An Approach for Assisting Diagnosis of Alzheimer's Disease Based on Multi-Model Features of Narrative Speech

Liu Ning
Hangzhou normal university https://orcid.org/0000-0002-9242-506X
Qingfeng Tang (tqf1013@sina.com)
University Key Laboratory of Intelligent Perception and Computing of Anhui Province
Kexue Luo
Tongde Hospital Of Zhejiang Province

Research article

Keywords: Alzheimer's disease, Natural Language Processing, linguistic feature, bert embedding, machine learning, acoustic feature

Posted Date: March 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-344336/v1

An Approach for Assisting Diagnosis of Alzheimer's Disease Based on Multi-Model Features of Narrative Speech

Liu Ning [1,2], Qingfeng Tang[3*], Kexue Luo [4]

1 School of Health Management, Hangzhou Normal University, Hangzhou, 311121, China

2 Department of mathematics and computer science, Quanzhou Normal University, Quanzhou, 362000, China

3 the University Key Laboratory of Intelligent Perception and Computing of Anhui Province. Anqing, 246113, China

4 Tongde Hospital of Zhejiang Province Geriatrics, attending doctor, Hangzhou, 362000,China

* Correspondence: tqf1013@sina.com; Tel.: 15856575232

Abstract

Background: Alzheimer's Disease (AD) is a common dementia which affects linguistic function, memory, cognitive and visual spatial ability of the patients. More and more studies have been done to access non-invasive, accessible, cost-effective methods for the detection of AD, Speech is proved to have relationship with AD, so a time that AD can be diagnosed in a doctor's office is coming.

Methods: In our study, the ADRess dataset in 2020 was used to detect AD which was balanced in gender and age. First we extract three categories of feature parameters: acoustic feature extracted by opensmile software, bert embeddings automatically and complicated linguistic feature extraction manually. Linguistic features are based on the POS tag, lexical Richness, fluency, semantic feature. Then seven different classifiers are used for identifying AD from normal controls, including SVM, Logistic Regress, Random forest, Extra Trees, Adaboost, LightGBM and a novel ensemble approach with majority voting strategy which is applied to overcome the error caused by a base classifier. Finally ten-fold cross validation is adopted for the evaluation of our approach. In addition, individual features and their combine features are fed to six base classifiers and ensemble of classifier.

Results: We get top-performing classify result on the test set with ensemble of classifiers, the best accuracy of which is 85.4%. The best performance of feature sets are linguistic features, the accuracy of which is 85.6% with LightGBM classifier, and SFS approach is used to manifest seven discriminative linguistic features.

Conclusions: The statistical and experimental results illustrates the feasibility by using speech to predict AD effectively based on acoustic and linguistic feature parameters. Stronger classifier and discriminate features are vital for the final results. We emphasise the best linguistic features for predicting AD disease are based on the POS tag, lexical Richness, fluency, semantic feature. Ensemble of classifiers usually has a better performance than single classifier.

Keywords: Alzheimer's disease; Natural Language Processing; linguistic feature; bert embedding; machine learning; acoustic feature

Background

With the population ages, the number of people with Alzheimer's disease (AD) is increasing. Jia Longfei [1] recently reported in Lancet Public Health that there were 15.07 million people over 60

years old sufferring from dementia in China, including 9.83 million AD patients, 3.92 million vascular dementia and 130,000 other dementia. In the meanwhile, there are over 500 million AD sufferings in America now (Https://www.alz.org/alzheimers-dementia/facts-figures). It is estimated that by 2030, 7,600 million people will be diagnosed with AD or other dementias. AD has gradually become a world wide problem. AD is a chronic, progressive disease characterized by losing the ability independently in daily life gradually. Although clinicians can differentiate people with AD from healthy controls by a combination of cognitive test scales [2], it is time-consuming and little uniform in selected measures. So it is essential to find a more reliable but simple test method to aid differentiating different cognitive people, especially for the early diagnosis of AD.

As a part of higher brain function, language function has an effective relationship with cognitive function [3]. Discourse represents one's psychological activities, which can manifest clearly the complicated relationship among cognitive, language and communication [4]. As the discourse reflects speaker's intention and attention, researchers have found that AD disease has great relationship with linguistic function [5]. The universal performance of AD sufferings is language barrier [6], the symptom of discourse disorder appears even earlier than memory and orientation damage, so Caplan [7] pointed that the most common method to study the relationship between language and brain is to analyze speech disorder caused by brain damage. Snowdon D [8] has demonstrated that low language proficiency is an important index for people with cognitive impairment in daily life, so language maybe a better identify indicator compared with other methods such as memory, study and cognitive function, automatic detection and screening method based on speech has great potential for AD patients.

Previous studies on the relationship between linguistic complexity and AD had demonstrated that low language ability had great relationship with cognitive impairment [12]. Larrieu S [13] and Howieson D [14] had found that some people will steady from mild cognitive impairment (MCI) to AD while others will remain stable for many years, and even a minority can return to normal cognitive status. Some studies attempt to quality the linguistic impairments by using computational techniques, as there are still many differences in the spoken language to supply discriminative markers. Guinn [15] took filled pauses, repetitions, and incomplete words as linguistic features, which was proved discriminative than POS tags and measures of lexical diversity, and finally got an accuracy of 79.5%. By using praat software, Meil´an [16] extracted acoustic features such as number of voice breaks, shimmer, number of periods of voice, the percentage of voice breaks, noise-to-harmonice ratio and so on, and got accuracy of 84.8% finally in distinguish 30 AD patients from 36 healthy controls. Jarrold [17] found that AD sufferings like to use more verbs, adjectives and pronouns and less nouns than healthy controls, and got a best accuracy of 88% by using POS features, acoustic features and psychologically motivated word lists. Orimaye [18] found that AD patients used less syntactic components and higher significantly lexical components. Yancheva [19] and Fraser [20] extracted 477 acoustic, semantic and lexico-syntactic features in a cookie theft picture task, forty most informative features in demantiabank database, finally the accuracy reached over 92% in distinguishing AD from healthy controls.

To summarize, precious work in this area mainly used two methods. The first one is to build feature engineering and then use machine learning classifier to recognize AD properly, which needs many expertise knowledge in order to get distinguishing features, so the integrity of features can not

be guaranteed, and features extraction mostly based on grammar, semantic and pragmatic and so on [9-11]. The second method is deep learning model, which uses powerful neural networks with multiple hidden layers to solve general machine learning tasks without feature engineering. Deep neural networks can learn representations from data by using cascades of multilevel nonlinear processing units for feature extraction. The manifestion of deep learning is better but interpretability is not better than the first method, which is of great importance for clinical diagnosis, however. As an unclear relationship between brain neurocognitive mechanism and language itself, the development of AD linguistics has affected. While the linguistic analysis of AD patients, especially, the relationship between linguistic features and patients' brain impair area, may explore the internal AD pathogenesis. So linguistic feature extraction has great significance for the diagnosis and treatment of AD patients, and we believe that oral linguistic markers to diagnose AD is a compelling method for future research.

Many work had done by using ADReSS dataset (including classification and MMSE prediction), while we are only intrested in the classification task between AD and normal controls [21][22][23]. The champion is Yuan [21] who used bert embeddings combined with encoded pauses. A Pompili [24] used both acoustic and textual feature embeddings and attained 81.25% accuracy in ADReSS challenge.

In all, most researches manifest that, by extracting clinical markers of acoustic and transcripts, the classification among AD, MCI and healthy controls is practicable. These study includes a small amount of specialized classifications, such as PPA, vascular dementia and so on. The task includes most common picture description task (e.g. Cookie Theft picture) or recall or repeat some stories(e.g. cinderella's story) and so on. The database used mainly includes public database such as Dementiabank or self-built corpus. transcript includes manual or automatic ways, and the accuracy of these studies is between 80%-92% or so.

In this work, we present the multi-modal system to the topic. We exploited syntatic, lexical and semantic features with measures of global and theme coherence, we also tried widely acclaimed bert method and acoustic features which is easily to get relatively by software and then combine different features in order to get the state-of-the-art statement. In the meanwhile, we used a novel ensemble approach with majority voting mechanism and got a best accuracy which is lower 4.2% than the champion (the accuracy is 89.6% [21]) in the ADReSS challenge. We try different model feature extracted method so as to provide a more comprehensive characterization in AD for speech and linguistic abilities and a more reliable identification.

The main purpose of our study is to explore different analyse measures to identify acoustic and linguistic feature unique to AD, and then classify them correctly by a stronger classifier. The main contributions of our work include the following four aspects.

(1) We combine the methods of the state of the art deep learning (bert embeddings) automatically and phonological / linguistic features manually in order to find suitable discriminative features.

(2) We extracted linguistic complicated measures manually from cookie theft picture task and used SFS method to find best discriminative features for different subjects in order to improve the interpretability of model.

(3) The linguistic features we extracted are based on the POS tag, lexical Richness, fluency, semantic feature, which got the best results for single classifier.

(4) It is the first time to use ensemble of Classifiers with majority voting strategy on features of narrative speech and language to discriminate AD from normal controls, and get best accuracy by this means.

Methods

One of the main contribution of the study is the idea of ensemble framework for AD prediction by using majority voting strategy. The approach includes three phases: preprocessing, feature extraction, ensemble of classifiers to get the final result. The detailed working of our method is shown in Figure 1.

