Speaker Recognition with Fully Connected Networks

Task 1

Daniel Troya Atancuri[1]

[1]School of Mathematical and Computational Sciences
Yachay Tech University

Neural Networks

Introduction

In this project the problem is audio classification. We use MLP neural networks to classify different voices. For this case we selected 5 speakers and took audio samples. Then the model learns to recognize and classify which person the voice belongs to.
About the dataset, there are not many problems. Each speaker has around 1500 samples, so the data is quite balanced. Only one speaker has a little less, but the difference is small.

Methodology

We planned and followed a structured workflow to avoid mistakes when programming the model:

- Import the needed libraries.
- Define the paths for each speaker's audio samples.
- Parse the data into variables X and y depending on the method.
- Split the data into training and testing sets (to avoid overfitting).
- Define and train the MLP model with 4 layers (Input, 150, 68, Output).
- Calculate the metrics.

Results

The results for each audio representation are similar, but there are differences in the number of correct predictions. The combination of metrics (Accuracy, F1, AUC) helps us understand the performance better.

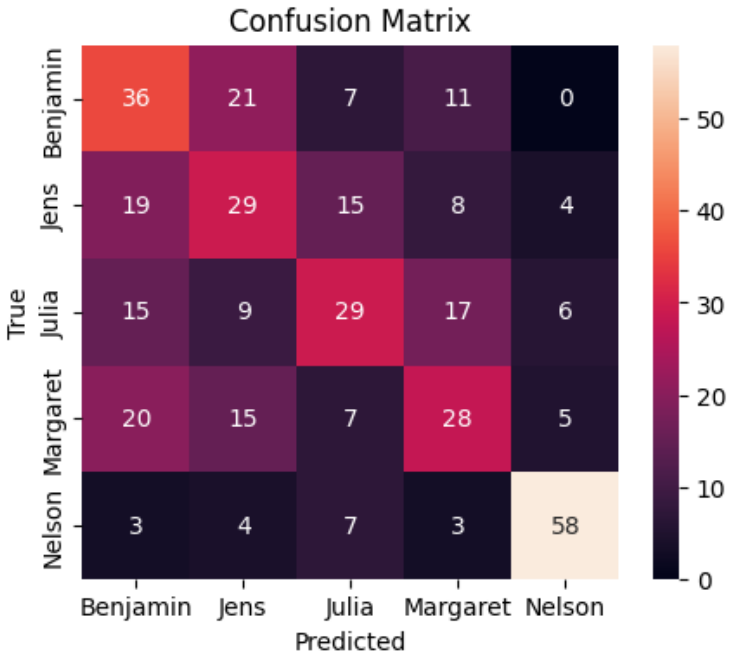


Figure: Time Domain

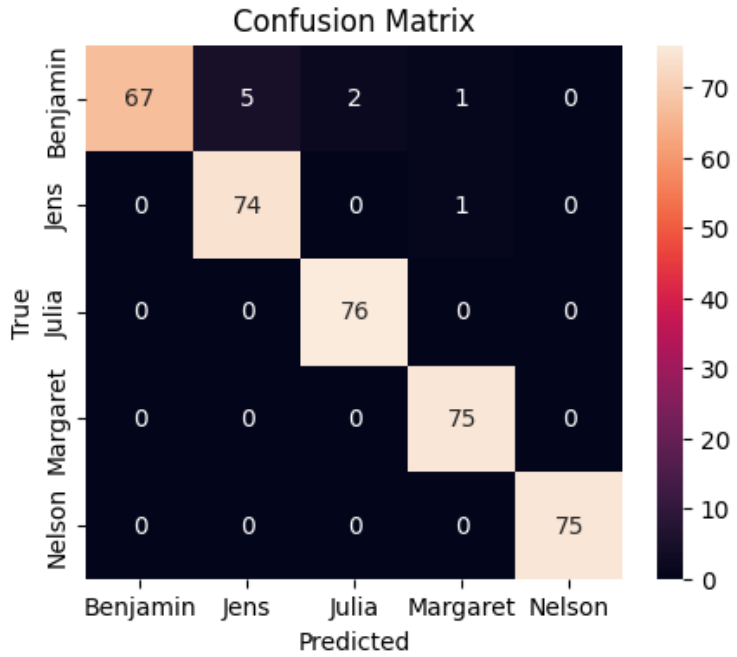


