Stuttering Event Detection from Podcasts with People Who Stutter

Amartyaraj Kumar 19029

Abstract

The ability to automatically detect stuttering events in speech could help speech pathologists track an individual's fluency over time or help improve speech recognition systems for people with atypical speech patterns. Also for many individual there may be a need for a device that could filter their stutter and make their voices bright. This study here shows a comprehensive machine learning modelling trial to find the best and most accurate supervised model that can detect stutters and note the voice as a stuttering voice.

Keywords : Stutter, Stuttering, repetition, prolongation, blocks, Dysfluencies.

Introduction

Stuttering is a speech disorder which leaves fatal impact on personal and professional lives of millions of (**appx 1% of the population**)[5] people worldwide. This speech disorder is basically a disruptive flow of speech that gets interrupted by involuntary blocks, pauses and repetition of sounds. Identification of stuttering is an interesting interdisciplinary research problem involving acoustics, and signal processing,pathology, psychology, making it complicated to detect. Recent developments in the machine and deep learning have dramatically revolutionized the speech domain via ***FFT, MFCC***[1]***, etc signal processing methods***[2], however not much attention has been given to stuttering identification. In this study, a comprehensively review acoustic features, and statistical and supervised learning-based stuttering/disfluency classification methods have been explored.

[1]MFCC: Mel-frequency cepstral coefficients, `https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9`

Overview

The goal of this project is to apply signal processing techniques to convert podcast audios into several features. Here we will use the `.wav` files as the recordings. After performing FFT sampling via `librosa.feature.mfcc, librosa.feature.melspectrogram`[2] there would be multiple features available to train the model and find predict the stuttering correlations.

Research Question

- Can we find a supervised machine learning model that can predict stuttering pattern or status from any audio clip?

Related Work

A systematic review of the literature on statistical and machine learning schemes for identifying symptoms of developmental stuttering from audio recordings was reported by Barrett et al. [2]. Another study on *Machine Learning for Stuttering Identification: Review, Challenges and Future Directions* was done by Sheikh et al. [4]. Also another study has been published on *Stuttering Disfluency Detection Using Machine Learning Approaches* by Abedal-Kareem Al-Banna and Hadi [1].

Hypotheses

We can have more than 80% accurate supervised model using Random Forest classifier.

[2] `https://librosa.org/doc/`

Dataset description

- **Dataset Name :** GSep-28k: A Dataset for Stuttering Event Detection from Podcasts with People Who Stutter
- **Link :** https://github.com/apple/ml-stuttering-events-dataset/
- **Number of observation :** `32321` collected sample voices but with `4144` labelled stuttering patterns.
- **Number of variables :** `13`
- **Description of the dataset :** This database contains people's voices from more than 350 television or radio shows. Most of them are stutterers, so the data set contains mostly the voices which are by any metric stuttering.

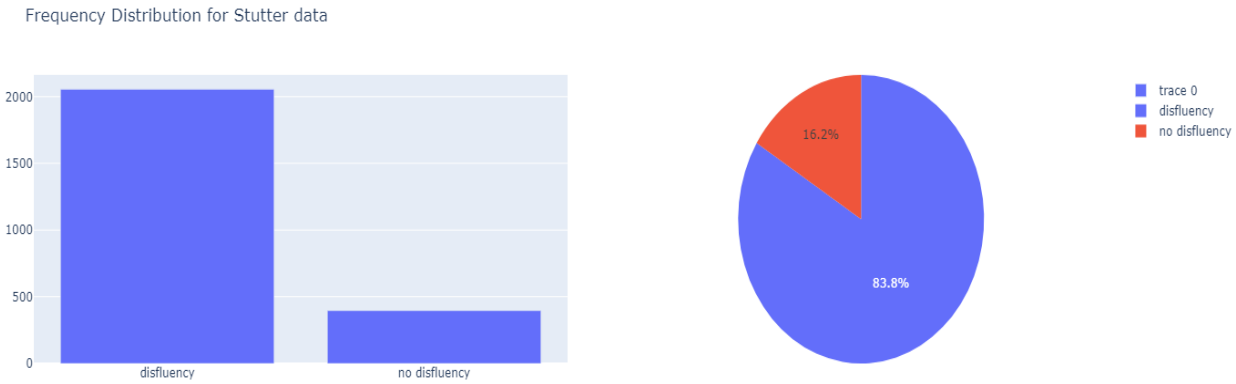

Figure 1: Distribution of voices with disfluency in the preprocessed dataset.

We have here the preprocessed data which contains two labels; **with disfluency** (encoded as '1'), **without disfluency** (encoded as '0'). The data is preprocessed through simple arithmetic operations like manual encoding and manual labelling. The preprocessed data has 3568 samples.

