Identification of Alzheimer's Disease using Non-linguistic Audio Descriptors

Chitralekha Bhat
TCS Reseach and Innovation
Mumbai, India
bhat.chitralekha@tcs.com

Sunil Kumar Kopparapu
TCS Reseach and Innovation
Mumbai, India
sunilkumar.kopparapu@tcs.com

***Abstract*—Dementia is an overall term used to describe the reduced cognitive functioning in human beings, that is severe enough to impact their daily activities. Early diagnosis of dementia is imperative to provide timely treatment, either medication or therapy to alleviate the effects and sometimes slow the progression of dementia. In this work, we use speech processing and machine learning techniques to automatically classify speech into (a) healthy (HC) (b) with mild cognitive impairment (MCI) or (c) with Alzheimer's disease (AD). Only acoustic non-linguistic parameters are used for this purpose, making this a language independent approach. We evaluate our work using dementia and healthy speech from Pitt corpus of DementiaBank database. The performance of a three class Random Forest classifier is compared with our system comprising multiple two-class Random Forest classifiers cascaded to form a three class classifier, wherein a combination of approximate posterior probabilities is used to obtain a final class probability estimate. additional, patient speech is classified at segment level as well as at overall conversation level. Post processing on the patient speech classification at segment level provides a classification accuracy of 82% which is a significant absolute improvement of 8% over a simple three-class classifier performance.**

***Index Terms*—Alzheimer's disease, Dementia, classification, feature selection**

I. Introduction

Alzheimer's disease (AD) is a progressive brain disease that begins well before clinical symptoms emerge. AD spans a continuum including those with dementia due to AD, with mild cognitive impairment (MCI) due to AD and asymptomatic individuals who have verified biomarkers of AD [1]. Advanced dementia renders a person incapable of performing everyday activities since it is mainly characterized by difficulties with memory, language, problem-solving and other cognitive skills. Timely diagnosis of AD is imperative to provide in-time treatment. Manual diagnosis of AD requires specialized skills of neurologists and geriatricians through a series of cognitive tests such as the mini mental state examination (MMSE) [2]. Other means of diagnosis involve collection/examination of cerebrospinal fluid from the brain and a magnetic resonance brain imaging (MRI), that can be invasive and painful as well as expensive and tedious. Hence a simple and non invasive approach is preferable. Speech is a good indicator of the cognitive state of a person [3] and can be acquired non invasively in the form of a natural audio conversation with no additional stress on the person.

Both linguistic (lexicon syntactic and semantic) and paralinguistic (acoustic) speech parameters have been harnessed to estimate AD stage. Earlier works used manual transcriptions to obtain linguistic features from dementia speech to classify into dementia stages [4], [5], recent works use automatic speech recognition to obtain the lexical and semantic features for this purpose [6], [7]. N-gram based approaches have been used for automatic detection of AD from speech [8], [9]. Working with acoustic features alone provides a language independent framework for dementia classification [10]–[12]. High accuracies have been reported for two-class classification, especially for the healthy control (HC) and the AD speech [12]. However, classification accuracy reduces when more than one stage of dementia is considered [8]. Moreover, classification of control `vs.` MCI and MCI `vs.` AD is non trivial wherein different sets of features provide better separation between the class pairs [11].

In this paper, we present a technique to classify a given utterance into one of the three classes, namely, healthy (HC), with MCI or with AD. We extend multiple binary classifiers (Random Forest) to form a three class classifier, using posterior probability estimates from the three two-class classifiers to make a decision. Classification is carried out on two sets of data (a) Speech turns belonging to the patient are concatenated and classified as a whole. (b) Each patient turn within a clinician-patient conversation is treated as a single utterance and classified separately, the complete utterance is then classified into the class to which maximum number of turns belong. The peformance of a conventional three-class classifier is compared with our proposed technique, where 3 binary cassifiers are fused by combining posterior probabilities. Results show that the proposed system along with post processing provides the best classification performance. Since classification is done using only acoustic descriptors computed from speech, this can be considered as a language independent approach. To the best of our knowledge this is the first attepmt at three-classification of AD, MCI and HC utterances using only acoustic parameters.

