ML Audio Classification of Parent-Infant Interactions

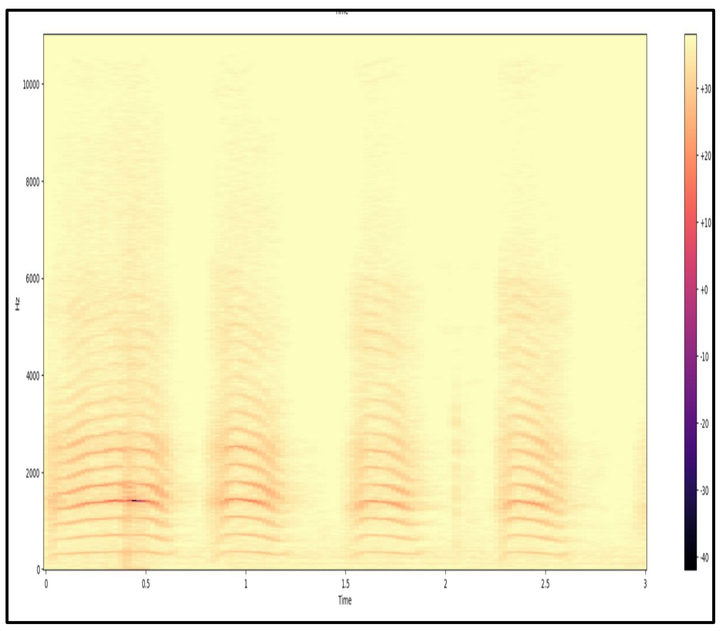
Analyzing Vocal Patterns within Latinx Families for Early Autism Detection

Presented by: Gwen Powers, Elisabeth Waldron, Will Sivolella | Mentors: Aiying Zhang, Jessica Zhang | Sponsor: Dr. Michaela DuBay, UVA STAR

UNIVERSITY OF VIRGINIA DATA SCIENCE INSTITUTE

INTRODUCTION

This project leverages Convolutional Neural Networks (CNN) to classify vocal interactions between parents and infants within the Latinx community to aid early autism research at UVA STAR. By processing audio recordings to differentiate between adult and infant vocalizations, this study aims to improve research efforts to evaluate conversational patterns so as to better inform culturally significant early autism intervention.



METHODS

Convolutional Neural Networks (CNNs):

- Utilized a ResNet50 model pre-trained on ImageNet, fine-tuned for classifying audio into 'infant', 'parent', and 'other' vocalizations.

Spectrogram Analysis:

- Employed the Short Time Fourier Transform (STFT) to convert audio signals into spectrogram images, which display frequencies vs. time with color intensity indicating amplitude.
- This visual representation is essential for CNNs to identify unique vocal signatures.
- Converted spectrograms to a decibel scale and normalized image matrix values to range between 0 and 1 to prepare them for effective CNN analysis.

Supervised Learning:

- Applied supervised learning techniques using labeled audio data, enhancing the training efficiency and performance of the model.
- This approach ensures the model learns to accurately differentiate between labeled categories (infant, parent, other).

DATA ANALYSIS & PROCESSING

Data Collection:

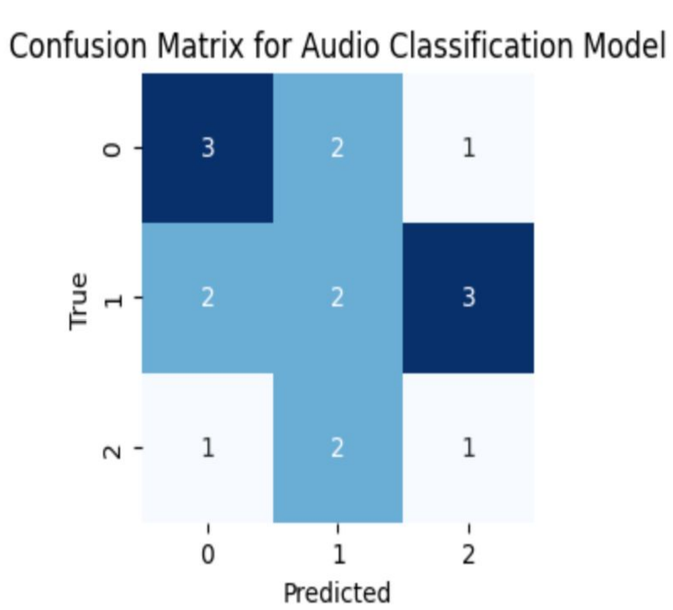
- **Source and Format:** 200+ hours of daily vocal interaction audio from 32 volunteer Latinx families, with 72 hours utilized at the time of analysis.
- **Segmentation and Labeling:** Segmented MP3 audio clips into 3-second slices, labeled as 'infant', 'parent', or 'other' to prepare data for supervised learning.

Training and Validation:

- **Configuration:** Configured batch size of 1 and resized images to 224x224 pixels to match the ResNet50 input specification.
- **Epochs:** The model was trained over 10 epochs, using real-time accuracy tracking.

