ICASSP 2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

Poster #: 3882

A Comparative Study of Acoustic and Linguistic Features Classification for Alzheimer's Disease Detection

Jinchao Li[1], Jianwei Yu[1], Zi Ye[1], Simon Wong[1], Manwai Mak[2], Brian Mak[3], Xunying Liu[1], Helen Meng[1]

[1]*The Chinese University of Hong Kong,* [2]*The Hong Kong Polytechnic University,* [3]*The Hong Kong University of Science and Technology*

Introduction

Motivation:

- Detection of Alzheimer's disease (AD) is crucial for timely intervention to slow down disease progression
- Speech offers non-intrusive, accessible and affordable means for detecting AD through automatic analysis of selected acoustic and linguistic features
- Most existing AD detection methods are based on manual transcripts [Balagopalan et al. 2020, Sarawgi et al. 2020, Yuan et al. 2020], and most acoustic feature-based methods need to be improved [Luz et al. 2020]

Current work: Presents a comparative study of selected features and classifiers for AD detection and highlights the use of automatically transcribed speech with AT-LSTM for improving detection

Approach

Overview:

- Input: speech audio and manual/automatic text
- Feature extraction
- Feature selection
- Classification

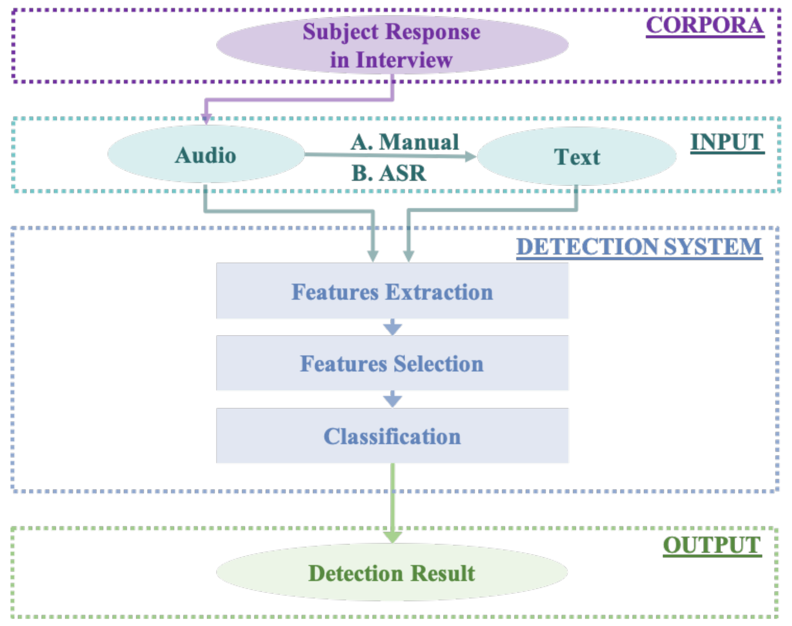

Fig 1. Overall process for AD detection.

Dataset: ADReSS Corpus [Luz et al. 2020].

- 156 speech samples and associated transcripts from AD and non-AD participants
- Divided into balanced training and testing subsets

		AD	non-AD
Train	Male	24	24
	Female	30	30
Test	Male	11	11
	Female	13	13

Tab 1. Composition of ADReSS dataset.

Benchmark features:

- ComParE [Weninger et al. 2013], which are low-level descriptors with temporal and voicing related features
- Linguistics [Snyder et al. 2018], which are linguistic measures, including POS, type-token ratio, etc.

Approach

Features Selection:

- Pearson's Correlation Test (threshold of 0.25).

Added in current work:

- X-vector [Snyder et al. 2017], which are Time Delay Neural Network (TDNN) embeddings for speaker verification
- TF-IDF [Ramos et al. 2003], which is textual vector representation.
- BERT [Devlin et al. 2018], which are bidirectional encoder representations from transformer.

Dimensionality Reduction:

- Principal Components Analysis with n_components = min(feature dim, data size)

Classifiers:

- Linear Discriminant Analysis (LDA)
- Support Vector Machine (SVM) with soft margin and linear kernel
- Attention-based Long Short-Term Memory Recurrent Neural Network (AT-LSTM) [Wang et al. 2018] cross-entropy loss and L-2 regularization

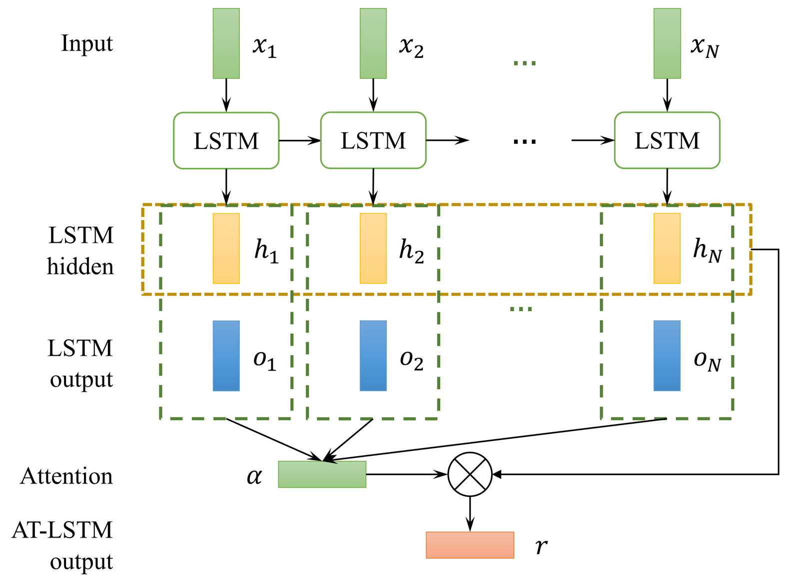

Fig 2. Architecture of AT-LSTM.