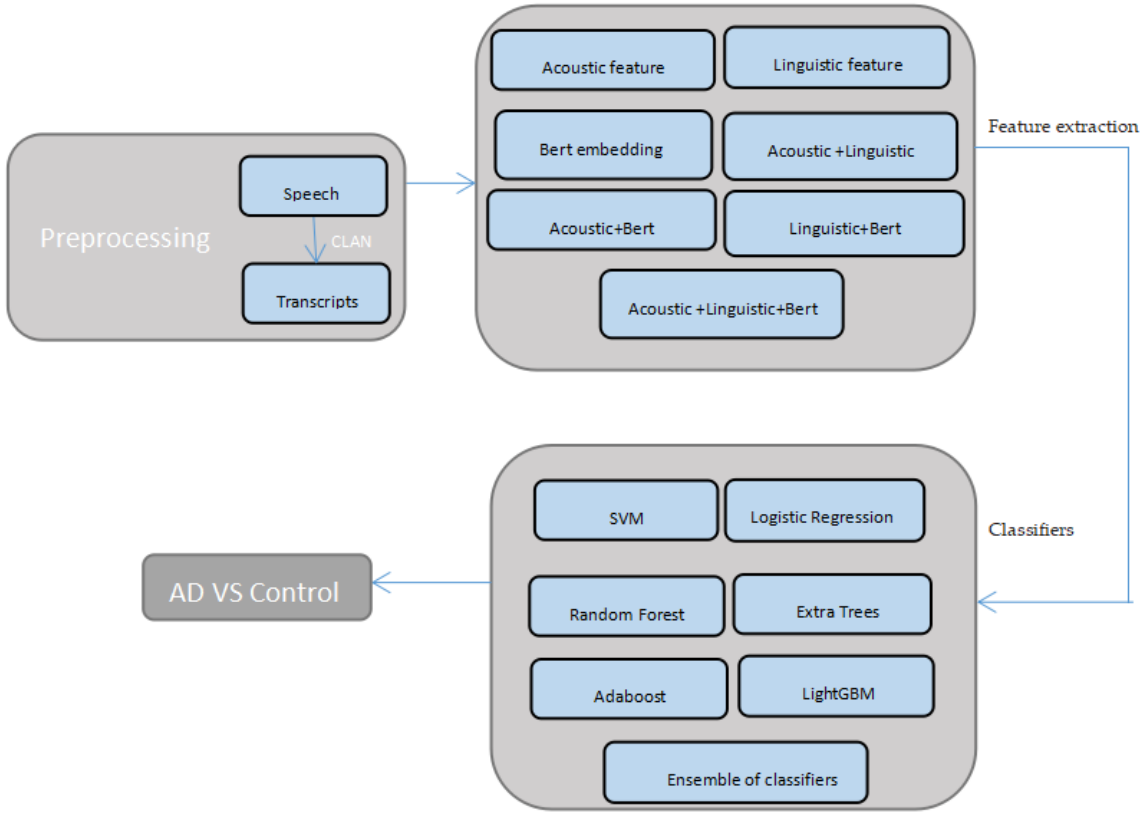


Figure 1. Detailed working of proposed approach

ADReSS Dataset

The data used in this study came from ADReSS Challenge in 2020. The data is composed of speech recordings and transcripts of the Boston Diagnostic Aphasia Exam [25], which is picture description task through the cookie theft picture. By using CHAT coding system [26], the text is transcribed and then annotated. The speech was segmented using a voice activity detection method based on signal energy value. The dataset includes 2,122 acoustic segments from 78 AD sufferings and 1,955 acoustic segments from 78 normal controls. All datasets have already been pre-processed by removing noise and normalling speech volume, and the corpus includes many user-defined tags. With those caveats in mind, the composition of the full dataset is shown in Table 1. the average and standard deviation(SD) of age and MMSE is shown in Table 2. There is not many differences in age between two groups, and both of the number of two groups is 78.

As the proprecessing process has already been dealt with by initiator, including transcripts and annotation, we do not need to do the first step again.

Table 1. Basic characteristics of the patients in every group

	AD		Non-AD	
Age Interval	**Male**	**Female**	**Male**	**Female**
[50,55)	2	0	2	0
[55,60)	7	6	7	6
[60, 65)	4	9	4	9
[65, 70)	9	14	9	14
[70, 75)	9	11	9	11
[75, 80)	4	3	4	3
Total	35	43	35	43

Table 2. The average and standard deviation of age and MMSE

	CTRL(n=78)		Dementia(n=78)	
Measure	Avg	SD	Avg	SD
Age	66.56	6.60	66.79	6.83
MMSE	29.01	1.16	17.79	5.48

Feature extraction methods

The feature extraction mainly includes three sections (acoustic features, bert embeddings and linguistic features manually) in this study. Linguistic features manually will be described in detail because it is most complicated relatively.

Acoustic features

We used 384 acoustic features which can got from opensmile software [27]. The method was proposed by Bjorn [28] in 2009 InterSpeech challenge. We simply describe the extract process of acoustic features. First it calculate 16 LLD, including the zero rate, Square root of energy, F0, HNR, MFCC1-2 and so on. Then the first order differential of 16 LLD is calculated and get 32 LLD. At last, 12 statistical function is applied on 32 LLD and got 32 * 12 = 384 features.

Linguistic Features on BERT Embeddings

A pre-trained Bidirectional Encoder Representations from Transformers (BERT) [29] model is used as a feature extractor. BERT models is a new pre-training language representations which obtains the state-of-the-art representation in many Natural Language Processing (NLP) tasks, the

performance outperforms other methods as it is the first deep bidirectional system for NLP. There are six layers in the whole architecture and every layer has two layers, one is self-attention layer, which is multi-head-Attention, the other is a fully connected layer. We just use self-attention layer to extract bert embeddings and get high level word embedding representations which can capture universal across different tasks information. We just use the encoder of transformer, composed of many transformer blocks, as feature extractor. The position embedding and word embedding are the input embedding, then we use BERT model and a pytorch deep learning framework to extract embeddings, the dimension of which is (156, 768) for every dialogue of every speaker in the transcripts, where 156 is the length of the dataset, 768 is the size of hidden size. The encoder structure of transformer is shown in Figure 2, x_1 and x_2 are input words. The configuration of BERT we used is shown in Table 3.

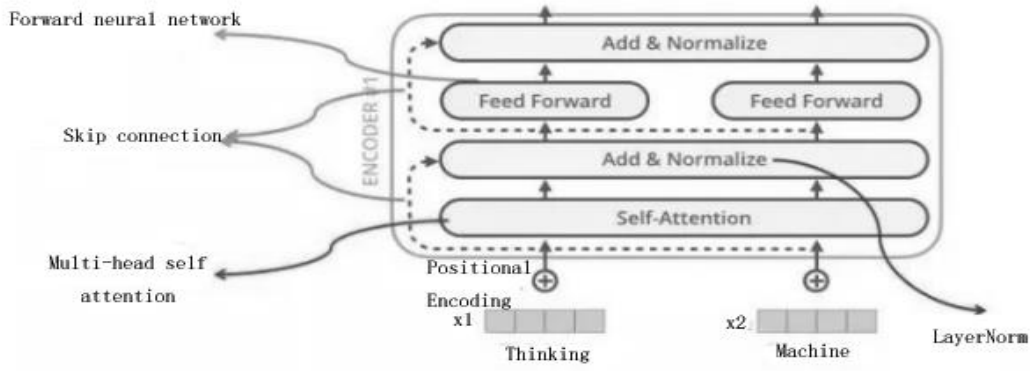


Figure 2. The encoder of transformer

Table 3. The configuration parameters of BERT

Parameter	Value
architectures	BertForMaskedLM
attention_probs_dropout_prob	0.1
hidden_act	gelu
hidden_dropout_prob	0.1
hidden_size	768
initializer_range	0.02
intermediate_size	3072
layer_norm_eps	1e-12
max_position_embeddings	512

model_type	bert
num_attention_heads	12
num_hidden_layers	12
pad_token_id	0
type_vocab_size	2
vocab_size	30522

Language analysis manually

Many methods are used to measure linguistics. For example, Frazier [30] proposed to track a path from a word up a tree until it reached either lowest node which isn't the leftmost child of their parent or the root of the tree. Rosenberg [31] used level based on presence of specific grammatical constructions. H. Scarborough [32] extracted linguistic features by using Index of Productive Syntax which was a scale used for language of child transcript analysis in recent years. Some approaches of assess phonological, lexical, semantic and semantic abilities based on language [33][34][35][36] have already been suggested. Different POS can detect some linguistic changes in different cognitive impairment. For instance, Jarrold [37] and Bucks [38] found an increase in the proportion of verbs, adjectives and pronouns and a decrease in the proportion of nouns for AD patients. Ahmed [39] found a change in the number of verbs and pronouns. AD sufferings often exhibit perseverative behavior in their daily life, including obvious linguistic differences [40] with healthy controls. For example, Tomoeda [41] and Nicholas [42] found that AD sufferings often repeated some words or phrases than normal, while the frequency was not relationship with the severity of disease. Pause is also a common phenomenon for AD patients, the frequency and location of pause can index some important information, for example, the expression of things in some scenes, coherence and organization of language, fluency and coherence are all important linguistic reference characteristics. There is a consistent conclusion that the impairment of word extraction and naming ability in AD patients is much related to the impairment of semantic memory and other factors.