Statistical Analysis and Visualization:

- **Accuracy Metrics:** Monitored and reported specific metrics.
- **Confusion Matrix:** Evaluates model precision across categories.
- **Visual Tools:** Utilized plots to track training/validation accuracy and heatmaps to visually represent the confusion matrix, aiding in the clear presentation and interpretation of the model's performance.



RESULTS (at time of printing)

Classification Accuracy:

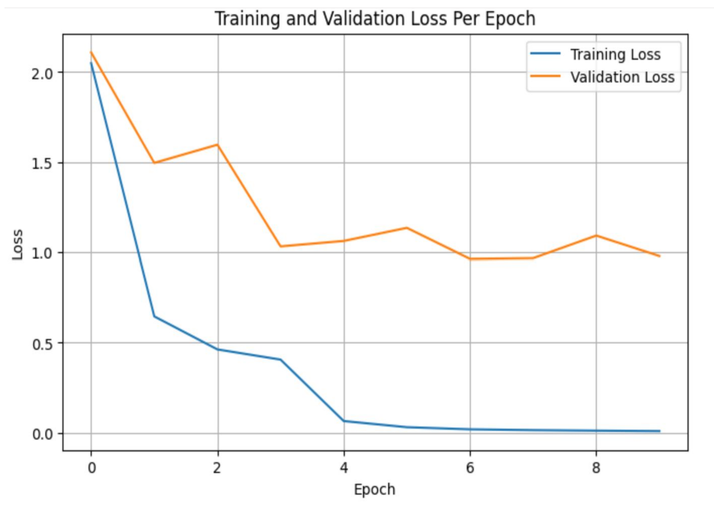
- Validation Accuracy of 76.47%
- Training Accuracy of 100%

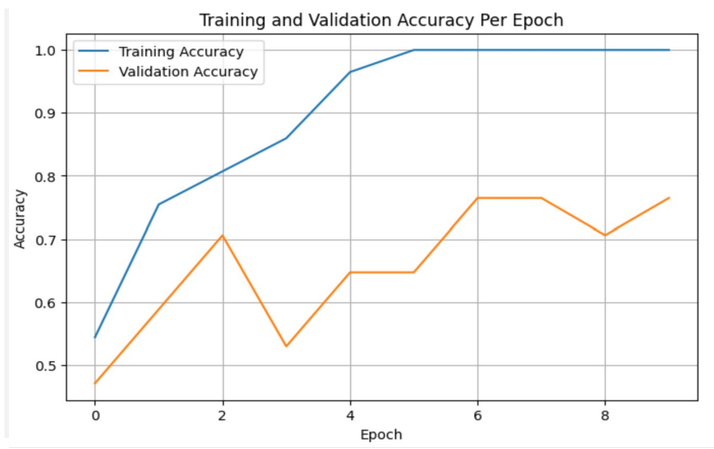
Confusion Matrix Outcomes:

- At least 50% misclassification; most confusion between 'parent' and 'other'.

Future Work:

- **Address overfitting:** Process more data and apply semi-supervised learning
- **Address overlapping data:** add 'both' label
- **Conversational Analysis:** Develop algorithm to identify sequences of parent and infant vocalizations and interactions.





REFERENCES

- Vishwkarma, D. (n.d.). CNN with Librosa implementation. Kaggle. Retrieved April 15, 2024, from https://www.kaggle.com/code/divyanshvishwkarma/cnn-with-librosa-implementatio
- Musikal Kemist. (n.d.). Music genre classification: Preparing the dataset. GitHub. Retrieved April 15, 2024, from https://github.com/musikalkemist/DeepLearningForAudioWithPython/blob/master/12-%20Music%20genre%20classification%3A%20Preparing%20the%20dataset/code/extract_data.py
- Tosi, A. [The Sound of AI]. (2020, August 3). The complete machine learning course with Python | learn machine learning [Video]. YouTube. https://www.youtube.com/watch?v=szyGiObZymo&list=PL-wATfeyAMNrtbkCNsLcpoAyBBRJZVlnf&index=14
- Jackson, C. W., & Callender, M. F. (2014). Environmental considerations: Home and school comparison of Spanish–English speakers' vocalizations. Topics in Early Childhood Special Education, 34(3), 165-174. https://doi.org/10.1177/0271121414536623
- Nittrouer, S., Caldwell-Tarr, A., Lowenstein, J. H., & Tarr, E. (2014). Enhancing the auditory and language development of children with cochlear implants. Clinical Linguistics & Phonetics, 28(1-2), 116-131. https://doi.org/10.3109/02699206.2013.809152
- Pérez-Leroux, A. T., Castilla-Earls, A., & Restrepo, M. A. (2015). Language development in bilingual children: A primer for pediatricians. Journal of Pediatric Health Care, 29(2), 147-154. https://doi.org/10.1016/j.pedhc.2014.09.003
- Spencer, E. J., Goldstein, H., Sherman, A., Nozzi, M., Schneider, N., & Hupp, S. (2018). The social interactions of children with disabilities in an inclusive pre-K classroom. Journal of Early Intervention, 40(3), 223-241. https://doi.org/10.1177/1053815118769400

STAR AUTISM INITIATIVE

UNIVERSITY of VIRGINIA | DATA SCIENCE