Team Octonet : Identifying Communication Mode from Classroom Recordings

Mehul Varma, Rakhi Shah, Vicram Rajagopalan, Reid O'Boyle, Evelyn Crowe
Dept. of Computer Science and Engineering, Texas A&M University

Introduction

The TAMU Education Department under Dr.Irby was assigned a project where various class-recordings from an ESL classroom were supposed to be analyzed using human codings taken according to the Transitional Bilingual Observation Protocol that was developed for this project. In this project we focus on making a model that can classify these codes on one of the codings called the Communication Mode which describes the type of physical interaction between students and the instructor.

Preprocessing Implementation

Audio Extraction:

In order to remove the audio from the video, we utilized the MoviePy editor in order to make the audio file into an object Python could work with. Then we used PyDub to extract the audio from the video file. We saved these as .wav files.

Split Audio Files:

To split the audio files into the 20 second files, we used the MoviePy object we created in the audio extraction and saved all the audio files into a new directory.

Match Audio and Code:

We took the TBOP codes corresponding to each video and matched the correct time frame to the corresponding time frame in the worksheet.

Filter Codes:

We filtered the codes to only contain the classifications aural, aural-verbal, and verbal-aural. We placed these codings into a 1D Numpy array to use the values.

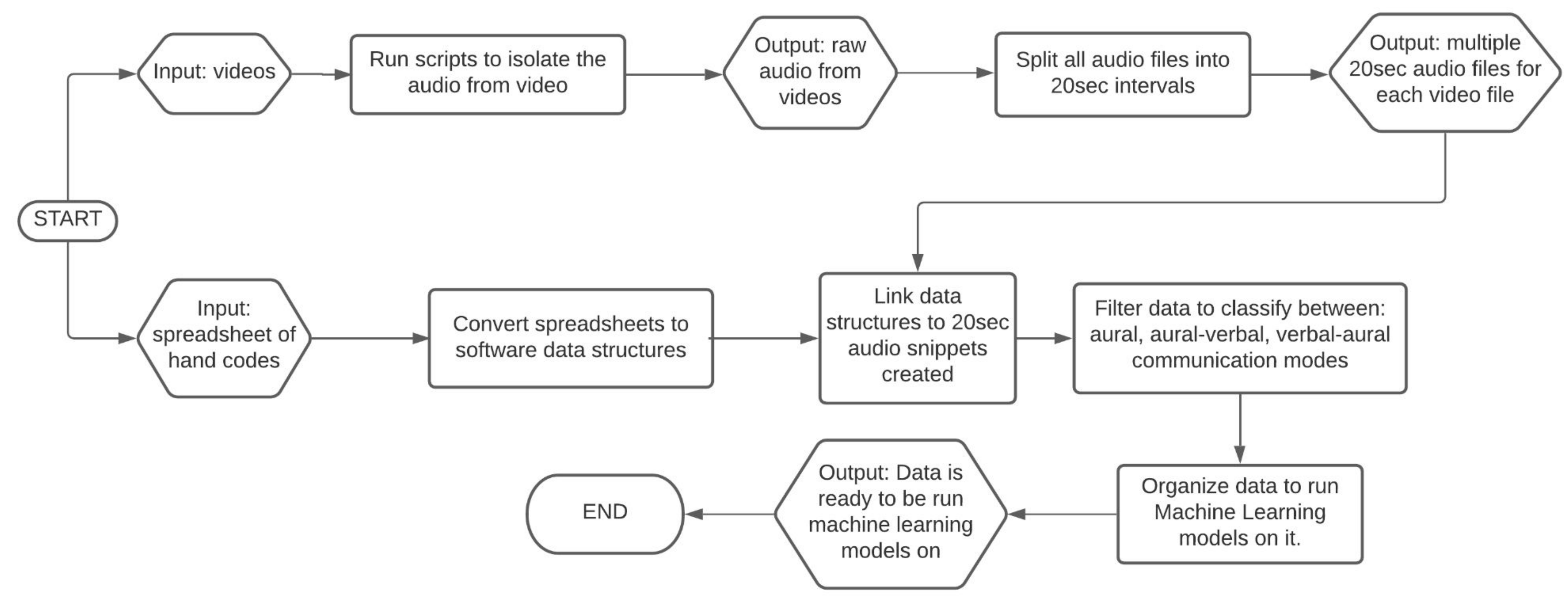
Transform Audio Files:

We used the PyDub library to convert the audio clips into 2D Numpy arrays, so we could feed these to the machine learning models.

