Stuttering Detection and Classification (SEP-28k)

Dataset

We used the SEP-28k dataset, consisting of ~28,000 podcast clips labeled with five dysfluency types: blocks, prolongations, sound repetitions, word repetitions, and interjections. Audio files were filtered, normalized, and split into training and testing sets.

Data Preprocessing

Preprocessing involved cleaning audio, segmenting longer clips, removing poor quality samples, normalizing waveforms, and stratified splitting. The resulting dataset was ready for MFCC feature extraction.

Feature Extraction

MFCCs were extracted from each audio clip using Librosa. These coefficients compactly represent spectral features crucial for classifying speech patterns.

Model Training

A neural network (TensorFlow) and traditional classifiers (scikit-learn) were trained using extracted MFCC features. Hyperparameters were optimized using validation data.

Evaluation & Results

The model achieved ~81% accuracy. A classification report showed F1-scores exceeding 0.75 for most classes, with results summarized using precision, recall, and confusion matrices.

Mobile App Deployment

Using TensorFlow Lite and Flutter, the model was deployed to an Android app for on-device inference. The app accepts user-recorded WAV files, processes them, and predicts stutter types.

Usage

1. Prepare data and extract features.
2. Train the model.
3. Evaluate with classification reports.
4. Convert model to .tflite.
5. Integrate and run using the Flutter app.

Results Summary

- Dataset: SEP-28k (28k clips)
- Features: MFCCs
- Accuracy: ~81%
- Classes: 5 dysfluency types
- Deployment: On-device Flutter app using TensorFlow Lite

Stuttering Detection and Classification (SEP-28k)

Citations

SEP-28k dataset
MFCC features with Librosa
Scikit-learn metrics
TensorFlow Lite + Flutter plugin