2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Selecting and Analyzing Speech Features for the Screening of Mild Cognitive Impairment

Qin Yang, Feiyang Xu, Zhenhua Ling, Xin Li*, Yunxia Li, Decheng Fang

Abstract — **The total number of patients with Alzheimer's Disease (AD) has exceeded 10 million in China, while the consultation rate is only 14%. Large-scale early screening of cognitive impairment is necessary, however, the methods of traditional screening are expensive and time-consuming. This study explores a speech-based method for the early screening of cognitive impairment by selecting and analyzing speech features to reduce cost and increase efficiency. Specifically, speech-based early screening models are built based on a feature selection method and a self-built dataset including AD patients, Mild Cognitive Impairment (MCI) patients, and healthy controls. This method achieves 10% relative improvement in F1-score to discriminate MCI patients from healthy controls on our dataset. The prediction F1-score reached 70.73% when discriminating MCI patients from healthy controls based on the feature importance list calculated by the auxiliary model that is built to discriminate AD from Control group. Besides, to further assist the medical screening of MCI, we analyze the correlation between brain atrophy features and speech features including acoustic, lexical and duration features. On the basis of key speech feature selection and correlation analysis, the reference interval of speech features is constructed based on the speech data from Control group to provide a reference for evaluating cognitive impairment.**

Clinical Relevance — **We build a speech-based dataset including AD, MCI and Control groups, and provide a feature selection method to improve the effectiveness of the screening of MCI. Apart from this, the correlation between speech features and brain atrophy features is analyzed. Finally, the reference interval of key speech features is established.**

I. Introduction

Alzheimer's Disease (AD) is the most common neurodegenerative disease. The main symptom of the disease is cognitive impairment, which accounts for more than 60% of all cognitive disorders. Memory impairment, visuospatial dysfunction, and executive dysfunction are, among many, observed clinically in patients with cognitive impairment. Patients with AD suffer from a drastic decline in living standards. Presently, China is the country with the largest number of Alzheimer's patients in the world, where the total number of patients has exceeded 10 million, and there are an average of 300,000 new cases each year. The prevalence of the disease among people over 65 years old is about 5%, and it reaches up to 30% among people over 85 years old. With the increasing aging of the population in China, it is estimated that by 2050, there will be 28 million people with Alzheimer's disease, which will not only negatively affect the population's health but also bring a significant economic burden to society [1]. Studies have shown that early diagnosis and treatment of cognitive impairment will be beneficial for maintaining a healthy brain activity, delaying the development of the disease, and extending the lifespan of patients [2, 3], which is of great help for reducing the family burdens and improving the population's health. Therefore, it is of great social significance to carry out screening for cognitive impairment among the elderly at an early stage.

The traditional methods of diagnosing cognitive impairment, such as scales, brain magnetic resonance imaging, cerebrospinal fluid, have deficiencies in balancing among the accuracy, cost, and convenience of the testing. The lack of professionals in the field, the low accuracy of evaluation as well as the inconvenient intrusive way of data collection from the patients are some of the contributing factors that make the testing of cognitive impairment hard to be widely applied [4]. Medical research shows that speech disorder is one of the typical clinical symptoms of cognitive impairment at the early stage. For example, the speech of patients with mild symptoms shows language characteristics such as difficulty in finding words, repetition, empty speech, vague language, etc. [5]. Such characteristics make it possible to make use of intelligent speech technologies to evaluate cognitive impairment by analyzing the characteristics of speech disorders [6-8].

In the existing studies on speech-based techniques for assessing the cognitive impairment of the elderly, there still exist three deficiencies. First, most existing studies focused on discriminating AD patients from healthy controls, while neglecting the category of MCI. Second, there was no correlation analysis between speech characteristics and the degree of brain atrophy. Third, the reference interval of speech features for cognitive impairment screening has not been built.

(1) In the realms of early screening of cognitive impairment, Koo et al. (2020) [9] based their research on the English dataset DementiaBank, utilizing acoustic and linguistic features and the attention mechanism as well as BiLSTM neural network, and achieved an accuracy rate of 81.25% on the binary classification between AD and healthy. Sarawgi et al. (2020) [10] used feature engineering to compare different ensemble learning methods and achieved 88% classification accuracy on the AD and Control binary classification task. By adopting the adaptive technology based on the ERNIE language model and the coding of the length of the pause, Yuan et al. (2020) [11] used the ensemble voting

Research supported by the China Postdoctoral Science Foundation (Grant No. 2021M693101). Qin Yang, Feiyang Xu and Decheng Fang are with the iFlytek Research, iFlytek Co.Ltd, Hefei, China (e-mail: qinyang5@iflytek.com, fyxu2@iflytek.com, dcfang@iflytek.com).

Zhenhua Ling and Xin Li are with NELSLIP, University of Science and Technology of China, Hefei, China (e-mail: zhling@ustc.edu.cn, leexin@ustc.edu.cn. Corresponding author: Xin Li, phone: +8613865901586; fax: +86055163606791).

Yunxia Li is with Department of Neuology, Shanghai Tongji Hospital, Shanghai, China (e-mail: doctorliyunxia@163.com).

method to achieve a classification accuracy of 89.6% in the ADReSS competition. Basing on the Chinese dataset, Liu et al. (2019) [12] utilized the Support Vector Machine (SVM) for the binary classification task of AD and Control and achieved a detection accuracy rate of 81.9%. At present, there exist researches focusing on the classification and prediction of AD and Control, while there is a lack of study on the early screening of patients with cognitive impairment including MCI.

(2) Regarding the analysis of the correlation between speech features and brain images, many studies have revealed the basis for neuroimaging behind the phenomenon of speech disorders in the elderly. Feng et al. (2018) [13] showed that the gray matter volume in the middle frontal gyrus, superior parietal lobe, inferior temporal gyrus, and fusiform gyrus of patients in the MCI group was significantly different from that in the Control group. There is a positive correlation between gray matter volume and the patient's language score. Joseph et al. (2017) [14] revealed that arcuate tract damage can produce long-term language impairment. Andrea et al. (2021) [15] found the prior relationship between the persistent speech disorder and Broca's area damage may be explained by the simultaneous damage of the white matter near the front of the arcuate tract above the insula. Li et al. (2017) [16] compared the AD and Control groups, and showed that the AD patients exhibit more atrophy in the gray matter volume of the hippocampus, parahippocampal gyrus, insula, and superior temporal gyrus, and their decrease in gray matter volume, loss of memory, attention disorder and language disorder are related. Although the above-mentioned research revealed the relationship between speech disorders and brain structure in the elderly, they focused on the group level and few studies have been conducted on the quantified speech features to associate them with brain imaging features. Thus, the extracted speech features usually lack explicit clinical interpretations.

(3) On the establishment of speech features' reference interval for cognitive impairment screening, there is no scholars have given the reference interval of acoustic and linguistic features for cognitive impairment screening at present. The cognitive state of the subjects can't be preliminarily judged based on the specific speech features, which reduced the convenience of early screening of cognitive impairment based on speech analysis.

In response to the above-mentioned problems, this article researches early screening of cognitive impairment based on intelligent speech analysis, providing a new technical path for the evaluation of cognitive impairment for the elderly in China, and promoting early large-scale screening. The contributions of this paper are mainly divided into the following points:

(1) Establishing a Chinese speech corpus for the elderly including Control, MCI, and AD subjects.

(2) Constructing the prediction models which are focusing on the early screening of MCI patients based on the feature importance list calculated by an auxiliary diagnosis task for discriminating AD and Control groups.

(3) Analyzing the correlation between speech features and brain atrophy features, and extract key speech features including duration, acoustic and lexical features which have positive correlation between the degree of atrophy in most of brain regions.

(4) Calculating key speech features' reference interval based on the data of Control group.

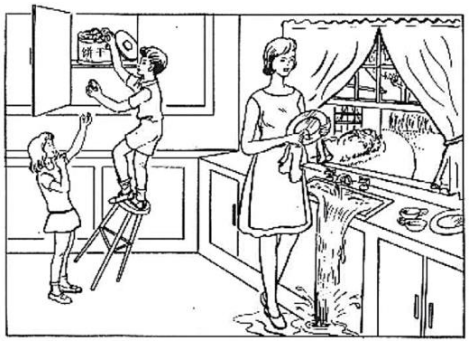

Figure 1. "Cookie Theft" Picture Description Task

II. METHODS

A. Research Participants

We collected speech recordings for this study based on the cookie theft picture description task (Figure 1) which was taken from Boston Diagnostic Aphasia Examination-Third Edition (BDAE-3) [17]. The dataset contains 515 subjects over 55 years old from the Department of Neurology, Shanghai Tongji Hospital. Transcripts, roles, and silence segments were annotated manually using the Praat coding system [18]. All subjects have signed an informed consent form, and this study has been approved by the Ethics Committee of Shanghai Tongji Hospital Affiliated with Tongji University. It is believed that this study complies with the Helsinki Principles and is in line with medical ethics.

The cookie theft picture description task requires each subject to say as much as possible about the content in the picture, and it is allowed that the subjects are encouraged by the examiner to restate when they encounter difficulties. Participants in the dataset included 142 AD patients, 201 MCI patients, and 172 Control (cognitive normal) subjects (Table I).

TABLE I. THE STATISTICS OF SUBJECTS IN OUR DATASET.

Group (number)	Gender (Male/Female)	Age mean(std)	Education mean(std)
Control(172)	84/88	68.98(7.24)	11.58(3.14)
MCI(201)	85/116	70.06(7.96)	9.70(4.10)
AD(142)	54/88	77.28(8.28)	7.84(5.02)
All(515)	223/292	71.69(8.56)	9.81(4.35)

B. Features

The features extracted in this paper can be divided into four categories: acoustic features, duration features, lexical features, and brain atrophy features.

1) Acoustic features: The eGeMAPS feature set was extracted by openSMILE toolkit [19], which contains 88-dimensional high-level statistical features calculated based on low-level descriptors, MFCC, Spectral flux, and frequency features [20].

2) Duration features: The duration features were calculated based on the results of manual tagging by Praat

toolkit [21], which contain 23-dimensional features including the length of speak time of the examiners and participants, the statistics of speech of each sentence of examiners, and so on.

3) Lexical features: The lexical features were calculated based on the results of word segmentation and part-of-speech (POS) tagging by HanLP toolkit [22], which contains 82-dimensional features including the number of words, POS rate, content density [23], word frequency [24], Quantitative Production Analysis variables [25], and so on.

4) Brain atrophy features: The Brain atrophy features were extracted based on Magnetic Resonance Imaging (MRI), which contains 58-dimensional features including the atrophy level of the hippocampus, lobus temporalis, lobus parietalis, and other brain regions.

C. Evaluation Metrics and Feature Selection Method

The experiments in our study contain two tasks named "auxiliary diagnosis" and "early screening". The task of auxiliary diagnosis aims to discriminate AD from Control, and the task of early screening is going to distinguish between MCI and Control. The subjects in our dataset were divided into training, development and test sets at a ratio of 7:1:2. And in the process of division, the training, development and test sets maintained similar age and education distributions.