Programs from CLAN [27] were used to extract many linguistic measures. Coding sheet on AphasiaBank (http://aphasia.talkbank.org/) and codes from CLAN manual (http : // childes.psy.cmu.edu/manuals/Clin-CLAN.pdf) were used to capture some mistakes. Additional codes were used to capture some potential deficits in people with AD. For example, the unfilled pause is marked with a symbol (.), (..) and (...). Trailing off is marked with '+…' which means that speaker forgets what they are to say, and then shift attention away from the topic talking about now. 'xxx' means unintelligible words and '[/]' means repetitions.

There is no uniform strandard for linguistic measure, the pros and cons of these linguistic measures have already beyond the scope of our study. After careful consideration, We will expound our extraction method from the following four aspects: Part-of-Speech(POS), lexical richness, fluency and semantic feature. We used Natural Language Toolkit (NLTK), exployed by University of Pennsylvania, to extract POS information of noun, gerund, pronoun, verb gerund

phrase automatically. We then compute the frequency of occurrence of those different POS tagging, normalized by the total number of words in every utterance. In the meanwhile we calculate ratios, for example, pronoun to noun, word-to-sentence, pauses and unintelligible count and so on. In addition, lexical Richness (TTR, ARI, CLI and so on), linguistic fluency, semantic understanding are also as an index in our study. More detailed description of 20 linguistic features is shown in Table 4.

Table 4. Linguistic analysis measures used in this study

Measure	Calculation Method
Part-of-Speech(POS) (8)	These features are used to investigate average rate of occurrence for every POS [43][44] category : Noun, Noun phrase, Verb phrase, Pronoun, Gerund, Verb gerund phase, e.g.: Noun / words. Pronoun-to-noun ratio is the ratio of pronoun to noun. Word-to-sentence ratio is num_words / num_sentences.
Lexical Richness(6) :	measure quantifies the richness of vocabulary or lexical diversity
Type-Token Ratio (TTR)	TTR is a ratio between the total vocabulary (V) which is used in a dialogue to total word count(N) in the dialogue. The formula is:TTR [45] = V / N. TTR is dependent on the text length.
Brunét's Index (Bruten)	Brunét's Index (BI) is unique from Type-Token Ratio because it tries to quatify vocabulary used without considering word count. The formula is : BI [46] = NV(-0.0165) where N is the total text length and V is the total vocabularies used by the participant. The value is usually between 10 to 20, the lower the value,the richer vocabulary the speaker.
R - Honor's Statistic	Honore's statistic(R) [47] is based on the conception that the larger the number of words by the speaker that only occur once, the richer vocabulary used by speaker. Words only used once(V1) and the total vocabulary used (V) have already been proved linearly associated. The formula is: R=100 log N / (1 - V1 / V) where N is text length, Which means, the higher the value of R,the richer that the vocabulary used by speaker.
ARI	Automated readability index (ARI) [48] means the interpretability of text.The calculate formula is: 4.71 * (num_char / num_words) + 0.5 * (num_words / num_sentence) - 21.43, where num_char is the number of letter, num_words is the number of word,num_words / num_sentence is the average sentence length of the article, The more difficult the article,The higher ARI.
CLI	The Coleman-Liau Index (CLI) is similar ARI, which means the difficulty of a text to understand,the formula is : CLI = 0.0588 * L- 0.296 * S - 15.8, L = (num_char / num_words) * 100, S = (num_sentence / num_words) * 100.

MLU	The formula of Mean Language Utterance (MLU) is the total of morphemes divided by total number of utterances in a sample EVAL utility in CLAN [27] provides the value.
Fluency (5)	
count_pauses	The number of pauses
count_unintelligible	The number of unintelligible
count_trailing	Trailing means the number of incomplete sentences
count_repetitions	The number of repetitions
SIM_score	1 - cos(sen1, sen2), calculate the cosine similiar of two sentences: sen1, sen2.
Semantic feature (1)	
Num_concepts_mentioned	The cookie theft biscuit is a dialogue between participants and doctor, the number of concepts mentioned is also an important information. In this section, we divide the theme into 10 subjects. Ten subjects are : list1 = ['mother','woman','lady'] list2 = ['girl','daughter','sister'] list3 = ['boy','son','child','kid','brother'] list4 = ['dish','plate','cup'] list5 = ['overflow','spill','running'] list6 = ['dry','wash'] list7 = ['faucet'] list8 = ['counter','cabinet'] list9 = ['water'] List10 = ['cookie', 'jar', 'stool', 'steal', 'sink', 'kitchen', 'window', 'curtain', 'fall'] We calculate the number of keywords (the list of the ten groups above) in the dialogue.

Table 5 is the average and Standard Deviation (SD) of linguistic features in two different groups. All the features are statistically significant ($P<0.05$) except four features: TTR, Prp_count, VP_count, SIM_score.

Table 5. Linguistic Measure Group Differences

		CTRL (n=78)		AD (n=78)		
Feature ID	Measure	Mean	SD	Mean	SD	P
1	TTR	0.54	0.09	0.53	0.09	0.37

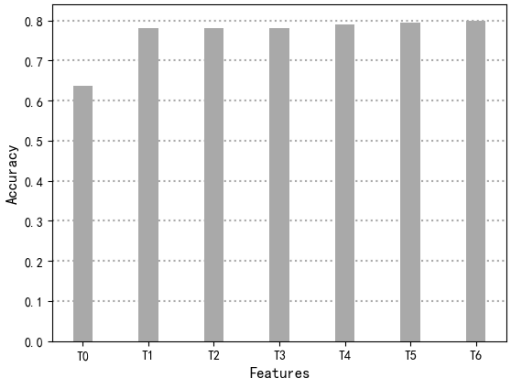
2	R	577.13	38.49	568.98	42.08	1.26e-07***
3	num_concept_mentioned	10.79	3.66	9.77	3.71	2.05e-15***
4	ARI	1.63	1.59	1.40	1.69	7.87e-08***
5	CLI	3.77	1.83	3.44	2.03	7.02e-08***
6	noun_count	22.87	13.45	21.55	14.54	0.0002***
7	vg_count	7.08	3.70	6.4	3.63	9.24e-07***
8	Prp_count	10.52	8.38	11.32	9.22	0.20
9	Prp_noun_ratio	0.49	0.32	0.56	0.35	3.56e-08***
10	Gerund_count	7.08	3.70	6.4	3.63	9.24e-07***
11	NP_count	12.17	7.30	11.22	7.60	0.0006***
12	VP_count	3.88	3.68	4.01	4.17	0.19
13	word_sentence_ratio	8.55	2.75	8.55	2.85	0.011*
14	MLU	7.73	2.65	7.57	2.72	0.007**
15	count_pauses	0.49	0.87	0.62	0.99	0.021*
16	count_unintelligible	0.36	0.97	0.48	1.16	0.008**
17	count_trailing	0.42	0.96	0.59	1.13	1.38e-5***
18	count_repetitions	1.38	2.43	1.68	2.86	0.001**
19	SIM_score	0.76	0.05	0.75	0.06	4.40
20	Bruten	43.17	14.2	41.83	15.21	0.002**

*p<0.05; **p<0.01; ***p<0.001

SFS algorithm

In order to improve the interpretation of the model, we need to identify most discriminant feature sets that influence the final result, we choose sequential forward selection (SFS) [49] method which is an iterative search strategy, it is also a simple greedy algorithm. The feature set starts from no features, the accuracy of the model is calculated by adding a feature at a time at each iteration, and the feature which yields the best result is then chose to add the final feature sets. When adding a new feature can not improve the accuracy of the model, the final iteration will end. In our work, we explored the variation of SFS, that is, the terminating condition was removed as soon as the first maximum is found. Finally seven feature sets which affect 80% performance were found to meet the performance convergence criterion. Figure 3 is the accuracy when adding T0, T1, T2, T3, T4, T5, T6 (T0: prp_noun_ratio, T1:num_concepts_mentioned, T2 : SIM_score, T3 : TTR, T4: noun_count, T5:

word_sentence_ratio, T6: ARI) linguistic feature respectively. T0 (prp_noun_ratio) accounts for about 63.63% performance for all 20 linguistic features, the accuracy improve 14.3% by adding T1. The rate of accuracy increases is slower from T2 to T6, which has only 2% improvement. Figure 4 is the boxplot of the seven features (T0-T6) , from which we can see the different score between AD and Control group.



T0: prp_noun_ratio T1: num_concepts_mentioned T2: SIM_score T3: TTR T4: noun_count T5: word_sentence_ratio T6: ARI

Figure 3. The accuracy of seven discriminative linguistic features on SFS algorithm

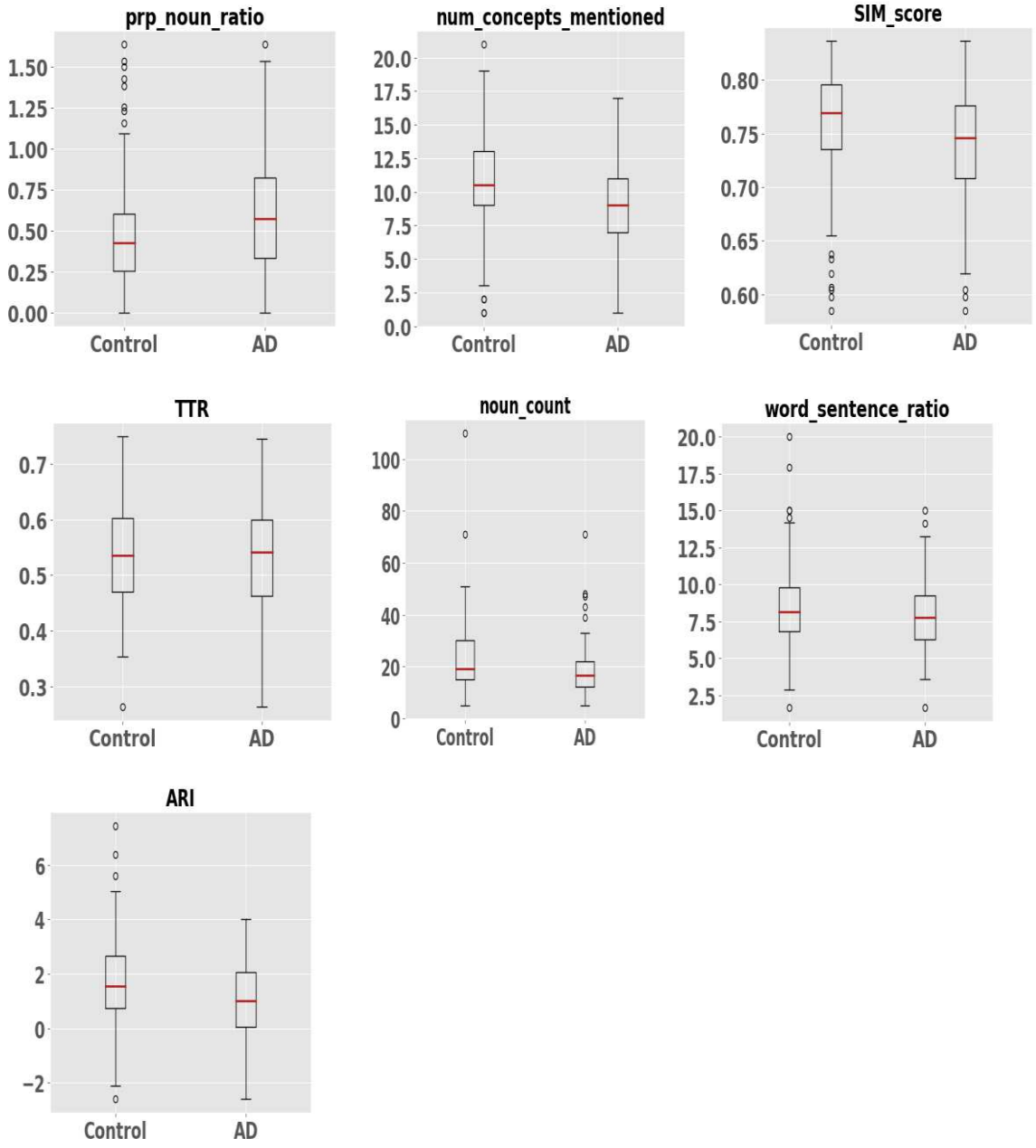


Figure 4. boxplot of seven features between AD and Control

Feature groups include three features above and their combinations: Bert embeddings (B), Linguistic features (L), Acoustic features (A), Acoustic plus Linguistic features (AL), Acoustic plus Bert (AB), Linguistic plus Bert embeddings (LB), Acoustic, Linguistic plus Bert embeddings (ALB). The dimension of acoustic features, linguistic features and bert embeddings is 384, 21 and 768 dimensional vectors, repectively. The dimension of combine features, including AL, AB, LB, ALB is 405, 1152, 789, 1173 respectively. These seven feature sets are feed to the following classifiers.

Classifier

In the following experiments, we choose SVM, Logistic Regression, Random Forest, Extra Trees, Adaboost and LightGBM six classifier to train datasets. The simple description of these classifiers is as follows:

(1) Support Vector Machines (SVM).

SVM are one of supervised classifiers [50] which has better performance in finite data. It work by finding a maximum margin hyperplane which can best seperate two datasets. We used SVM equipped with RBF kernel for our work, the performance with RBF kernel is better than linear kernal for a small dataset.

(2) Logistic Regression

Logisitc Regress establishes the cost function and solves the best model parameters iteratively through the optimal method, then verifies the performance of the model we solved. It is often used in binary classification and popular in industry as its simple, parallelized and strong explanation. The formula of Logistic Regression is:

$$g(x) = \frac{1}{1+e^{-x}} \tag{1}$$

Where x is the training datasets.

(3) Random Forest

Random Forest is a supervised learning algorithms, it is the ensemble of decision trees which train by bagging method in some random way. It can get more accuracy and stable performance by corporate the result of these trees.

(4) Extra Trees

Extremely Randomized Trees (Extra Trees) is a combine method which is similiar to Random Forest, there are mainly two differences:

Random Forest uses random sampling boosting method to choose the training sets of each decision tree, while extra trees use the original training sets.

Random forest chooses a best eigenvalue partitioning points based on gini coefficient or mean square error, while extra trees choose an eigenvalue randomly to divide the decision tree. The variance of Extra Trees is decreased further and bias is further increased than Random Forest, so the generalization ability of Extra Trees is better relatively.

(5) Adaboost

Adaboost (adaptive boosting) is an iterative algorithm, the main idea is to train different weak classifiers for the same training sets, and then assemble these weak classifiers to construct a final stronger classifier. By adjusting the weight of the sample and weak classifier, the classifier with the smallest weight coefficient is selected from the trained classifier to form a final strong classifier.

The run process is as follows: Train every samples and give it a weight which construct a vector D. First the weight is equal, the weight after training will be changed. Adaboost allocates every base

classifier with a weight *alpha* which is calculated based on the error rate of every base classifier, the definition of error rate *m* and weight *alpha* is:

$$alpha = \frac{1}{2} \ln \left(\frac{1-m}{m} \right) \tag{2}$$

Where m is the number of misclassified samples divided by the number of all samples. After calculating the value of *alpha*, the weight vector D will be updated by reducing the weight of samples classified correctly and increasing the weight of misclassified samples.

If the sample is classified correctly, the weight of the sample is:

$$D_i^{(t+1)} = \frac{D_i^{(t)} e^{-alpha}}{Sum\,(D)} \tag{3}$$

While if the sample is classified wrongly, the weight of the sample is:

$$D_i^{(t+1)} = \frac{D_i^{(t)} e^{alpha}}{Sum\,(D)} \tag{4}$$

After calculating D, Adaboost iterates the process again and again and adjusts the weight until the error rate of training is zero or the number of weak classifiers reaches the value specified by the user.

(6) LightGBM [51]

LightGBM is an efficient Gradient Boosting Decision Tree (GBDT) which is a popular machine learning algorithm. It mainly uses Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling (GOSS) algorithm to raise training speed and reduce memory consumption. Every layer of the model is made up of GBDT, and deepened layer by layer. GOSS algorithm (algorithm 1) deletes the samples which have small gradient. The EFB algorithm (algorithm 2 and algorithm 3) reduces the number of the features by binding exclusive features mutually together.

Algorithm 1 : GOSS

Input :

I : Training data; d : iterations ; a : sampling rate of large - gradient data;

b : small gradient data sampling rate;

loss : loss function; L : weak learner.

mod el ← { }, fact ← (1 − a) / b

*1) topN ← a * len (I), randN ← b * len (I)*

2) for i = 1 to d do

3) preds ← mod els.predict (I)

4) g ← loss (I , preds) , w ← { 1 , 1 ,... }

5) sorted ← GetSorted Indices (abs (g))

6) topSet ← sorted [1 : topN]

7) randSet ← RandomPick (sorted [topN : len (I)], randN)

8) usedSet ← topSet + randSet

*9) w [randSet] * = fact*

10) newModel ← L (I [usedSet] , −g [usedSet] , w [usedSet])

11) mod els . append (newModel)

Alg orithm 2 : Greedy Bundling .

Input :

F : feature set; K : maximum conflict count

Output :

Bound sets : bundles.

Structure graph G :

1) search Order ← G.sortByDegree ( )

2) bundles ← { } , bundles Conflict ← { }

3) for i in search Order do

4) need New ← True

5) for j = 1 to len (bundles) do

6) cnt ← Conflict Cnt (bundles [j] , F [i])

7) if cnt bundles Conflict [i] ≤ K then

8) bundles [j] . add (F [i]), need New ← False

9) break

10) if needNew then

11) Add F [i] to bundle sets bundles as a new bundle

Algorithm 3 : Merge Exclusive Features.

Input : The number of data numData, A bundle of binding features about mutually exclusive features F;

Output : new histogram newBin, histogram interval binRanges.

```
1)  binRanges ← { 0 },totalBin ← 0
2)  for f in F do
3)    totalBin += f.numBin
4)    binRanges.append ( totalBin )
5)    newBin ← new Bin ( numData )
6)    for i = 1 to numData do
7)      newBin [ i ] ← 0
8)      for j = 1 to numData do
9)        if F [ j ].bin [ i ] ≠ 0 then
10)         newBin [ i ] ← F [ j ].bin [ i ] +
            binRanges [ j ]
```

(7) Ensemble of classifiers

Ensemble of classifiers is some base classifiers that classify one new instance on voting strategy of base classifiers [52], the main aim of which is to improve the performance. The strategy we adopted is based on majority voting. The result of every base classifier is used as a vote, the final class is one that has majority votes from the base classifers. Every base classifier we used is independent, we just want to known the voting result of many classifiers, so we choose some base classifiers randomly, the strategy of more classifiers maybe a better choice than single classifier.