Figure: Frequency Domain

Results

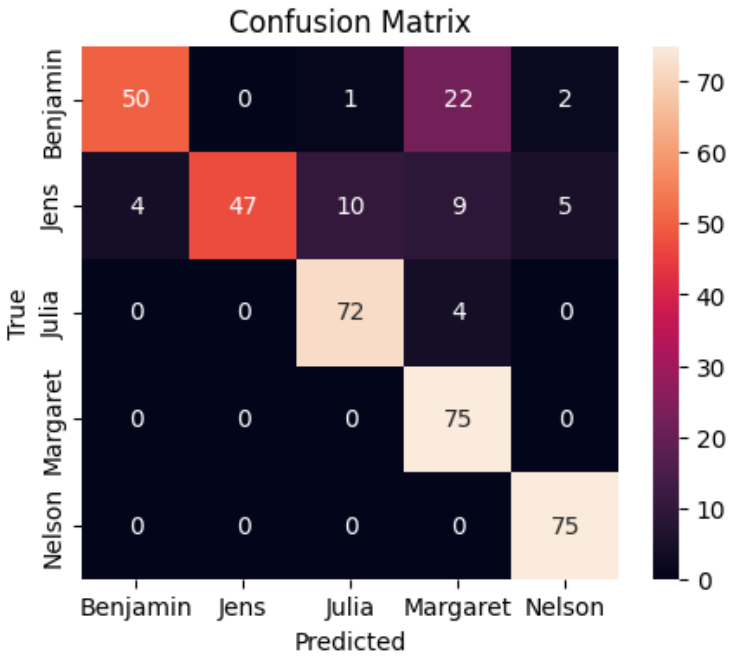


Figure: Spectrograms

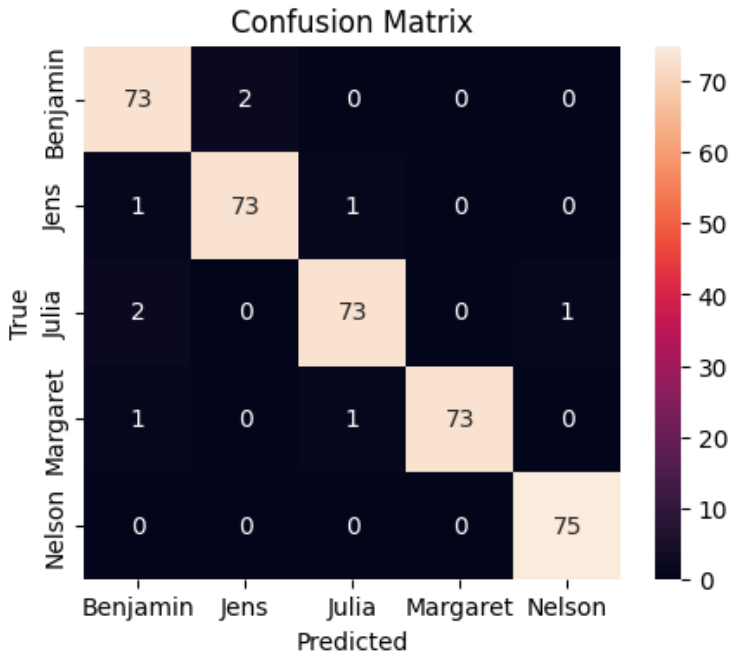


Figure: Mel Spectrograms

Results

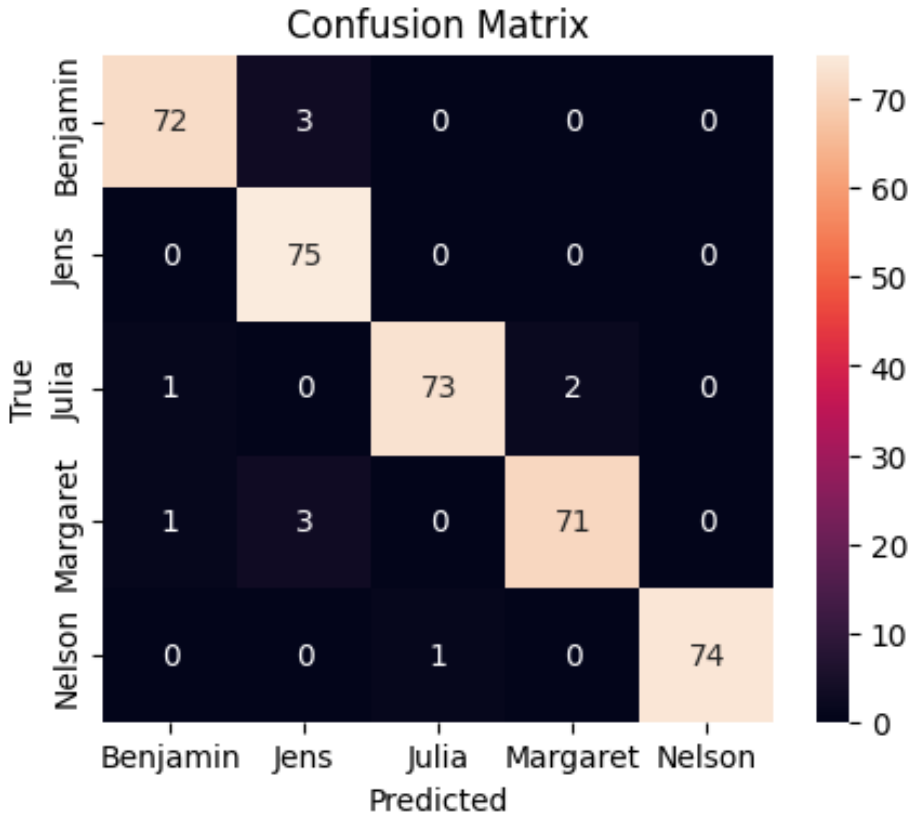
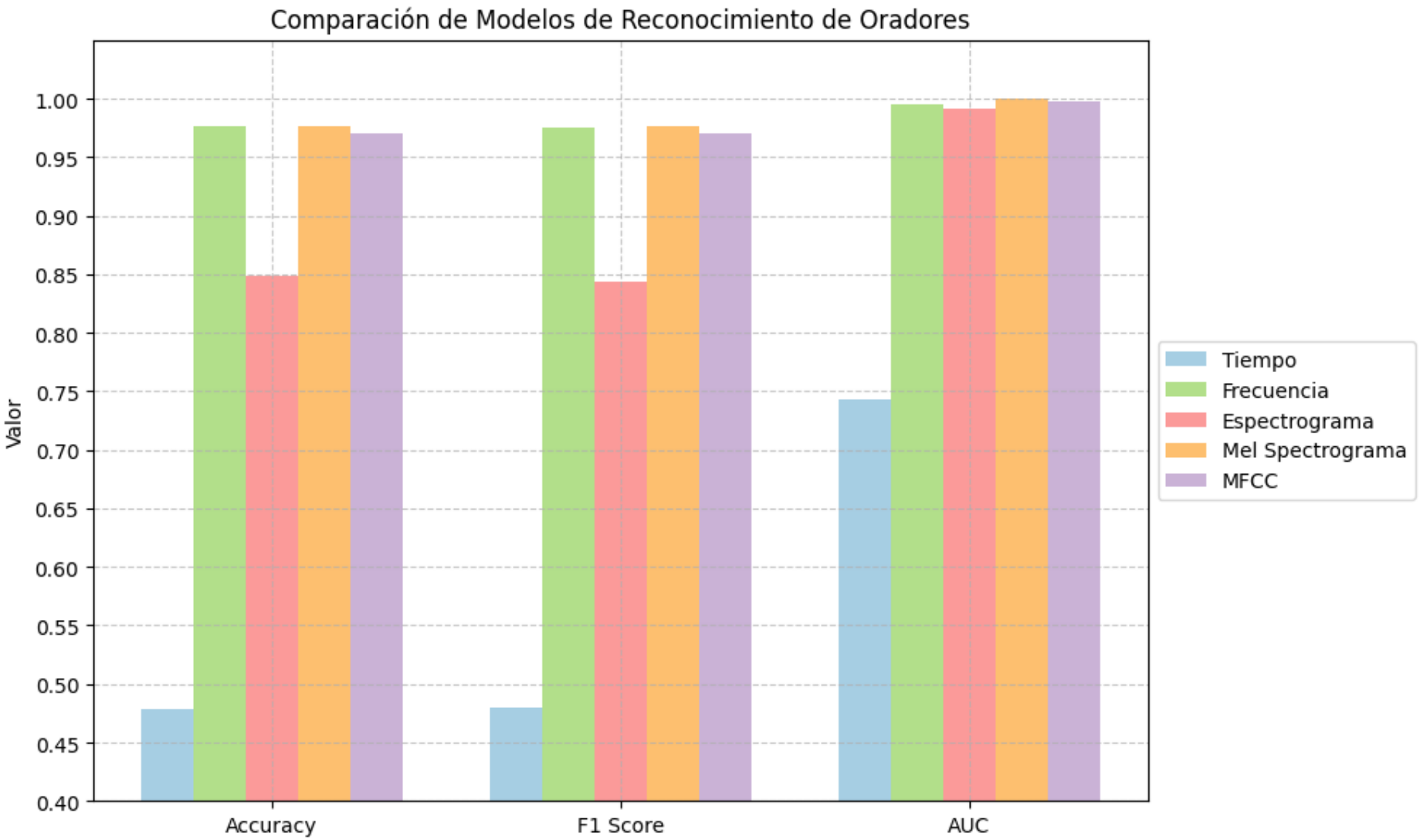


Figure: Cepstral Coefficients

Final Results



Conclusions

In summary, different audio representations can change the results because the model interprets the data in different ways. Also, the configuration of the model (layers, activation functions, amount of data) can affect the final performance. MFCCs are the most effective features, but the dataset size and balance are still important limitations.