In the aspect of model training and evaluation, the experiments used the XGBoost classifier in the Python package scikit-learn [26] to establish the baseline model and adopted the Accuracy, Precision, Recall and F1-Score as evaluation metrics. During the establishment of the XGBoost classifier, the Gini coefficient was used to calculate the purity of each split point, and the importance of each decision tree was calculated by the improved performance from the split point, and then all the decision trees were averaged to obtain the importance of the features. On this basis, the Mann-Whitney-U (MWU) test method [27] was employed to analyze the significant difference of the features, and the features with no significant difference were deleted. Based on the feature importance and significance difference analysis results, an appropriate threshold was selected according to the development set, and the feature set that achieved the highest accuracy was selected. This study used different random seeds on the training set to perform the 10-folds cross-validation 20 times. The mean and standard deviation of each evaluation among the 10-folds cross-validations were recorded and used for hyperparameter selection. Then the feature importances were ranked and the best feature subset was selected using the development set, and the evaluation results were reported on the test set.

D. Spearman Correlation Analysis

Based on the feature selection, this experiment used Spearman's correlation analysis method to analyze the correlation between the key speech features and the brain atrophy features. After the correlation analysis, the feature combinations with significant levels less than 0.05 were extracted to study the clinical interpretability of the speech features for detecting cognitive impairment.

E. Reference Interval of Key Speech Features

Targeting at the determined key speech features, their means and standard deviations were calculated by analyzing their distributions of the Control group. According to the percentile method, the reference interval of these features was established.

III. Experimental Results

A. Cognitive Impairment Screening

Briefly, we use datasets includes Control Group and AD patients to sort the importance of features to get an ordered feature set of high importance to low importance, with Table II exhibiting the metrics by the model trained with full features. Then the top N features which are selected by MWU test form a subset of key features for the early screening of AD, which differentiates the MCI patients from Control Group and the results are exhibited in Table III.

Firstly, for the large-scale application of early screening of cognitive impairment, we did experiment on the auxiliary diagnosis of AD. Specifically, we trained an XGBoost classifier to distinguish AD patients from the Control group with all features (Full Features), and ranked the features by their importance weight. Then, MWU test on training set was utilized to remove features of which the means of AD and Control groups were not significantly different from each other to acquire the feature subset. The evaluation results on the test set of the auxiliary diagnosis task are reported in Table II.

TABLE II. Evaluation Results on the Test Set of the Diagnosis Task.

Feature Set	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Full Features	82.26	75.00	84.00	79.25

Then, to investigate whether feature subset formed by auxiliary diagnosis of AD is effective for the early screening task, we built another XGBoost classifier to discriminate MCI from the Control group on all features as baseline. Similarly, an early screening feature subset was obtained by feature importance weight and MWU test using the training and development data of MCI and Control groups.

From Table III, we can see that the classification accuracy was increased from 62.16% to 63.51% by using the first 19 features in early screening feature subset, while it was increased to 67.57% by using the first 22 features in the feature subset of auxiliary diagnosis task.

TABLE III. Evaluation Metrics on Test Set of Screening Task using different feature sets

Feature Set	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Full Features	62.16	72.50	63.04	67.44
Early Screening Feature Subset	63.51	75.00	63.83	68.97
Auxiliary Diagnosis Feature Subset	**67.57**	**72.50**	**69.05**	**70.73**

It can be seen that the model established using the features generated by the feature subset of the auxiliary diagnosis task can improve the early screening effectiveness, which achieves 10% relative improvement in F1-score to discriminate MCI patients from healthy controls. The feature subset which achieves 70.73% F1-Score contains 22 dimensional features which are listed in Table IV.

TABLE IV. DESCRIPTIONS OF THE FEATURE SUBSET FOR THE SCREENING TASK DETERMINED USING AN AUXIALIARY DIAGNOSIS TASK.