The base classifiers that we used in ensemble classifer include SVM with a linear kernel coefficient for 'RBF', Logistic Regress (LR), Random forest (RF), Extra Trees (ET), Adaboost (Ada) and LightGBM (LGB). Consider

$$EnsembleClass = \begin{cases} 0, & \text{if } \sum_{i=1}^{n} \text{Base\_Classifer\_Class} > \frac{n}{2}, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

Where "0" stands for AD class, "1" stands for normal controls, "n" is the number of base classifiers. Six base classifiers were trained on seven feature sets and got 6*7= 42 results in all. If the voting result is 21:21, The result of the classifier that perform best in current feature is used as the final result.

Results and Discussion

Model evaluation and Cross-Validation

The computer configuration we used is Intel (R) Core (TM) i7-6700 CPU @3.40GHZ CPU and 8.00GB RAM. The experiment environment in the paper is windows 10 operating system with python 3.7.0 and scikit-learn library. we use a 10-fold cross-validation method in which a 10% test

set is used in every iteration for evaluation, the remaining 90% training set is used to construct the models. The result is the average value across 10 folds. That is to mean, data from any speaker in a given fold maybe training set or test set, but not both. As ADReSS is a balanced dataset, which has 78 AD sufferings and 78 normal, we use accuracy, AUC and F1 value as the final evaluate strandard. The relationship between the actual class and predicted class is shown in Table 6, the evaluate metrics in this study are defined as Eqs.(2) to (6). The classify result is provided in Table 7, 8, 9.

Table 6. Relation between the predicted and true classes.

	True class	
Predicted class	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{6}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{7}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{8}$$

$$FPR = \frac{FP}{FP + TN} \tag{9}$$

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

ROC (Receiver Operating Characteristic Curve) is a coordinate figure, the horizontal axis and vertical axis of which are FPR and TPR respectively. AUC is the area under ROC. The larger the AUC, the more effective a classifier is.

Table 7. Accuracy on different classifiers and feature sets

Features	SVM	LR	RF	ET	Ada	LGB
Acoustic	0.5	0.582	0.61	0.598	0.551	**0.609**
Linguistic	0.629	**0.822**	0.782	0.765	0.771	0.801
Bert	0.606	**0.755**	0.7	0.594	0.659	0.705
AL	0.5	0.576	0.628	0.617	0.731	**0.782**
AB	0.5	0.576	0.664	0.587	0.629	**0.724**
LB	0.752	**0.806**	0.659	0.621	0.754	0.788

ALB	0.5	0.576	0.681	0.701	0.745	**0.788**

Table 8. AUC on different classifiers and feature sets

Features	SVM	LR	RF	ET	Ada	LGB
Acoustic	0.5	0.582	0.61	0.598	0.61	**0.609**
Lingusitic	0.629	0.822	0.782	0.765	0.771	**0.856**
Bert	0.606	**0.755**	0.7	0.594	0.659	0.705
AL	0.5	0.576	0.628	0.617	0.731	**0.850**
AB	0.5	0.576	0.664	0.587	0.629	**0.772**
LB	0.752	0.806	0.659	0.621	0.788	**0.835**
ALB	0.5	0.576	0.681	0.701	0.745	**0.837**

Table 9. F1 on different classifiers and feature sets

Features	SVM	LR	RF	ET	Ada	LGB
Acoustic	0.333	0.574	**0.609**	0.587	0.545	**0.609**
Lingusitic	0.605	**0.819**	0.778	0.757	0.767	0.801
Bert	0.597	**0.753**	0.688	0.578	0.656	0.705
AL	0.333	0.568	0.621	0.594	0.721	**0.782**
AB	0.655	0.566	0.655	0.579	0.619	**0.724**
LB	0.749	**0.803**	0.649	0.602	0.747	0.788
ALB	0.333	0.568	0.677	0.694	0.741	**0.788**

The bold in Table 7, 8, 9 means the best result in a row, from which we can find that linguistic features manifest best, the accuracy of which is 82.2%, although the process is complicated relatively and needs more professional knowledge, while the interpretability is also better. From the manifestion of the classifiers, LightGBM has better performance in Acoustic, Linguistic, AL, AB and

ALB features, Logistic Regress has better performance in Bert and LB features. Horizontally, the performance of acoustic features is lower than Linguistic and Bert for single feature sets. We use four combination, and linguistic plus Bert features get the best accuracy, which is 80.6%. As the performance of acoustic is very low for binary classification (the accuracy of random guess is about 50%), so the combine features including acoustic features are not better relatively. Maybe acoustic features have a bad effect for the final decision-making. Longitudinally the best performance of classifier is lightGBM, and Logistic Regression comes next. The first three lines in Table 7, 8, 9 is the result of three Individual features. We can find that linguistic features manifest best in LR, Bert embedding manifest the next, and acoustic feature is worst. The performance of the combine features is not all better than single features, maybe the dimension of combine features is much higher, for example, the dimension of ALB is 384 + 786 + 18 = 1188, the dimension reduction method such as Principal Component Analysis (PCA) maybe a good choice, just as many studies choose dimension reduction strategy. While in our study we try another approach: ensemble method with maximum voting in order to better classify AD from normal controls.

The result of ensemble of classifiers

In order to compare the results with the champion(the best accuracy is 89.6%), we use accuracy as a general evaluate index. As the datasets is balance, the number of AD and normal is all 78, the threshold of the binary classifier is 0.5. so it is rational to evaluate the performance of classifier by using accuracy. The accuracy of maximum voting is shown in Figure 5, from which we got the best accuracy of 85.4%.

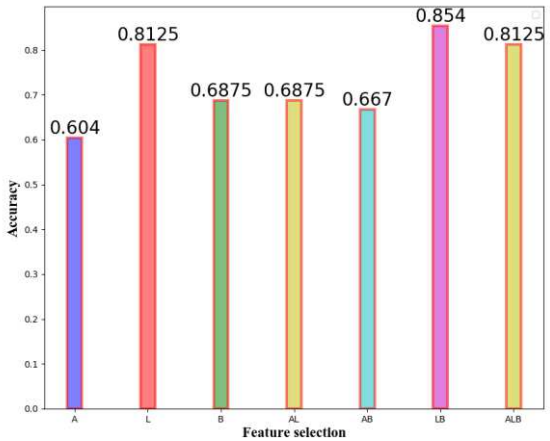


Figure 5. The accuracy of different features with ensemble of classifiers

From the result we can get two conclusions, on the one hand, we use majority voting strategy (the minority is subordinate to the majority) for ensemble approach in this study, which maybe not impartial. The base classifier that has better performance should have more weight, as far as we think. So majority voting maybe not a better strategy. While how to allocate the weight of every base classifier is also a worth-study subject. On the other hand, ensemble approach maybe get better result by getting rid of the feature with poor performance. From Table 6 we known that the

performance of acoustic features worst among three single features. Five combine features: AL, AB, LB, ALB, LB, the best performance is LB feature sets without acoustic features, and ALB second.

Conclusions

This study is an essential step in developing a simple but practical, low-cost reliable tool for the early detection of AD or other dementias based on multi-model Features of Narrative Speech and language. Also, we hope the tool can detect the change of AD gradually with the development of the disease in real time. In our work, we try to explore some speech and linguistic complex measures in order to explore discriminative markers of AD, and then validate the effect of those markers in discriminating clinically different cognitive function groups. We tried three different feature extraction methods, especially, we spend many enery on linguistic feature extraction method manually, and give a detailed explanation for different statistical measures and best seven linguistic features by using SFS algorithm. Finally, through ensemble of classifiers with majority voting strategy in different feature sets, we got a best accuracy of 85.4%.

From the study we can find that acoustic feature extracted by opensmile is easier relatively than other feature extract menthods, as it does not need transcript, while the performance is worst from the final result. Bert embedding features is paid more attention in this area since transformers emerged in 2017, especially the best result of ADReSS with bert structure in 2020. The result with bert embeddings is not bad but not the best, but it is a simple approach. Linguistic features manually is most complicated because first it need transcript and annotation by CLAN, it needs professional knowledge and is hard to comprehensive, the accuracy usually 80%-88% or so in the studies nowadays.

Differences in oral language do supply markers of better discriminative utility for elderly subjects. The topic-focused and narrow use of oral language makes automatic feature extraction to be much accurate for many elders. Based on the above considerations, we believe the use of spoken language markers to diagnose AD is a exploring and compelling area for the further research.

Future, acoustic feature is an explorable area, for example, the importance of pause such as the location and number of pauses in acoustic has already been proved. Deep learning model is also a better study orientation, as the method is easy relatively without complicated feature extraction process, while the performance is usually better, but the explainable is worse of course as deep learning is a “black box” for us after all. We think the following aspects may be the future research idea for the detection of cognitive impairment based on speech:

(1) How early can we detect cognitive impairment by speech method?

With the deepen of the research in this area, we hope to detect cognitive change as early as possible. Early diagnosis means early treatment at the least cost. Now SCD (Subject cognitive impairment) has already been proposed, Some domestic teams, such as Beijing Xuanwu Hospital, are leading the research in this field. Language can detect SCD? This is another worth-study problem, while there is a long way as medical concept on SCD is still in the infancy after all.

