Coding AI To Enhance Speech Therapy

Finch, Ethan (School: Stillwater Area High School)

Natural Language Processing has led to the development of voice recognition technology, and taking speech analysis further can lead to the development of technology to aid in treatment of speech-language pathology. The purpose of this experiment is to train four different machine learning models and compare their performance in identifying correct and various incorrect /s/ phone pronunciations by using accuracy as the scoring method. A total of 1200 audio files, with 240 correct pronunciations, 240 dental pronunciations, 240 interdental pronunciations, 240 labiodental pronunciations, and 240 plosive pronunciations were created. The 1200 audio files were transformed into Mel-Frequency Cepstral Coefficients and split into 70% training, 15% validation, and 15% testing datasets. A Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) model, a Support Vector Machine (SVM), and a K-Nearest Neighbors Classifier (KNN) were created and run through hyperparameter optimization. One model of each type was trained and evaluated using the optimal hyperparameters. The hypothesis was the CNN would identify correct pronunciations and each type of mispronunciation with the best accuracy because of its ability to classify images. Each trained model was evaluated and compared based on its average accuracy in identifying correct pronunciations and different types of mispronunciations in the testing dataset. The CNN and LSTM had an average accuracy of 20% on the test data, while the SVM had an average accuracy of 95% and the KNN had an average accuracy of 92%, proving the hypothesis incorrect. The analysis and conclusion discuss considerations why the SVM performed the best.

Awards Won: