Project Proposal:

Title:
Audio Classification

Objective:
The goal of audio classification is to enable machines to automatically recognise numerical pronunciation and detect the numerical values between 0-9 from the audio samples.

Data:
We utilised a web interface to collect audio data from diverse subjects, employing a microphone for voice recording and organising the files into labeled folders (0-9). A total of 350 samples were gathered, approximately with 300 allocated for training purposes and 50 reserved for testing the model's performance. The recorded voices were associated with labels ranging from 0 to 9 for classification. This data acquisition process aimed to enhance the model's ability to recognise and differentiate various audio patterns.

Method:
Initially, we processed audio information in waveform format, subsequently transforming it into a Mel spectrogram. Following this, we conducted training and testing on the resulting numpy array.

We implement a convolutional neural network (CNN) using TensorFlow's Keras API with sequential layers. It consists of convolutional layers with activation functions, max-pooling layers, and dense layers. The model is configured for image data with RGB channels. Early stopping is implemented to prevent overfitting during training.
The architecture includes three sets of convolutional layers with increasing filters and max-pooling for feature extraction. Dropout is applied to reduce overfitting, followed by dense layers for classification. The final layer uses softmax activation for multi-class classification.
The model is compiled with categorical crossentropy loss, the Adam optimizer, and accuracy metrics. Training is performed using

a generator for image data, specifying the number of epochs and incorporating early stopping to monitor and halt training when performance plateaus.