The rest of the paper is organized as follows. Section II describes the methodology and the proposed system to extend two class classifiers into multi-class classifiers. Section III describes the experimental setup, section IV gives the results and analysis of results. We conclude in section V.

978-9-0827-9703-9/19/$31.00

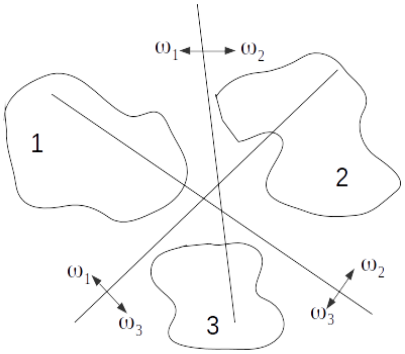

Fig. 1: Approach to convert multiple-class database into multiple two-class discriminant boundaries

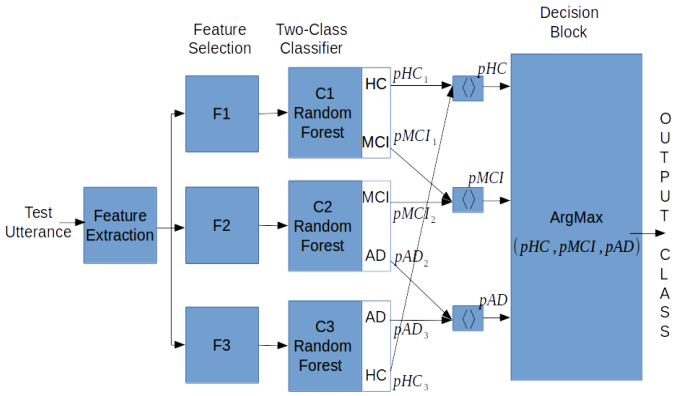

Fig. 2: Proposed system to extend 3 two-class classifiers into three-class classifier

II. Methodology

In this paper we consider classification of an utterance into one of the three classes, namely HC, MCI and AD. A straight forward technique would be to use a three class classifier for the task. However, it can be seen from literature that two-class classfication into HC and MCI or classification into MCI and AD is not a trivial task, though classification of HC and AD have shown high accuracies [10]. However, when three class classification is used, the accuracy drops further. We hypothesize that this can be attributed to the fact that speech features that represent a clear separation between the three class pairs (HC-MCI; MCI-AD; AD-HC) are different. In order to tackle this problem, we propose a system that uses multiple two class classifiers to form a three class classifier as described in [13].

Let us consider a database with N classes represented by class label ω_n where $n = 1, 2, 3 \cdots N$. The goal is to classify, a new object from the same distribution into one of the N classes.

Further, we train a discriminant function $f_{i,j}(x;\omega)$ where x is the object from the database, on a two-class classification problem involving classes ω_i and ω_j. The discriminant is optimized such that:

$$f_{i,j}(x;\omega) = \begin{cases} \geq 0 & \text{if } x \in \omega_i \\ < 0 & \text{if } x \in \omega_j \end{cases} \quad (1)$$

Figure 1 depicts the approach wherein the database of $N = 3$ classes is classified into 3 two-class discriminant boundaries. The posterior probabilities for a particular object x belonging to a class k is estimated as $p(x_k)$, for each of the two-class classifiers. The final class of an object x is estimated based on the averages of the posterior probabilities of multiple two-class estimates and the object is classified into the class with maximum average posterior probability.

Figure 2 shows the system design to extend three two-class Random Forest classifiers, trained for two-class pairs (1) HC-MCI, (2) AD-MCI and (3) AD-HC, to form a three-class classifier, which provides an estimate for a particular test utterance as belonging one of the three classes HC, MCI or AD.

Consider a patient utterance s with N segments such that $s = cocatenation(s_1, s_2 \ldots s_N)$. Let the classifiers HC-MCI, AD-MCI and AD-HC be denoted as 1, 2 and 3 respectively and the posterior probabilities be denoted as p<$class$>$_{<classifierID>}$ where class would be HC, MCI or AD and classifier ID would 1, 2 or 3. The posterior probabilities for each utterance s as well as the individual segments s_n are obtained from the two-class classifiers 1, 2 and 3 and are averaged as shown below.