Feature Name	Type	Import-ance	Description
A_speak_duration_median	duration	1.000	The median of all sentences of the examiner
A_nsyll_sum	duration	0.755	Number of syllables of the examiner
A_speak_duration_mean	duration	0.496	The average length of all sentences of the examiner
F0_sma3nz_per50	acoustic	0.485	50% quantile of pitch
F0_sma3nz_amean	acoustic	0.469	Mean of pitch
n_unique_words	lexical	0.405	Number of different words
F0_sma3nz_per80	acoustic	0.383	80% quantile of pitch
age	-	0.379	The age of the subject
n_characters	lexical	0.378	Number of Chinese characters
n_n	lexical	0.346	Number of nouns
HNRdBACF_sma3nz_amean	acoustic	0.341	Mean of harmonics-to-noise ratio
A_sr_mean	duration	0.298	The average of speech rate of the examiner
O1	lexical	0.272	Word frequency of words containing the meaning of 'cookies' in Chinese
curtain	lexical	0.257	Word frequency of 'curtain' in Chinese
slopeUV0-500_sma3nz_amean	acoustic	0.226	Mean of Spectral Slope 0-500 Hz
O12	lexical	0.225	Word frequency of words containing the meaning of 'curtain' in Chinese
B_speak_duration_proportion	duration	0.218	Percentage of the total time that the subject speaks in the entire test
ratio_nv	lexical	0.205	Ratio of nouns to verbs
alphaRatioV_sma3nz_stddevNorm	acoustic	0.204	Standard deviation of Alpha Ratio
B_speak_duration_mean	duration	0.200	The average length of all sentences of the subject
ratio_n	lexical	0.191	Ratio of nouns to total words
O7	lexical	0.190	Word frequency of words containing the meaning of 'water in Chinese

B. *Correlation Analysis*

After extracting the key speech features for MCI/Control classification, the correlation coefficients between key speech features and brain atrophy features were calculated to analyze the relationship between speech features and the etiology of cognitive impairment. The speech features used in correlation analysis were the ones listed in Table IV, including acoustic features, duration features, and lexical features. The brain atrophy features used including scores of brain atrophy in different brain regions were evaluated by the Medial Temporal Lobe Atrophy Scale (MTA) [20] and Global Cerebral Atrophy Scale (GCA) [28]. The MTA mainly used MRI to evaluate the height of the hippocampus, the width of the choroidal fissure, and the size of the temporal horn. The scores of MTA can be divided into 4 levels (0-3), in which 0 level indicates no atrophy while 3 level indicates that the medial temporal lobe is significantly atrophic and the hippocampal volume is severely reduced. The GCA mainly assesses the extent of atrophy in the cerebral cortex and records the location. The scores of GCA also can be divided into 4 levels (0-3), in which 0 level indicates no atrophy while 3 level indicates severe atrophy in the cerebral cortex.

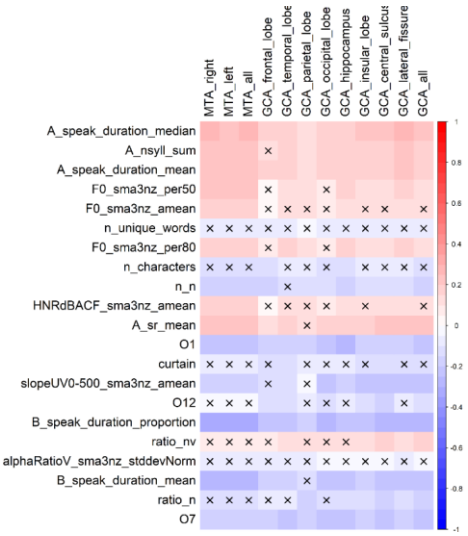

Figure 2. Results of the Correlation Analysis between Speech Features and Brain Atrophy Features.

The results of correlation analysis are shown in Figure 2, where the vertical axis represents speech features which are sorted by feature importance, and the horizontal axis represents brain atrophy features. Darker red or blue indicates higher positive or negative correlation. The cross indicates that the correlation is not significant. It can be seen from the results that the speech duration of the examiner is positively correlated with the atrophy of each brain region, while the speech duration of the subject is negatively correlated. In terms of acoustic features, two frequency and spectral parameter features, "F0_sma3nz_per50" and "F0_sma3nz_per80" are positively correlated with most brain atrophy features, while the spectral feature "slopeUV0-500_sma3nz_amean" is negatively correlated with most brain atrophy features. In terms of lexical features, "A_nsyll_sum" and "A_sr_mean" are positively correlated with with most brain atrophy features, while 'n_n', 'O1' and 'O7' are negatively correlated with most brain atrophy features.