Methods

Here, given the dataset, mainly four supervised learning classifiers are considered for training and building the model. Those are k-NN Classification, Support Vector Machine (SVM), Logistic Regression, and Random Forest Ensemble Classifier.

At first, we extracted 25 features from the audio files by using FFT sampling of 1024 per bit. The we preprocessed our training data by adding all the patterns of stuttering as **'Stutter** column and all the extracted features to the main labelled data. The correlation heat-map of those 25 scaled features is given below [Figure 2].

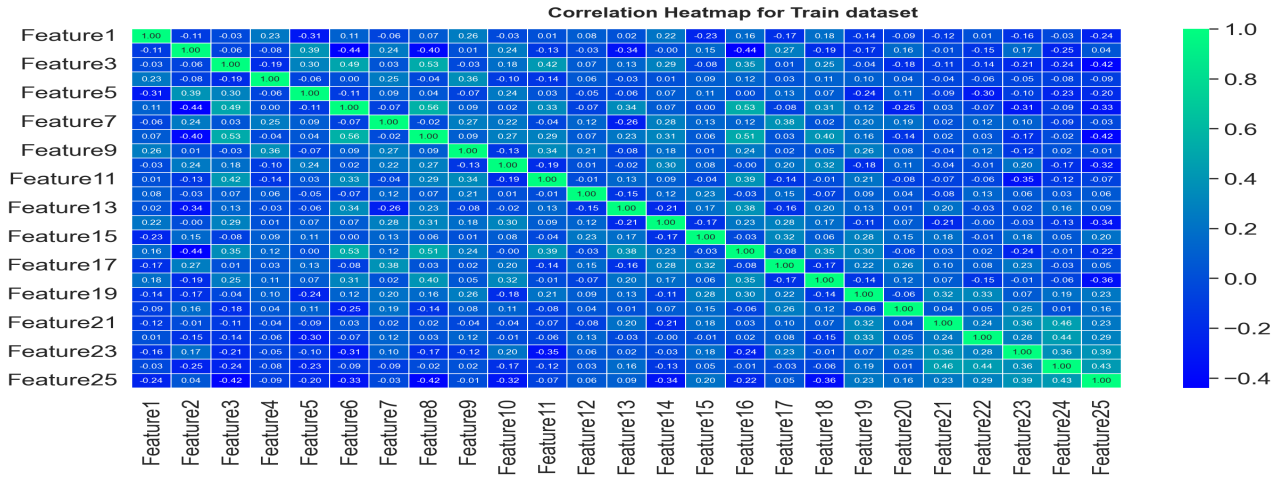

Figure 2: Correlation heat-map of those 25 scaled features.

Then we removed the labels used `StandardScaler` to scale the data, and then proceeded with `train_test_split`.We kept a training size of 80%.

KNN Classifier

We started with k-NN classifier using our preprocessed & split training data which contains all the 25 feature vectors. With the speedrun and no parameter tuning we got an accuracy of **82.28%**.

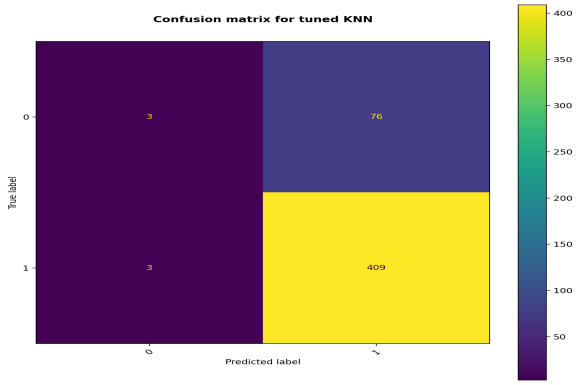

Figure 3: Confusion Matrix

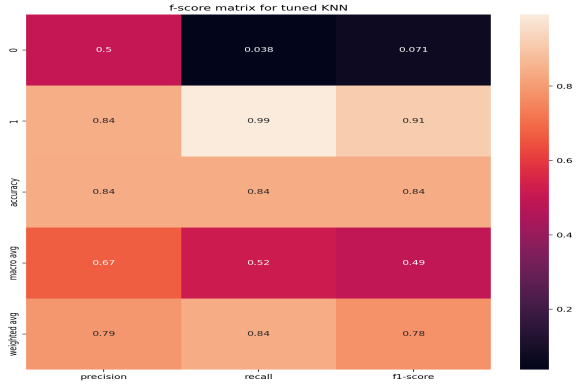

Figure 4: f-score and Accuracy

Then We opted for hyperparameter tuning with 10 cross validation of the classifier parameters and found the best k-NN fit with accuracy = **83.91%**.