1) $pHC = \langle pHC_1, pHC_3 \rangle$
2) $pMCI = \langle pMCI_1, pMCI_2 \rangle$
3) $pAD = \langle pAD_2, pAD_3 \rangle$

Where the operation $\langle x, y \rangle$ computes the average of x and y. The class decision for s is made on the basis of $maximum(pHC, pMCI, pAD)$. A similar process is carried out for each of the segments s_n and a class decision is made for each segment as $s_n \in \{HC, MCI, AD\}$. Overall class descision for an utterance s is made using the individual segment class decision as shown below:

$$s \quad \in \quad argmax(n_{HC}, n_{MCI}, n_{AD}) \quad (2)$$

where $(n_{HC} + n_{MCI} + n_{AD} = N)$

III. Experimental Setup

A. Data

Pitt corpus [14] from the DementiaBank data set, collected at University of Pittsburgh School of Medicine was used. It comprises clinician-patient interviews in the form of audio, manual transcripts and subjective assessment of the patient's cognitive state as a longitudinal study over the span of four years. Data corresponds to four different linguistic-cognition tasks namely, picture description, fluency, recall and sentence construction. In this work we use the audio, transcription and subjective assessment from the picture description task, which is a verbal description of the *Boston Cookie Theft* picture. It was recorded from people with different types of dementia with an age range from 49 to 90 years as well as from healthy (HC) subjects with an age range from 46 to 81 years. During the interviews, patients were asked to discuss everything they

TABLE I: Top 15 Features selected - Concatenated utterances

AD-HC	AD-MCI	HC-MCI
F0env_sma_nnz	mfcc_sma_de_de[7]_qregerrA	pcm_fftMag_spectralRollOff90.0_sma_percentile98.0
mfcc_sma_de_de[7]_qregerrA	mfcc_sma_de_de[7]_qregerrQ	pcm_fftMag_spectralRollOff90.0_sma_percentile95.0
mfcc_sma_de_de[10]_numPeaks	mfcc_sma_de_de[11]_qregerrQ	pcm_fftMag_spectralMaxPos_sma_range
pcm_fftMag_spectralRollOff90.0_sma_percentile95.0	mfcc_sma_de_de[8]_qregerrQ	pcm_fftMag_spectralRollOff90.0_sma_quartile3
mfcc_sma[0]_meanPeakDist	mfcc_sma_de[11]_qregerrA	pcm_fftMag_spectralMaxPos_sma_de_range
mfcc_sma_de[0]_kurtosis	pcm_LOGenergy_sma_nnz	pcm_fftMag_spectralMinPos_sma_percentile98.0
mfcc_sma_de_de[0]_zcr	mfcc_sma_de_de[6]_numPeaks	pcm_fftMag_spectralMaxPos_sma_de_minameandist
mfcc_sma_de[0]_meanPeakDist	pcm_fftMag_spectralRollOff90.0_sma_range	pcm_fftMag_spectralMaxPos_sma_percentile98.0
pcm_fftMag_spectralRollOff90.0_sma_percentile98.0	pcm_fftMag_spectralRollOff90.0_sma_percentile98.0	pcm_fftMag_spectralMaxPos_sma_de_maxameandist
pcm_fftMag_melspec_sma[4]_linregc1	pcm_fftMag_spectralMaxPos_sma_de_range	pcm_fftMag_spectralMinPos_sma_percentile95.0
F0_sma_linregc1	pcm_fftMag_melspec_sma[22]_linregerrQ	mfcc_sma[12]_minameandist
F0env_sma_linregc2	pcm_fftMag_melspec_sma[22]_variance	pcm_LOGenergy_sma_minameandist
F0_sma_de_qregc2	pcm_fftMag_spectralMaxPos_sma_range	pcm_fftMag_spectralRollOff90.0_sma_de_minameandist
voiceProb_sma_de_de_stddev	pcm_fftMag_spectralRollOff75.0_sma_de_range	pcm_fftMag_spectralRollOff90.0_sma_range

