MEDINFO 2025 — Healthcare Smart × Medicine Deep
M.S. Househ et al. (Eds.)

doi:10.3233/SHTI250991

A Proof-of-Concept Development on Speech Analysis for Concussion Detection

Upeka De Silva[a,1], Samaneh MADANIAN [a], Ajit NARAYANAN [b], John Michael TEMPLETON [c], Christian POELLABAUER [d], Sandra L. SCHNEIDER [e] and Rahmina RUBAIAT [d]

[a] *Department of Data Science and Artificial Intelligence, AUT, New Zealand*
[b] *Department of Computer Science and Software Engineering, AUT, New Zealand*
[c] *University of South Florida, FL 33620, USA*
[d] *Florida International University, Miami, FL 33199, USA*
[e] *Saint Mary's College, Notre Dame, IN 46556, USA*
ORCiD ID: Upeka De Silva https://orcid.org/0000-0002-6913-3464

Abstract. Speech signal analysis to support objective clinical decision-making has gained immense interest, especially in neurological disorders. This research assessed the feasibility of speech analysis on the detection of concussions. Using a speech dataset from 82 concussed and 82 healthy participants, we extracted two speech feature sets focusing on Mel Frequency Cepstral Coefficients (MFCCs) to characterize speech articulation. A machine learning pipeline was developed to discriminate concussion speech from healthy speech by applying Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Tree (DT) classifiers. All three classifiers trained on the MFCC-based feature set achieved Matthew's correlation coefficient score above 0.5 on the holdout data set. DT model achieved a 78% sensitivity and 75% specificity. The findings of this research serve as proof-of-concept for speech analysis of concussion detection.

Keywords. Speech Analysis, Machine Learning, Concussion Detection

1. Introduction

Traumatic Brain Injuries (TBI) result from an external physical hit to the head, harming normal brain function. Among them, mild TBI (mTBI), known as concussions, is are frequent occurrence with potentially long-lasting health impacts. Globally, around 69 million individuals suffer a TBI each year [1] while it is estimated that almost 90% of concussions remain undetected due to difficulties in detecting brain injuries and lack of reliable, quick, and non-invasive biomarkers [2]. Concussions are predominantly diagnosed based on self-reported symptoms [3], history of the injury and previous concussions, pre-existing medical and neurological conditions. The effectiveness of neuro-cognitive and self-reported symptoms-based concussion assessments are questioned due to their potential subjectivity [2]. These clinical challenges are encouraging the investigation of novel approaches to support clinicians with their concussion diagnosis.

[1]Corresponding Author: Upeka De Silva, Auckland University of Technology, New Zealand, email: Upeka.kalumith@autuni.ac.nz

Consequently, the investigation of speech pattern changes following concussions has gained interest in the objective detection of concussions. In speech, we can examine both the 'content' (what is said) and the 'form' (how it is said). Content includes thoughts and ideas conveyed through linguistic elements like syllables and words. Form encompasses properties such as pitch and volume, which can be analyzed through acoustic properties like frequency and amplitude. Analysis of the form of speech instead of speech content has the benefit of the possibility of generalizing across languages due to similar vocal anatomy. It is especially useful for low-resource languages when natural language processing technology is unavailable. Moreover, speech-based clinical assessments can be operated non-invasively and non-intrusively, both in real-time and offline at a low cost. Further, speech assessments can be easily automated to monitor variability over time compared to one-off snapshots from other diagnostic approaches. Therefore, this research aims to explore machine learning (ML) techniques to investigate content-independent speech signal analysis to detect concussion.

2. Methods

2.1. Dataset Description

We used a speech dataset collected within the US from a group of young high school athletes, aged of 15-24, at baseline and during a sports season discussed in [2] and [4] using a microphone attached to a mobile device. For this experiment, speech recordings of reading a set of multisyllabic words (Application, Participate, Education, Difficulty, Congratulations, Possibility, Mathematical, Opportunity) were obtained. This semi-structured speech task demands complex motor patterns that go from the front to the back of the mouth and, therefore, focus on fluency/prosody and articulation in speech production [2]. Concussion status had been clinically verified by either an athlete trainer or a physician and concussion speech had been recorded during the acute period (within 48 hours of head impact). Table 1 summarizes the details of the study sample. A single recording file from each participant was obtained to maintain speaker independence during training and testing.

Table 1. Summary of participants in the study sample

Column1	Concussed participants	Healthy participants
Number of participants	Male:70, Female:12	Male:407, Female:66
Speech test file duration (s)	Male:13.03 +/- 0.08	Male: 13 +/-0.05
	Female:13.00 +/-0.06	Female: 13+/0.06

2.2. Speech Feature Extraction

Speech features were extracted using the Pyaudioanalysis library [5] taking middle 10-second of each utterance. Pyaudioanalysis supports the extraction of 34 speech features (see Table 2) and their first-order derivative per frame, further summarizing them through mean and standard deviation to get 136 speech features per speech utterance.