Experimental Setup

ASR [Ye et al. 2021]

- Two ASR systems with respectively participant WER 44.89% and 33.17% on the ADReSS test subset.

Training for AT-LSTM

- Adam optimizer with 0.0005 learning rate and 0.01 weight decay.
- Dropout rate of 0.2, early stopping of 16 epochs, batch size of 32.

Evaluation

- The detection systems are evaluated by 10-fold cross-validation (CV) on the train set and evaluated on test set.
- Metrics of accuracy(ACC), precision (PRE), recall (REC), F1 and Area Under Curve (AUC) scores.

Experimental Results

Model	Feature	ACC	PRE	REC	F1	AUC
LDA [13]	ComParE	0.56 / 0.62	0.57 / 0.60	0.52 / 0.75	0.54 / 0.67	N/A
LDA [13]	Linguistics	0.77 / 0.75	0.77 / 0.83	0.76 / 0.62	0.77 / 0.71	N/A
SVM [9]	BERT	0.82 / 0.83	0.84 / 0.81	0.79 / **0.88**	0.81 / 0.84	N/A
LDA	ComParE	0.66 / 0.65	0.65 / 0.64	0.62 / 0.62	0.64 / 0.64	0.71 / 0.66
	X-vector	0.63 / 0.58	0.62 / 0.59	0.66 / 0.54	0.62 / 0.57	0.66 / 0.63
	Linguistics	0.81 / 0.83	0.86 / 0.94	0.73 / 0.71	0.78 / 0.81	0.90 / 0.90
	TF-IDF	0.76 / 0.71	0.79 / 0.81	0.73 / 0.54	0.74 / 0.65	0.84 / 0.88
	BERT	0.76 / 0.79	0.74 / 0.79	0.80 / 0.79	0.76 / 0.79	0.83 / 0.89
SVM	ComParE	0.71 / 0.58	0.73 / 0.62	0.68 / 0.42	0.68 / 0.50	0.76 / 0.60
	X-vector	0.61 / 0.58	0.62 / 0.60	0.61 / 0.50	0.60 / 0.55	0.62 / 0.62
	Linguistics	0.80 / 0.83	0.82 / 0.90	0.75 / 0.75	0.76 / 0.82	0.89 / **0.90**
	TF-IDF	**0.86** / 0.71	**0.91** / 0.73	**0.82** / 0.67	**0.85** / 0.70	**0.93** / 0.83
	BERT	0.75 / **0.88**	0.74 / **0.91**	0.79 / 0.83	0.75 / **0.87**	0.83 / 0.89
AT-LSTM	ComParE	**0.80** / 0.64	0.81 / 0.64	0.80 / 0.64	0.79 / 0.64	0.87 / 0.71
	X-vector	0.58 / **0.67**	0.58 / 0.66	0.65 / 0.69	0.59 / 0.67	0.65 / 0.71
	Linguistics	0.82 / 0.81	0.88 / 0.88	0.76 / 0.73	0.79 / 0.79	0.90 / 0.88
	TF-IDF	0.82 / 0.66	0.84 / 0.67	0.79 / 0.65	0.80 / 0.66	0.87 / 0.77
	BERT	0.80 / 0.83	0.80 / **0.91**	0.80 / 0.74	0.78 / 0.81	0.89 / **0.90**

Tab 2. Results of proposed vs. benchmark models, formatted as CV / Test scores.

System	Feature	ACC	PRE	REC	F1	AUC
Sys. 4 (0.45)	TF-IDF	0.69	**0.74**	0.58	0.65	0.85
	BERT	**0.79**	0.72	**0.96**	**0.82**	**0.87**
Sys. 10 (0.33)	TF-IDF	0.69	0.74	0.58	0.65	0.82
	BERT	**0.88**	**0.82**	**0.96**	**0.88**	**0.92**
Manual	TF-IDF	0.71	0.73	0.67	0.70	0.83
	BERT	**0.88**	**0.91**	**0.83**	**0.87**	**0.89**

Tab 3. Test results of manual vs. ASR-based features.

Discussion & Conclusions

- Presented a comparative study of different acoustic and linguistic features extracted from transcribed speech for AD detection using different classifiers.
- Results reflect viability of using speech for AD detection, extracted linguistic features outperform acoustic features, and feature selection methods are useful for improving the performance.
- Results indicate the feasibility of a fully automatic AD detection from speech based on ASR-derived transcriptions.
- Future work will include applying the system to Cantonese data and improving detection performance by features fusion methods.

Acknowledgement

- This project is partially supported by the HKSARG Research Grants Council's Theme-based Research Grant Scheme (Project No. T45- 407/19N).