TABLE II: Top 15 Features selected - Segmental utterances

AD-HC	AD-MCI	HC-MCI
pcm_LOGenergy_sma_numPeaks	pcm_fftMag_melspec_sma[25]_quartile1	pcm_fftMag_spectralRollOff90.0_sma_percentile98.0
mfcc_sma[12]_numPeaks	pcm_fftMag_spectralRollOff90.0_sma_percentile98.0	pcm_fftMag_spectralRollOff90.0_sma_percentile95.0
pcm_fftMag_melspec_sma[5]_numPeaks	mfcc_sma[12]_numPeaks	pcm_fftMag_spectralRollOff90.0_sma_quartile3
pcm_fftMag_melspec_sma[4]_numPeaks	pcm_LOGenergy_sma_numPeaks	pcm_fftMag_spectralMinPos_sma_percentile98.0
pcm_fftMag_melspec_sma[3]_numPeaks	pcm_fftMag_melspec_sma[5]_numPeaks	pcm_fftMag_spectralMaxPos_sma_range
pcm_fftMag_fband250-650_sma_numPeaks	pcm_fftMag_melspec_sma[25]_quartile2	pcm_fftMag_spectralRollOff90.0_sma_quartile2
pcm_fftMag_fband0-650_sma_numPeaks	pcm_fftMag_spectralRollOff90.0_sma_percentile95.0	pcm_fftMag_spectralRollOff90.0_sma_range
pcm_fftMag_melspec_sma[6]_numPeaks	pcm_fftMag_melspec_sma[6]_numPeaks	pcm_fftMag_melspec_sma[25]_quartile1
pcm_fftMag_melspec_sma[2]_numPeaks	pcm_fftMag_fband0-650_sma_numPeaks	mfcc_sma[10]_quartile2
pcm_fftMag_melspec_sma[22]_quartile1	pcm_fftMag_melspec_sma[4]_numPeaks	mfcc_sma[12]_minameandist
pcm_fftMag_melspec_sma[24]_quartile1	pcm_LOGenergy_sma_de_de_qregerrQ	pcm_fftMag_melspec_sma_de_de[1]_skewness
pcm_fftMag_melspec_sma[7]_numPeaks	pcm_fftMag_fband250-650_sma_numPeaks	pcm_fftMag_fband0-650_sma_numPeaks
pcm_fftMag_melspec_sma[1]_numPeaks	pcm_fftMag_melspec_sma[3]_numPeaks	pcm_fftMag_melspec_sma[9]_centroid
pcm_fftMag_melspec_sma[20]_quartile1	pcm_fftMag_melspec_sma[7]_numPeaks	pcm_fftMag_melspec_sma_de_de[0]_zcr

could see happening in the cookie theft picture. We consider a sample data with a total of 597 recordings from 97 HC participants, 168 AD patients and 19 patients diagnosed with MCI. For HC and AD participants, recordings from only *cookie theft* task was considered, whereas for MCI, recordings from all except *sentence* task was considered so as to make up for the data insufficiency. We use the speaker timing information provided in the transcripts, to remove the clinician turns from the recordings, retaining only the participant speech.

We work with two sets of data, namely:

- All the patient turns within a clinician-patient conversation are concatenated to form one single utterance.
- Each patient turn within a clinician-patient conversation is segmented and classified as a separate utterance.

B. Feature extraction

The openSMILE toolkit [15] was configured to use the large openSMILE emotion feature set (emolarge), which gives 6552 acoustic features. Time domain and frequency domain acoustic descriptors such as signal energy, loudness, MFCC, pitch, voice quality (jitter, shimmer), spectral shape descriptors and their statistical functionals such as means, extremes, moments, segments, percentiles etc. are some of the feature used. This set of speech features was used as the super set for all 3 two-class classifiers, namely HC-MCI, AD-MCI and AD-HC as well as the three-class classifier for HC-MCI-AD classification.

C. Feature selection and Classification

We use the attribute evaluator *CfsSubsetEval* with the search method *BestFirst* specified in the Weka toolkit [16] for feature selection and the Random Forest classifier with 500 trees was used along with 10-fold cross validation for classification. Feature selection process was carried for each of the two-class classfiers; namely HC-MCI, AD-MCI and AD-HC. Features that provide the best discrimination between the classes were selected for each class pair. Similar exercise was conducted for the three-class classifier as well. The two-class classifiers were then trained using their respective selected features. The proposed system uses the posterior probability estimates provided by each of the two-class classifier for individual utterances to enable the final class decision as shown in Figure 2. The two-class, three-class and proposed system classification results are discussed in the Section IV.