SVM Classifier

With Support Vector Machine classifier, firstly we had our speedrun with all the scaled feature vectors and no parameter tuning. and as a result it produced an accuracy of **83.91%**.

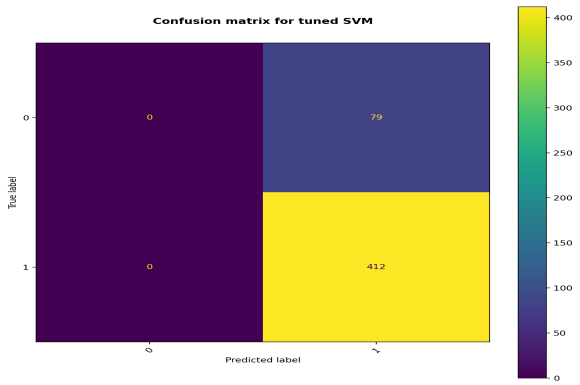

Figure 5: Confusion Matrix

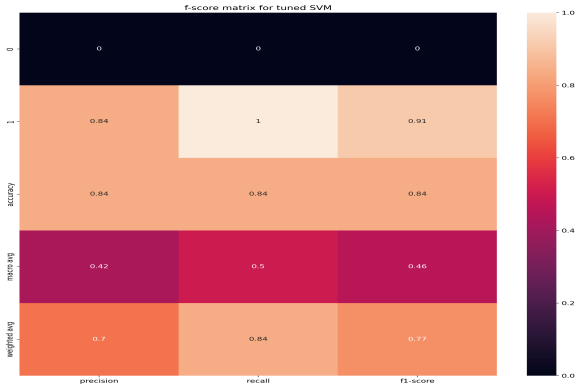

Figure 6: f-score and Accuracy

Then we opted for hyperparameter tuning with 5 cross validation of the classifier parameters including `'linear', 'rbf' & 'poly'` kernels, and found the best SVM fit with accuracy = **83.91%**.

Logistic Regression

Using our preprocessed & split training data which contains all the 25 feature vectors, we first had our no-tuning speedrun accuracy for Logistic Regression = **84.11%**.

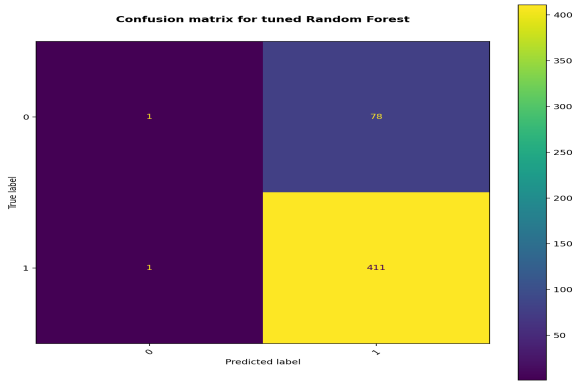

Figure 7: Confusion Matrix

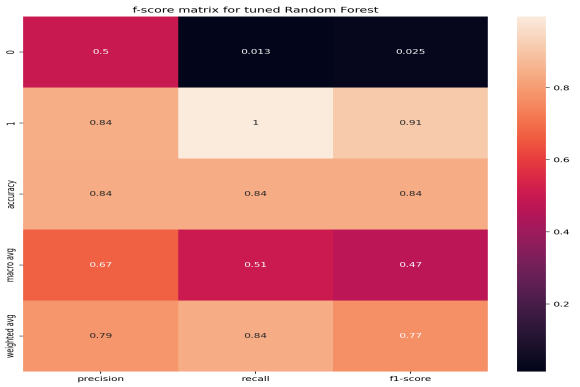

Figure 8: f-score and Accuracy

Then We opted for hyperparameter tuning of the classifier parameters with 5 cross validation and found the best Logistic Regression fit with accuracy = **84.11%**.

Random Forest Classifier

Using our preprocessed & split training data which contains all the 25 feature vectors, we first had our no-tuning speedrun accuracy for Random Forest classification = **84.11%**.

