Mallaber, Viktoriya Skidanova, Timothy J Gardner (2022) Automated annotation of birdsong with a neural network that segments spectrograms
3. Morita T, Koda H, Okanoya K, Tachibana RO (2021) Measuring context dependency in birdsong using artificial neural networks. PLoS Computional Biolology, 17(12), e1009707

Conclusion

- Feature extraction from spectrogram images provides more acoustic information
- Supervised classification models significantly out-perform K-Means clustering

Discussion

To understand how organisms learn language, we need a reliable way to measure acoustic similarities. The optimal model for this is pairing a convolutional neural network (CNN) for pooling spectrogram images with a logistic regression for classification. Frequency-domain matrices provide more holistic representations of birdsong than time-domain sound waves. Employing a supervised learning algorithm (i.e., logistic regression) will streamline time-consuming labeling and mitigate human biases.

Limitations

- CNNs may become computationally burdensome when dealing with larger datasets
- If using a subset of labeled syllables to implement a supervised algorithm, the researcher assumes that the bird will not introduce new syllables in the future
- Supervised learning algorithms are bird-specific

Acknowledgements

1. My mentor, Kofi Vordzorgbe, and PI, Sam Sober
2. Simon's Foundation (SURFiN)
3. Georgia Tech and Emory University

Contact: Felipe Oliveira **Email**: foliveira8@gatech.edu