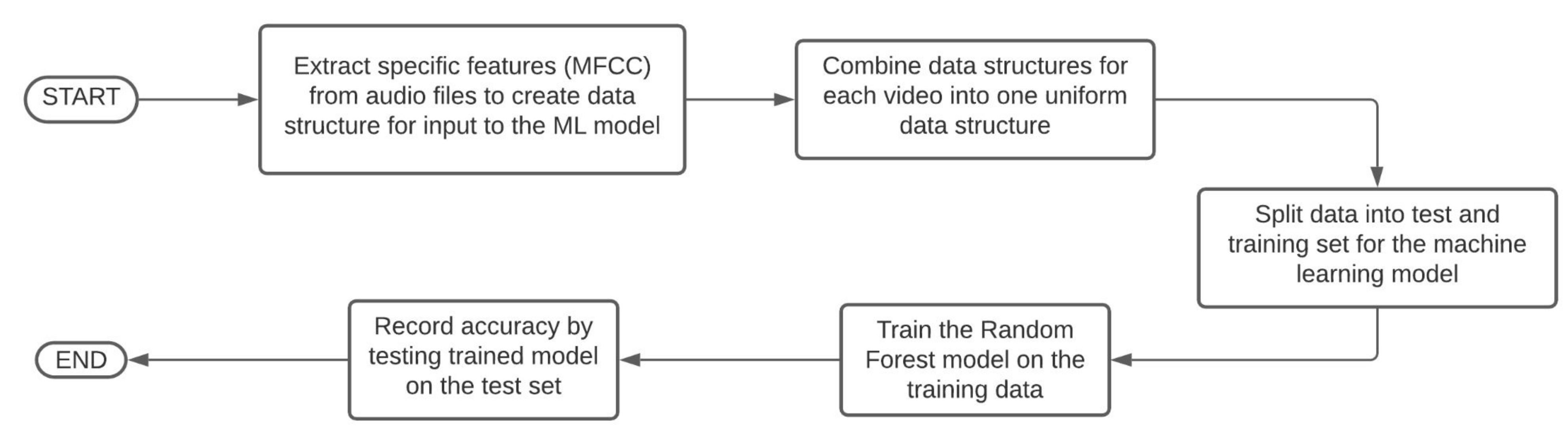
Machine Learning Implementation

Feature Extraction: We extracted Mel-Frequency Cepstral Coefficients from each audio file to use as our main feature for the model and combined the features from every file into one data structure.
Training: We split our training and testing data into 90% training data and 10% testing data. We did this 25 times randomly selecting the training and test sets and trained the model on each set.
Testing: We ran predictions of the data on our testing set and compared those results to the true data to get an accuracy and record the results.