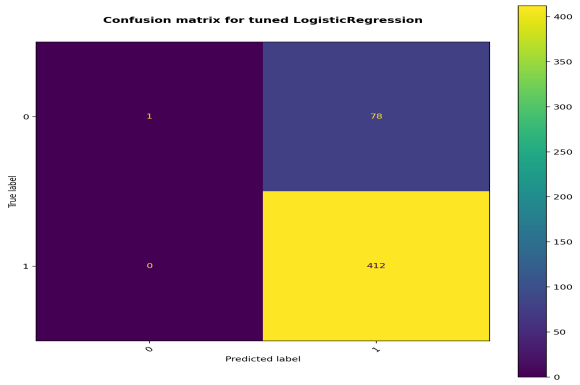

Figure 9: Confusion Matrix

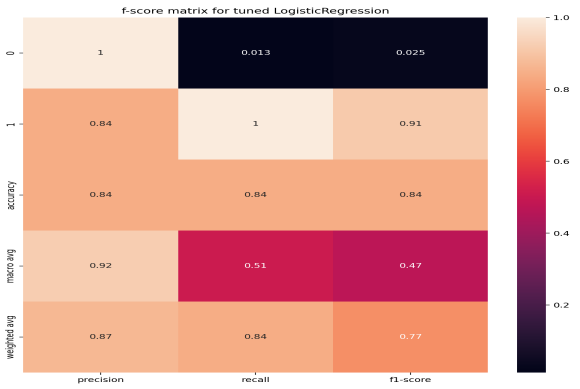

Figure 10: f-score and Accuracy

Then We opted for hyperparameter tuning of the classifier parameters with 5 cross validation and found the best Logistic Regression fit with accuracy = **84.11%**.

Accuracy Table

Accuracy Comparison		
Classifier Name	Speedrun accuracy	Accuracy after tuning
K -Nearest Neighbour	82.28%	83.91%
Support Vector Machine	83.91%	83.91%
Logistic Regression	84.11%	84.11%
Random Forest	84.11%	84.11%

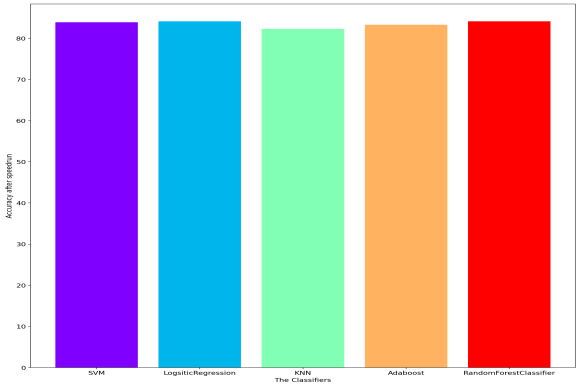

Figure 11: Initial accuracy comparison between models

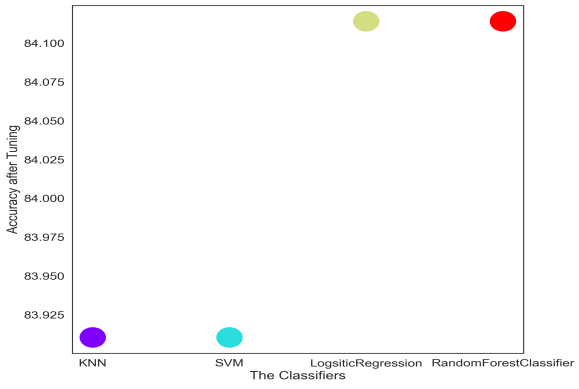

Figure 12: Final accuracy comparison between models

So, our best baseline model after hyperparameter tuning is any of **Logistic Regression or Random Forest Classifier**. So, our hypothesis here satisfies the initial assumption.

GitHub Link

Link

Discussions

We have performed some supervised learning techniques on the dataset Lea et al. [3] and got a maximum of 84.11% of accuracy in our best model. The whole dataset is noise free but the main issue is that the feature vectors are generated by FFT processing. So, having direct experimental features would definitely help more. It can be made more accurate by using neural network methods and deep learning techniques..

References

[1] H. F. Abedal-Kareem Al-Banna, Eran Edirisinghe and W. Hadi. Stuttering disfluency detection using machine learning approaches. *Journal of Information Knowledge Management*, 21(02):2250020, 2022. doi: 10.1142/S0219649222500204.

[2] L. Barrett, J. Hu, and P. Howell. Systematic review of machine learning approaches for detecting developmental stuttering. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:1160–1172, 2022. doi: 10.1109/TASLP.2022.3155295.

[3] C. Lea, V. Mitra, A. Joshi, S. Kajarekar, and J. Bigham. Sep-28k: A dataset for stuttering event detection from podcasts with people who stutter. 2021. URL `https://arxiv.org/pdf/2102.12394.pdf`.

[4] S. A. Sheikh, M. Sahidullah, F. Hirsch, and S. Ouni. Machine learning for stuttering identification: Review, challenges & future directions. *CoRR*, abs/2107.04057, 2021. URL `https://arxiv.org/abs/2107.04057`.

[5] E. Yairi and N. Ambrose. Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, 38(2):66–87, 2013. ISSN 0094-730X. doi: https://doi.org/10.1016/j.jfludis.2012.11.002. URL `https://www.sciencedirect.com/science/article/pii/S0094730X12001052`.