The feature extraction, selection and classfication processes described above were carried out for both the datasets mentioned in Section III-A The top 15 ranked features selected for each of the two-class classifiers are shown in Tables I and II. We have retained the openSMILE names of the features for the sake of reproducibility. For the concatenated utterances, the AD-HC classifier features comprised MFCC, pitch, spectral and voice quality-based features while the classifiers AD-MCI and HC-MCI comprised majorly of spectral features. However, for segmental utterances spectral features were selected for all the three classifiers. It can be observed that there are only a few common features between the two-class classifiers for both concatenated as well as segmental data sets.

IV. Results and Analysis

Firstly, we analyze the performance of our system using the concatenated patient utterances. Next, we use the predicted

classes of the individual segments within a clinician-patient conversation in such a way that we classify the entire conversation into the class to which maximum number of segments have been classified as described in Section II.

We present a comparison between the classification accuracies of a three-class Random Forest (RF) classifier and the proposed system which arrives at a class decision using posterior probability estimates from three two-class classifiers. Post processing on segment-wise classification showed significant improvements over concatenated utterance-based classification at the three-class, individual two-class as well as the final proposed system with cascaded two-class classifiers.

Table III shows the precision (Pr), recall (Re) and F-score for a three-class RF classifier designed to classify for HC, MCI and AD classes using the selected features from the emolarge superset.

TABLE III: Three-class classification performance

Class	Concatenated			Segmental		
	Pr	Re	F-Score	Pr	Re	F-Score
HC	76.8	91.7	83.6	76.8	95.5	85.1
MCI	80.6	43.1	56.2	97.8	38.8	55.6
AD	80.9	83.1	82.0	83.5	87.1	85.2
Overall	79.4	72.6	73.9	86.0	73.8	75.3

It can be seen from the Table III that although the selected features provide good discrimination between AD and HC, the same features do not fare well for MCI. However, the classification of segmented patient turns followed by post processing performs better than the concatenated patient turns.

Table IV shows the precision, recall and F-score for the three individual two-class RF classifiers shown in Figure 2 using the selected features from the emolarge superset. In

TABLE IV: Two-class classification performance

Class	Concatenated			Segmental		
	Pr	Re	F-Score	Pr	Re	F-Score
HC	85.3	96.3	90.5	83.2	100.0	90.8
MCI	89.4	65.5	75.6	100.0	57.8	73.2
Overall	87.4	80.9	83.1	91.6	79.0	82.0
AD	82.5	94.1	87.9	85.3	100.0	92.1
MCI	81.3	56.0	66.3	100.0	62.1	76.6
Overall	81.9	75.1	77.1	92.6	81.0	84.3
AD	93.2	85.9	89.4	95.8	90.1	92.8
HC	86.3	93.4	89.7	89.9	95.7	92.7
Overall	89.8	89.7	89.6	92.9	92.9	92.8

each of the individual classifiers, the F-score for the class MCI is low as compared to AD and HC, however, it shows significant improvements of the order of 20% over the three class classifier. This indicates that the feature selection process for a three-class classifier to distinguish between the three classes HC, MCI and AD is non-trivial. In order to overcome this, we propose to use the posterior probabilities of the two-class classifiers to arrive at a class decision for an utterance.

Additionally, we apply the *cascaded two-class to three-class* to both concatenated and segmental utterances as described in Section II. Final decision regarding the class of an utterance is computed as shown in Figure 2 and is shown in Table V.

TABLE V: Proposed method three-class classification performance 1

Class	Concatenated			Segmental		
	Pr	Re	F-Score	Pr	Re	F-Score
HC	81.9	95.5	88.2	82.6	98.3	89.8
MCI	79.5	53.4	63.9	100.0	48.3	65.1
AD	87.7	87.1	87.4	88.5	93.3	90.8
Overall	83.1	78.7	79.8	90.4	80.0	81.9

We observe that the proposed system performs with a 9% improvement on both the concatenated as well as segmental utterances as compared to the three-class classifiers for MCI utterances, 6% and 9% improvement for AD and HC classes respectively and with an overall improvement of 8%.