Preprocessing Flow



Machine Learning Flow

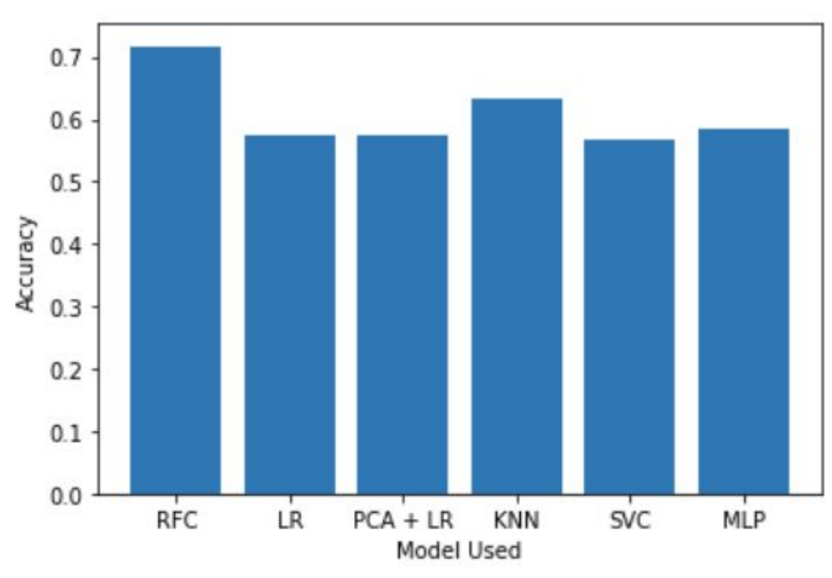


Machine Learning Models' Results

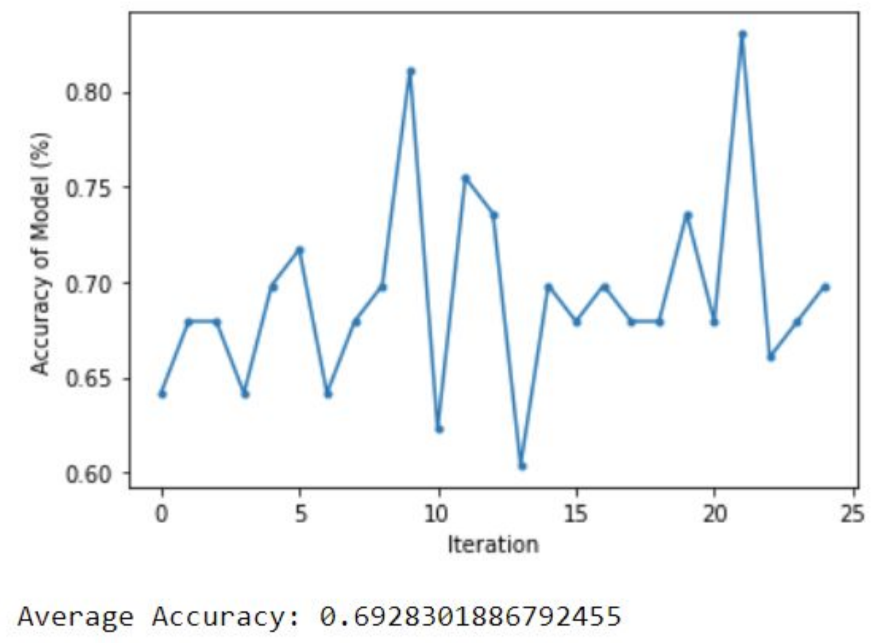
Model Name	Accuracy (%)
Random Forest Classifier	71.7
Logistic Regression	57.5
Principle Component Analysis + Logistic Regression	50
K-nearest neighbors	63.2
Support Vector Classifier	56.6
Multi-Layered Perceptron	58.5

Experimental Results

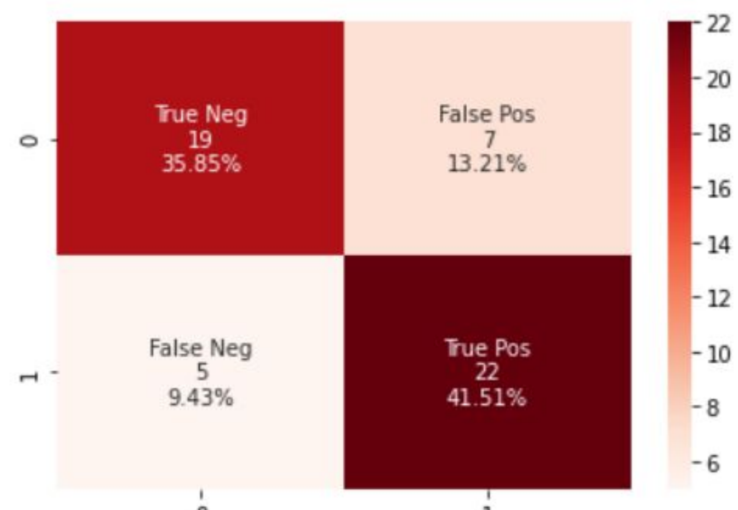
Result 1: Accuracies of ML Models tested



Result 2: Accuracy of RFC Model on 25 different randomly selected test sets



Result 3: Confusion Matrix of Results



In Result1 you can see the accuracies of each different machine learning model that we tested. You can see that a Random Forest gave us the best results over the other model we tried using. We then ran the Random Forest Classifier 25 times on randomly selected training and test sets and recorded the accuracy of each one to get a strong average accuracy. You can see the changes in accuracy from each iteration in Result 2. In Result 3 you can see the distribution of the true positive, true negative, false positive, false negative from our models predictions. True positive means if correctly predicted the Mode as 3, True Negative means it correctly predicted the Mode as 15/18, and false positive/negative meaning it incorrectly predicted the Mode.

Conclusion

From our results we can conclude that a Random Forest is the most effective for our dataset and that MFC coefficients are an effective feature for Mode prediction on the dataset. In future work, we think the accuracy can be improved by using more features as input to the model and spending more time tuning the model. In addition, exploring deep models and neural networks on this data could be future work to attempt to improve accuracy. We expect that exploring a more complex model while using other useful features such as energy, amplitude, wavelengths and other extractable features could result in better accuracy.