Table 2. Summary of speech features

Speech feature domain	Supported speech features	Feature count
Time domain	Zero Crossing Rate, Energy, Entropy of Energy	3
Frequency domain	Spectral Centroid, Spectral Spread, Spectral Entropy, Spectral Flux, Spectral Rolloff, 12 Chroma Vectors, Chroma Deviation	18
Cepstral domain	13 MFCCs	13

Experiments were carried out on two main speech feature sets (presented in Table 3) including the entire speech feature set and a set of Mel Frequency Cepstral Coefficients (MFCCs) based features. MFCCs compute the energy variations between frequency bands of a speech signal and can characterize the spectral envelope, especially related to articulation [6].

Table 3. Speech feature sets used in the experiments

Speech feature set	Speech features	Summary statistics	Feature count
Feature set 1 (FS1)	34 speech features and first-order derivatives	Mean, std	136
Feature set 2 (FS2)	13 MFCCs and first-order derivatives	Mean, std	52

2.3. ML Algorithms and Implementation

We developed an ML pipeline (Figure 1) with three popular classification models: Support Vector Machines (SVM), Decision Tree (DT), and k-Nearest neighbors (KNN).

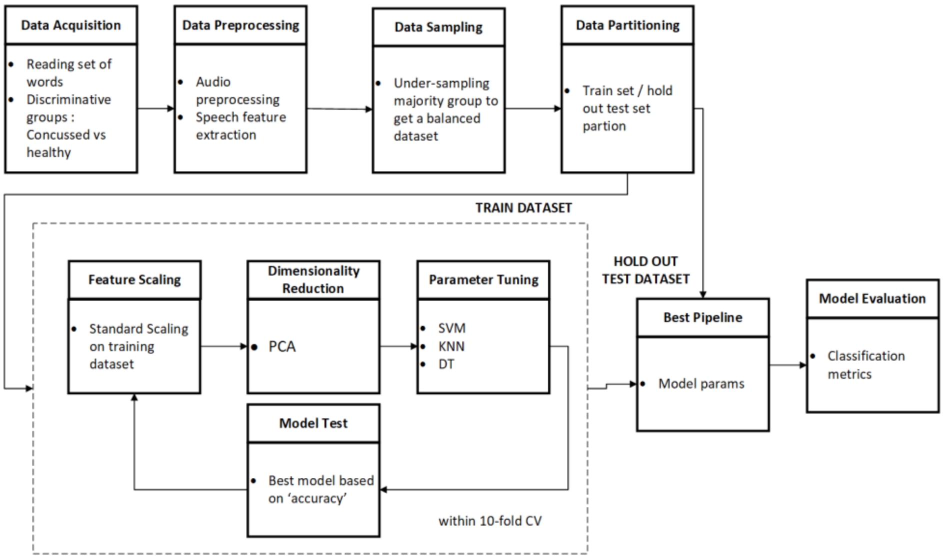

Figure 1. Architecture of ML pipeline.

From a dataset of 473 healthy and 82 concussed participants, a balanced dataset (82 participants per group) was obtained through under sampling, and was partitioned into a training and a hold-out test set, keeping 90% and 10% of the data respectively. Then, the experiments were conducted with and without applying dimensionality reduction using principal component analysis (PCA) on the standardized speech features. The number of principal components to consider by the ML pipeline was empirically tuned. Hyper-parameters were tuned using grid search with 10-fold cross-validation on the training dataset. The performance of the best model was evaluated on the hold-out test set on their accuracy, sensitivity, specificity, Matthew's correlation coefficient (MCC), and area under the receiver operating characteristic curve (ROC AUC).

3. Results

In total, 12 classification models were trained (3 classifiers * 2 feature sets * with/without PCA). Table 4 summarizes model evaluation results on the hold-out test set.

Table 4. Model evaluation results on the hold-out test set

Feature set and feature set reduction	ML model	Accuracy	Specificit y	Sensitivity	ROC-AUC	MCC
FS1-No PCA	DT	0.41	0.50	0.33	0.38	-0.17
	KNN	0.59	1.00	0.22	0.43	0.34
	SVM	0.59	0.50	0.67	0.75	0.17
FS1-With PCA	DT	0.41	0.25	0.56	0.42	-0.20
	KNN	0.59	1.00	0.22	0.53	0.34
	SVM	0.59	0.63	0.56	0.58	0.18
FS2-Without PCA	DT	0.76	0.75	0.78	0.82	0.53
	KNN	0.71	1.00	0.44	0.82	0.52
	SVM	0.76	1.00	0.56	0.92	0.61
FS2-With PCA	DT	0.65	0.88	0.44	0.55	0.35
	KNN	0.76	1.00	0.56	0.78	0.61
	SVM	0.65	0.75	0.56	0.75	0.31

For FS1 (entire speech feature set), none of the models could achieve an MCC score above 0.5. For FS2, four models (grey background in Table 4) have achieved an MCC score above 0.5. All 3 classifiers achieved an MCC score above 0.5 without applying PCA for feature reduction, while a KNN model could achieve an MCC score above 0.5 when PCA is applied. Out of them, the DT model gave the best performance with 0.78% sensitivity, 75% specificity, 0.82 AUC ROC, and 0.53 MCC score.