(2) Cooperation between different experts in related fields is very important, medical, linguistics and computer should get better deeper study in this area as this is an interdisciplinary research.

(3) We hope the study method in this area, diagnose disease by speech and linguistics, can apply to other neurological diseases.

(4) Multi classification is also a trend, such as common vascular dementia and so on, not just common AD and MCI.

Abbreviations

ADReSS: Alzheimer's Dementia Recognition through Spontaneous Speech; AD: Alzheimer's disease; SCD: Subjective cognitive Decline; MCI: Mild Cognitive Impairment; POS Part-of-Speech; SVM: Support Vector Machine; LightGBM: Light Gradient Boosting Machine; ARI: Automated readability index; R: Honore's statistic; Bruten: Brunét's Index; TTR: Type-Token Ratio; CLI: The Coleman-Liau Index; SD: Standard Deviation; TTR: Type-Token Ratio; Bruten: Brunét's Index; MLU: Mean Language Utterance; SFS: sequential forward selection; B: Bert embeddings; L: Linguistic features; A:Acoustic features; AL: Acoustic plus Linguistic features; AB: Acoustic plus Bert; LB: Linguistic plus Bert embeddings; ALB: Acoustic, Linguistic plus Bert embeddings; ROC: Receiver Operating Characteristic Curve

Acknowledgments

We thanks the support of the school of Health Management, Hangzhou Normal University, for an experimental environment to conduct the study. We also acknowledge the funding support from the big data platform of department of mathematics and computer science, Quanzhou Normal University to facilitate this study.

Funding

This research was funded by initial fee for introducing doctoral research of Anqing Normal University in 2020 (No. 201023).

Availability of data and material

The public available Dementiabank data was supported by NIH grants AG005133 and AG003705. Please visit https://talkbank.org/access/DementiaBank/English/Pitt.html for more details. The processed materials and the data used in this study can be downloaded from github website: https://github.com/lzy1012/Dataset-of-ADReSS-2020-/releases.

Authors' contributions

Conceptualization, Liu Ning; methodology, Qingfeng Tang; software, Qingfeng Tang; validation, Liu Ning and Kexue Luo; formal analysis, Kexue Luo; investigation, Kexue Luo; writing—original draft preparation, Liu Ning; writing—review and editing, Qingfeng Tang; visualization, Kexue Luo; supervision, Kexue Luo; project administration, Qingfeng Tang; and funding acquisition, Qingfeng Tang.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All the authors declare that there is not any competing interests.

References

1. Jia Longfei, Du Yifeng, Chu Lan, et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study[J]. The Lancet Public Health, 2020, 5(12).

2. Mesulam, M. M., Wieneke, C., Thompson, C., Rogalski, E., & Weintraub, S. Quantitative classification of primary progressive aphasia at early and mild impairment stages. Brain, 2012, 135, 1537–1553.

3. Folstein, M., Folstein, S., McHugh, P.,Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J. Psychiatric Res. 12 (3), 189–198. 1975.

4. Ulatowska, Hanna K . The Aging brain : communication in the elderly. College-Hill Press, 1985.

5. Sabat,S.R..Language function in Alzheimer's disease:a critical review of selected literature.Language & Communication, 4, 1994, 331 - 351.

6. Appell,J.,Kertesz,A.,&Fisman,M.Astudy of language functioning in Alzheimer patients. Brain and language,1982,17,73-91

7. Caplan,D. Syntactic and Semantic Structures in Agrammatism. M.L. Kean (Ed.), Agrammatism, New Yark: Academic Press 125-152. [CrossRef]

8. Snowdon D, Kemper S, Mortimer J, Greiner L, Wekstein D, Markesbery W. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. J Amer Med Assoc. 1996; 275(7):528–532.

9. Kemper S, LaBarge E, Ferraro F, Cheung H, Cheung H, Storandt M. On the preservation of syntax in Alzheimer's disease. Archives of Neurol. 1993; 50:81–86.

10. Lyons K, Kemper S, LaBarge E, Ferraro F, Balota D, Storandt M. Oral language and Alzheimer's disease: A reduction in syntactic complexity. Aging and Cogn. 1994; 1(4):271–281.

11. Singh S, Bucks R, Cuerden J. Evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech. Aphasiology. 2001; 15(6):571–584.

12. Snowdon D, Kemper S, Mortimer J, Greiner L, Wekstein D, Markesbery W. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. J Amer Med Assoc. 1996; 275(7):528–532.

13. Larrieu S, Letenneur L, Orgogozo J, Fabrigoule C, Amieva H, LeCarret N. Incidence and outcome of mild cognitive impairment in population-based prospective cohort. Neurology. 2002; 59:1594– 1599. [PubMed: 12451203]

14. Howieson D, Camicioli R, Quinn J, Silbert L, Care B, Moore M, Dame A, Sexton G, Kaye J. Natural history of cognitive decline in the old old. Neurology. 2003; 60:1489–1494. [PubMed: 12743237]

15. Guinn CI, Habash A. Language analysis of speakers with dementia of the Alzheimer's Type. AAAI Fall Symposium: Artificial Intelligence for Gerontechnology,2012, 8-13. [CrossRef]

16. Meil´an JJG, Mart´inez-S´anchez F, Carro J,L´opez DE,MillianMorell L, Arana JM (2014) Speech in Alzheimer's disease: Can temporal and acoustic parameters discriminate dementia. Dement Geriatr Cogn Disord 37, 327-334.

17. Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno Tempini ML, Ogar J (2014) Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology, pp. 27-36. [CrossRef]

18. Orimaye S O , Wong S M , Golden K J , et al. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers[J]. BMC Bioinformatics, 2017, 18(1):34.

19. Yancheva, M., Fraser, K., Rudzicz, F., 2015. Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias. 6th Workshop on Speech and Language Processing for Assistive Technologies. [CrossRef]

20. Fraser, K. C., Meltzer, J. A., Rudzicz, F., 2015. Linguistic Features Identify Alzheimer s Disease in Narrative Speech. Journal of Alzheimer's Disease 49, 407–22.

21. Yuan, J., Bian, Y., Cai, X., Huang, J., Ye, Z., and Church, K. “Disfluencies and fine-tuning pre-trained language models for detection of Alzheimer’s disease,” in Proceedings of Interspeech 2020 (Shanghai), 2162–2166. [CrossRef]

22. Edwards, E., Dognin, C., Bollepalli, B., and Singh, M. “Multiscale system for Alzheimer’s dementia recognition through spontaneous speech,” in Proceedings of Interspeech 2020 (Shanghai), 2197–2201. [CrossRef]

23. Pompili, A., Rolland, T., and Abad, A. (2020). “The INESC-ID multi-modal system for the ADReSS 2020 challenge” in Proceedings of Interspeech 2020 (Shanghai), 2202–2206. [CrossRef]

24. A Pompili, T Rolland, A Abad. The INESC-ID Multi-Modal System for the ADReSS 2020 Challenge. 2020. arXiv:2005.14646.

25. J. T. Becker, F. Boller, O. L. Lopez, J. Saxton, and K. L. McGonigle, “The Natural History of Alzheimer’s Disease,” Archives of Neurology, vol. 51, no. 6, p. 585, 1994.

26. MacWhinney, B. The CHILDES Project: Tools for Analyzing Talk, Volume I: Transcription Format and Programs. New York, NY; Hove, ES: Psychology Press. 2014. [CrossRef]

27. F. Eyben F. Weninger F. Gross and B. Schuller "Recent developments in opensmile the munich open-source multimedia feature extractor" Proceedings of the 21st ACM International Conference on Multimedia, pp. 835-838 2013. [CrossRef]

28. Bjorn Schuller, Steidl S, Batliner A. The Interspeech 2009 Emotion Challenge [J]. Interspeech, 2009:312-315. https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/deliver/index/docId/66757/file/i09_0312.pdf.

29. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: pre-training of deep bidirectional transformers for language understanding, 2018. [CrossRef]

30. Dowty D R , Karttunen L , Zwicky A M . Natural Language Parsing[M]. Cambridge University Press, 1985. [CrossRef]

31. S. Rosenberg and L. Abbeduto, “Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults”, Appl. Psycholinguist., vol. 8, pp. 19–32, 1987.

32. H. Scarborough, “Index of productive syntax,” Appl. Psycholinguist., vol. 11, pp. 1–22, 1990.

33. Lee, L. L. (1966). Developmental sentence types: A method for comparing normal and deviant syntactic development. Journal of speech and hearing disorders, 31(4), 311–330.

34. MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analyses of the Cinderella story. Aphasiology, 24, 856–868.

35. MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analyses of the Cinderella story. Aphasiology, 24, 856–868.

36. Thorne, J., & Faroqi-Shah, Y. (2016). Verb production in aphasia: Testing the division of labor between syntax and semantics. Seminars in Speech and Language, 37(1), 23–33.

37. Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, GornoTempini ML, Ogar J (2014) Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology, pp. 27-36. [CrossRef]

38. Bucks RS, Singh S, Cuerden JM, Wilcock GK (2000) Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. Aphasiology 14, 71-91.

39. Ahmed S, Haigh A-MF, de Jager CA, Garrard P (2013) Connected speech as a marker of disease progression in autopsy-proven Alzheimers disease. Brain 136, 3727-3737.