V. Conclusion and Future work

Alzheimer's disease (AD) is a slow progressive brain disease that begins well before clinical symptoms emerge. Timely diagnosis of AD is imperative to provide in-time treatment be it medication or therapy. Also, in the initial stages the progression of the disease can be prolonged in some cases. In this work, we use speech processing and machine learning techniques to automatically classify speech into healthy (HC), with mild cognitive impairment (MCI) or with Alzheimer's disease (AD). Only acoustic parameters are used for this purpose making this a language independent approach. Speech parameters provided by the emorlarge feature set have been used. The performance of a three class Random Forest classifier is compared with our proposed system comprising multiple two-class Random Forest classifiers cascaded to form a three class classifier, wherein a combination of approximate posterior probabilities is used to obtain a final class probability estimate. Two methods of classification have been explored, (a) Speech turns belonging to the patient are concatenated and classified as a whole. (b) By treating each patient turn within a clinician-patient conversation as a single utterance and classified separately, the complete utterance is then classified into the class to which maximum number of turns belong. We evaluate our work using dementia and healthy speech from Pitt corpus from DementiaBank database. This system shows an absolute improvement of 8% over the three class classifier. Improvement in F-score was seen for each class with MCI improving the most. Future work would involve an investigation into speech features such as linguistic features to improve the classification.

References

[1] Alzheimer's Association, "2017 Alzheimer's Disease Facts And Figures," 2017, https://www.alz.org/documents_custom/2017-facts-and-figures.pdf.

[2] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh, ""mini-mental state": A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189 – 198, 1975.

[3] S. Ahmed, A. Haigh, C. A. de Jager, and P. Garrard, "Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease," in *Brain : A journal of neurology*, 2013.

[4] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[5] M. Yancheva, K. Fraser, and F. Rudzicz, "Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias," in *6th SLPAT*, 2015, pp. 134–139.

[6] Gábor Gosztolya, László Tóth, Tamás Grósz, Veronika Vincze, Ildikó Hoffmann, Gréta Szatlóczki, Magdolna Pákáski, and János Kálmán, "Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection," in *INTERSPEECH*, 2016, pp. 107–111.

[7] R. Sadeghian, J. D. Schaffer, and S. A. Zahorian, "Speech processing approach for diagnosing dementia in an early stage," in *INTERSPEECH*, 2017, pp. 2705–2709.

[8] C. Thomas, V. Keselj, N. Cercone, K. Rockwood, and E. Asp, "Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech," in *IEEE International Conference Mechatronics and Automation, 2005*, 2005, vol. 3, pp. 1569–1574 Vol. 3.

[9] S. Wankerl, E. Nöth, and S. Evert, "An N-gram based approach to the automatic diagnosis of Alzheimer's disease from spoken language," in *INTERSPEECH*, 2017.

[10] A. Satt, R. Hoory, A. König, P. Aalten, and P. H. Robert, "Speech-based automatic and robust detection of very early dementia," in *INTERSPEECH*, 2014.

[11] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, and R. David, "Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112 – 124, 2015.

[12] S. Al-Hameed, M. Benaissa, and H. Christensen, "Simple and robust audio-based detection of biomarkers for Alzheimer's disease," in *7th SLPAT*, 2016, pp. 32–36.

[13] D. M. J. Tax and R. P. W. Duin, "Using two-class classifiers for multiclass classification," in *Object recognition supported by user interaction for service robots*, 2002, vol. 2, pp. 124–127 vol.2.

[14] J. T. Becker, F. Boller, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis," vol. Archives of Neurology, 51(6), pp. 585–594, 1994.

[15] E. Florian, W. Felix, G. Florian, and B. W. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *ACM Multimedia*, 2013, pp. 835–838.

[16] E. Frank, M. A. Hall, and I. H. Witten, "The Weka workbench. Online appendix for data mining: Practical machine learning tools and techniques," 2016, Morgan Kaufmann, Fourth Edition.