4. Discussions

In this study, we developed a proof-of-concept to show the feasibility of speech analysis for concussion detection. From a brief speech task focusing on speech articulation, we studied MFCC-based speech features' contribution to concussion discrimination. Our results align with the findings from the literature on the potential role of MFCC-based speech features in concussion assessments. MFCC features were explored from the speech of rugby players (7 mTBI, 39 HC) using bidirectional long short-term memory attention (Bi-LSTM-A) based deep learning network leading to an overall classification accuracy of 89.5% [7]. Subclinical mTBI detection was also studied with MFCC-based features [8].

Several studies have focused on potential speech-motor control deficits following mTBI for disease prediction [2; 4; 9]. From a dataset of 580 youth athletes (95 concussed players), the statistical significance of a set of speech features in concussion prediction was explored [2] and a mTBI classification model was built with an AUC of 0.86 [4]. Another study (105 players including 7 concussed) had classified mTBI with an accuracy of 98.2% [9]. Despite limited performance of this study at 78% sensitivity and 0.82 AUC, the findings highlight the role of MFCC-based features in concussion discrimination.

Speech-based concussion detection and identification of specific speech changes bring cost-effective clinical decision support in concussion management. Importantly, it can empower digital health with telehealth management for low-resource environments where evidence of mTBI risks is high. In addition, analysis of speech pattern changes over time can enable low-cost monitoring and follow-up of this time-evolving pathology with a low burden to patients and healthcare systems during rehabilitation.

5. Conclusions

This study developed an ML pipeline to predict concussion from speech features derived from a short speech task. The results show the feasibility of concussion detection through speech signal analysis and the importance of MFCC features in concussion discrimination. We plan to expand our speech feature set and carry out feature importance analysis to improve model performance and characterize speech feature changes associated with concussion with respect to gender.

References

[1]. [1] M.C. Dewan, A. Rattani, S. Gupta, R.E. Baticulon, Y.-C. Hung, M. Punchak, A. Agrawal, A.O. Adeleye, M.G. Shrime, A.M. Rubiano, J.V. Rosenfeld, and K.B. Park, Estimating the global incidence of traumatic brain injury, *J Neurosurg* (2018), 1080-1097.

[2]. [2] C. Poellabauer, N. Yadav, D. Louis, S.L. Schneider, C. Busso, and P.J. Flynn, Challenges in Concussion Detection Using Vocal Acoustic Biomarkers, *IEEE Access* **3** (2015), 1143-1160.

[3]. [3] A.R. Mayer, D.K. Quinn, and C.L. Master, The spectrum of mild traumatic brain injury, *Neurology* **89** (2017), 623-632-632.

[4]. [4] L. Daudet, N. Yadav, M. Perez, C. Poellabauer, S. Schneider, and A. Huebner, Portable mTBI Assessment Using Temporal and Frequency Analysis of Speech, *IEEE Journal of Biomedical and Health Informatics* **21** (2017), 496-506.

[5]. [5] T. Giannakopoulos, PyAudioAnalysis: An open-source python library for audio signal analysis, *PLoS ONE* **10** (2015).

[6]. [6] R. Khaskhoussy and Y.B. Ayed, Speech processing for early Parkinson's disease diagnosis: machine learning and deep learning-based approach, *Social Network Analysis and Mining* **12** (2022).

[7]. [7] C. Wall, D. Powell, F. Young, A.J. Zynda, S. Stuart, T. Covassin, and A. Godfrey, A deep learning-based approach to diagnose mild traumatic brain injury using audio classification, *PLoS ONE* **17** (2022).

[8]. [8] T. Talkar, S. Yuditskaya, J.R. Williamson, A. Lammert, H. Rao, D. Hannon, A. O'Brien, G. Vergara-Diaz, R. DeLaura, D. Sturim, G. Ciccarelli, R. Zafonte, J. Palmer, P. Bonato, and T.F. Quatieri, Detection of subclinical mild Traumatic Brain Injury (mTBI) through speech and gait, in, 2020, pp. 135-139.

[9]. [9] M. Falcone, N. Yadav, C. Poellabauer, and P. Flynn, Using isolated vowel sounds for classification of Mild Traumatic Brain Injury, in: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 7577-7581.