40. Bayles KA, Tomoeda CK, McKnight PE, HelmEstabrooks N, Hawley JN (2004) Verbal perseveration in individuals with Alzheimer's disease. Semin Speech Lang 25, 335-347.

41. Tomoeda CK, Bayles KA, Trosset MW, Azuma T, McGeagh A (1996) Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. Alzheimer Dis Assoc Disord 10, 204-215.

42. Nicholas M, Obler LK, Albert ML, Helm-Estabrooks N (1985) Empty speech in Alzheimer's disease and flfluent aphasia. J Speech Lang Hear Res 28, 405-410.

43. Holmes, D.I., Singh, S., 1996. A stylometric analysis of conversational speech of aphasic patients. Literary and Linguistic Computing 11 (3), 133–140.

44. Bucks, R. S.; Singh, S.; Cuerden, J. M. & Wilcock, G. K. (2000), 'Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance', Aphasiology 14(1), 71-91.

45. Holmes, D.I., Singh, S., 1996. A stylometric analysis of conversational speech of aphasic patients. Literary and Linguistic Computing 11 (3), 133–140.

46. Raymond A G , Brunet E , Dugast D . Le Vocabulaire de Jean Giraudoux. Structure et Evolution[J]. Modern Language Journal, 1983, 66(1):85.

47. Yun Yuan, Songsong Zhang, Wei Zhang. Readability Index and English Reading Teaching ——Analysis of Long and Difficult Sentences from the Perspective of Core Sentence Theory [J]. Foreign Languages and Literature, 2015, 32(03) : 208-215. [CrossRef]

48. Marcano-Cedeo A , J. Quintanilla-Domínguez, Cortina-Januchs M G , et al. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network[C]// IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society. IEEE, 2010. [CrossRef]

49. V. N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, NY, USA, 1995. [CrossRef] .

50. Guolin Ke1,Qi Meng,et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, NIPS 2017,32,1-9. [CrossRef]

51. L. Rokach, "Ensemble-based classifiers," Artificial Intelligence Review,vol.33,no.1-2,pp.1–39,2010.

Figures

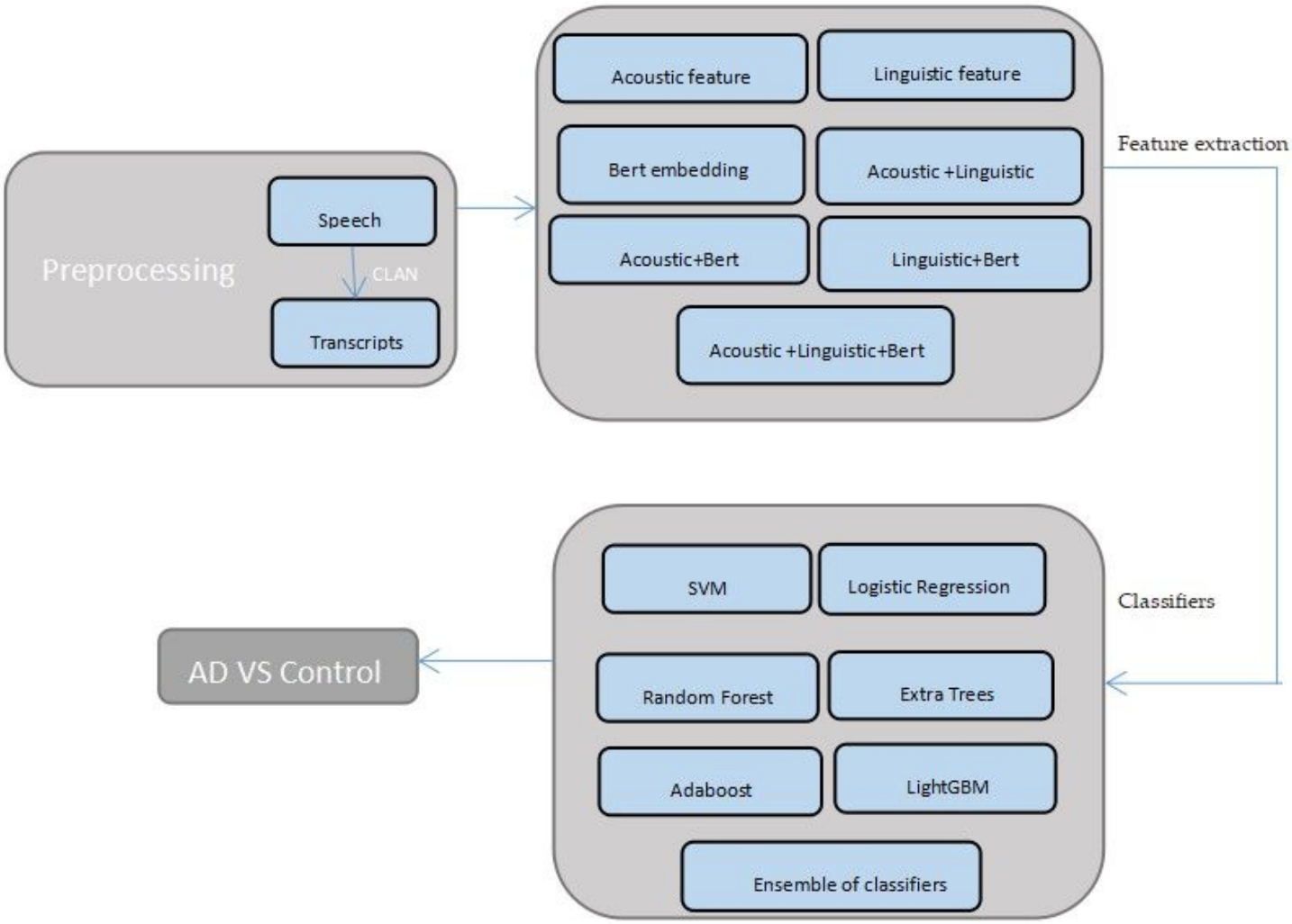


Figure 1

Detailed working of proposed approach

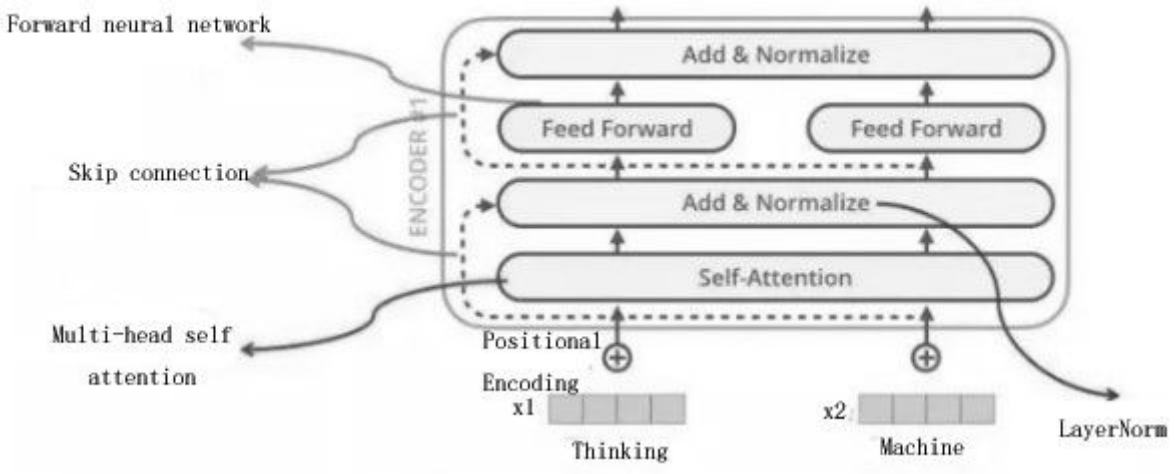


Figure 2

The encoder of transformer

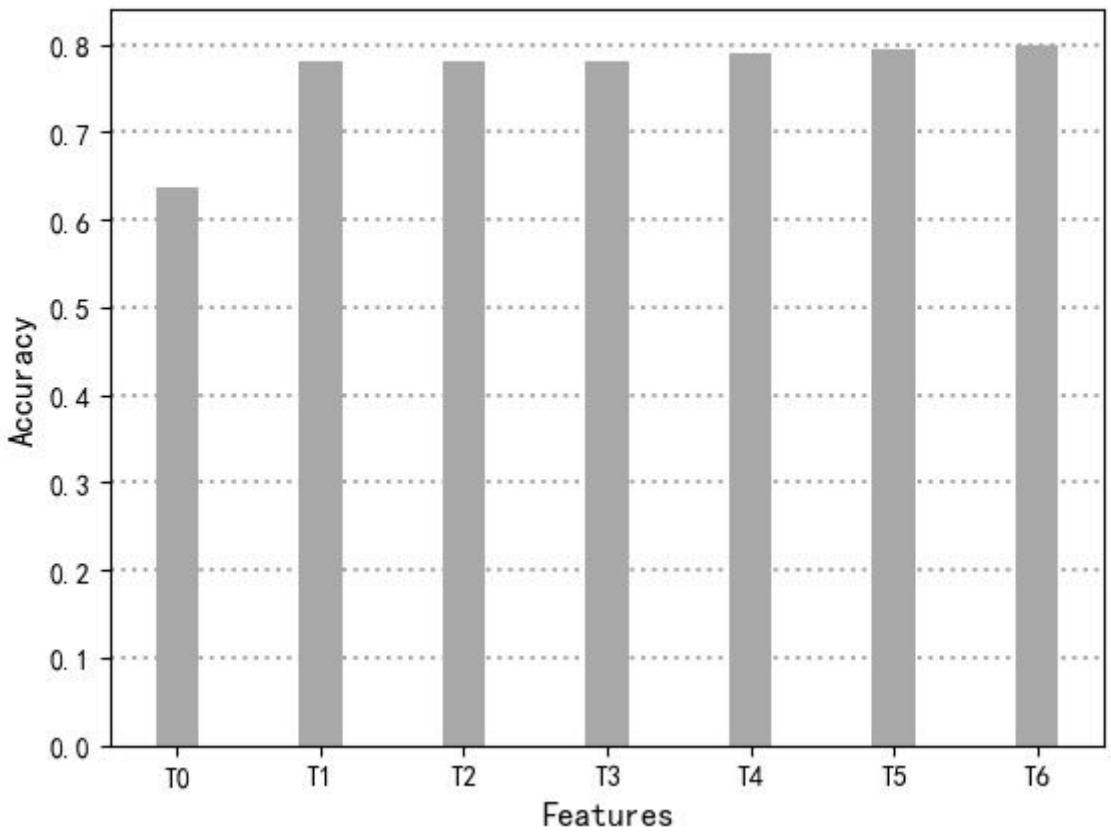


Figure 3

The accuracy of seven discriminative linguistic features on SFS algorithm T0: prp_noun_ratio T1: num_concepts_mentioned T2: SIM_score T3: TTR T4: noun_count T5: word_sentence_ratio T6: ARI

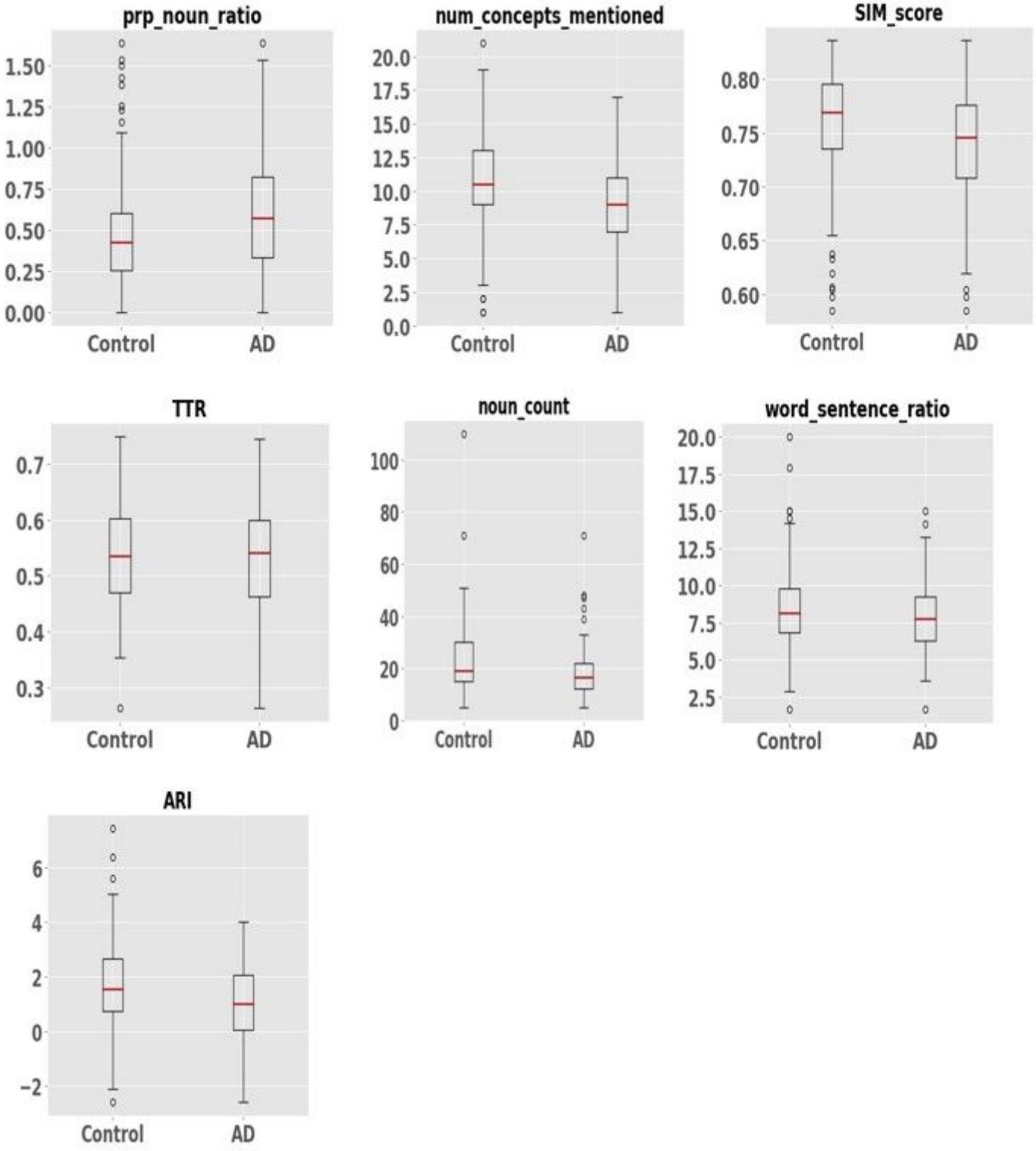


Figure 4

boxplot of seven features between AD and Control

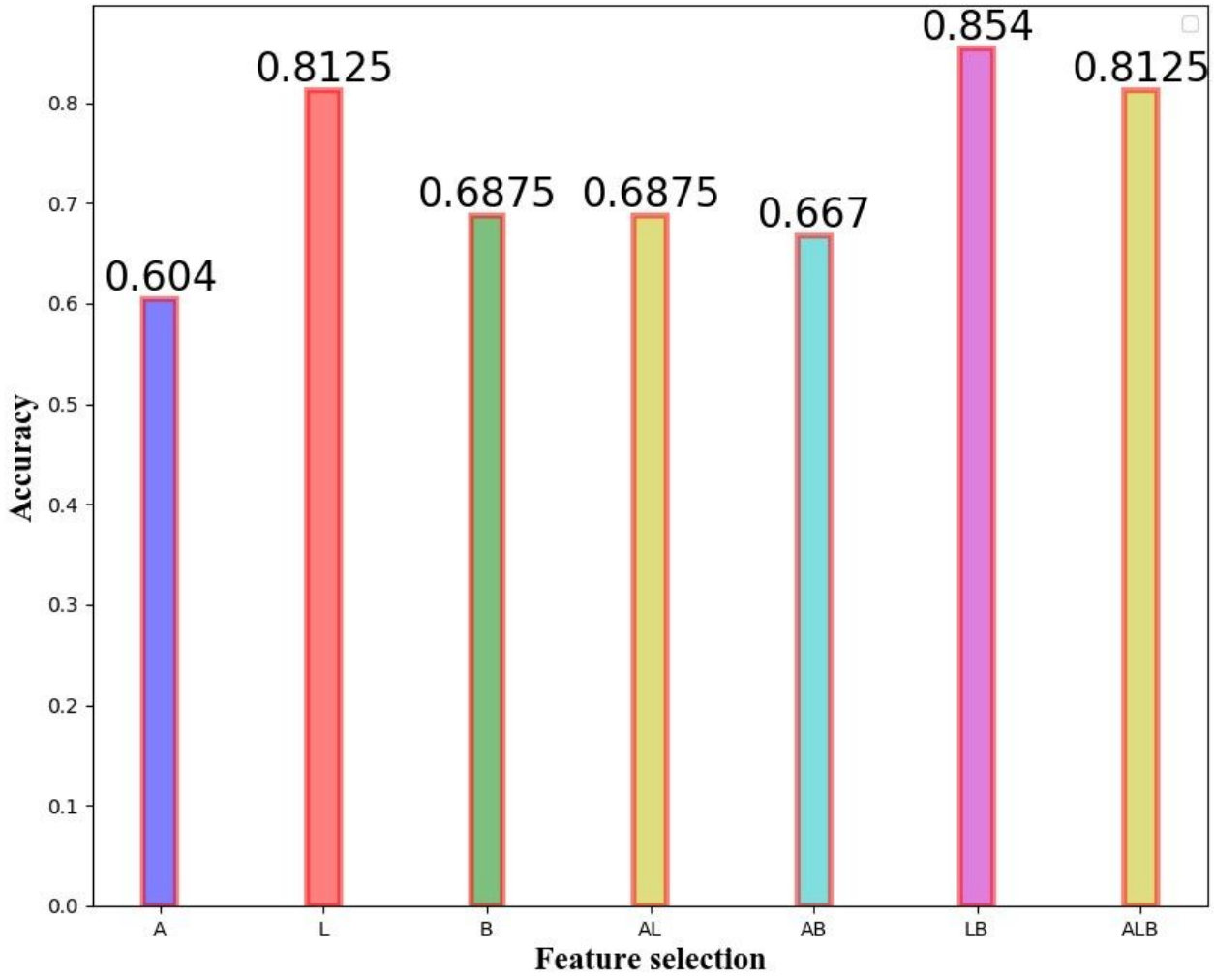


Figure 5

The accuracy of different features with ensemble of classifiers