C. *Reference Interval of Speech Features*

For the convenience of cognitive impairment screening based on intelligent speech analysis, we established the reference interval of key speech features based on the percentile method due to the non-normality of the data [29]. These features were selected as the ones that were listed in

Table IV and also had significant correlations with three main brain atrophy features "MTA_all", "GCA_all" and "GCA_hippocampus". The statistics of each feature are shown in Table V, where 2.5 and 97.5 percentiles represent the reference interval of each feature calculated in the Control group respectively.

TABLE V. REFERENCE RANGE OF SPEECH FEATURES.

Feature	Mean	Standard Deviation	2.5th Percen -tile	97.5th Percen -tile
A_speak_duration_media n	0.84	0.36	0.48	1.75
A_nsyll_sum	136.63	67.56	41.30	322.20
A_speak_duration_mean	1.38	0.43	0.88	2.29
F0_sma3nz_per50	31.78	3.97	23.15	35.92
F0_sma3nz_per80	37.15	3.51	34.25	39.69
n_n	24.81	10.00	10.00	49.00
A_sr_mean	3.30	0.61	2.07	4.33
O1	1.08	0.42	0.00	2.00
slopeUV0-500_sma3nz_amean	0.01	0.04	-0.05	0.06
B_speak_duration_propo rtion	0.43	0.12	0.24	0.65
B_speak_duration_mean	1.79	0.73	1.05	3.82
O7	1.08	0.71	0.00	2.00

IV. CONCLUSION

To investigate the methods of MCI screening based on speech analysis, this paper first built up a speech dataset containing the audio recordings from AD, MCI patients and healthy controls. Then, this paper have proposed a method to select the speech features for MCI detection with the help of an auxiliary AD diagnosis task. Based on the selected speech features, we analyzed their correlations with brain atrophy features, and established the reference interval of the key speech features for convenient cognitive impairment screening.

REFERENCES

[1] Jia L, Du Y, Chu L, 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): e661-e671.

[2] Dubois B, Padovani A, Scheltens P, et al. Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges[J]. Journal of Alzheimer's disease, 2016, 49(3): 617-631.

[3] Wang T, Xiao S, Chen K, et al. Prevalence, incidence, risk and protective factors of amnestic mild cognitive impairment in the elderly in Shanghai[J]. Current Alzheimer Research, 2017, 14(4): 460-466.

[4] Minoshima S, Mosci K, Cross D, et al. Brain [F-18] FDG PET for Clinical Dementia Workup: Differential Diagnosis of Alzheimer's Disease and Other Types of Dementing Disorders[C]//Seminars in Nuclear Medicine. WB Saunders, 2021.

[5] Eyigoz E, Mathur S, Santamaria M, et al. Linguistic markers predict onset of Alzheimer's disease[J]. EClinicalMedicine, 2020, 28: 100583.

[6] Hernández-Domínguez L, García-Cano E, Ratté S, et al. Detection of Alzheimer's disease based on automatic analysis of common objects descriptions[C]//Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning. 2016: 10-15.

[7] Fraser K C, Meltzer J A, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech[J]. Journal of Alzheimer's Disease, 2016, 49(2): 407-422.

[8] Almubark I, Chang L C, Shattuck K F, et al. A 5-min Cognitive Task With Deep Learning Accurately Detects Early Alzheimer's Disease[J]. Frontiers in aging neuroscience, 2020, 12.

[9] Koo J, Lee J H, Pyo J, et al. Exploiting Multi-Modal Features From Pre-trained Networks for Alzheimer's Dementia Recognition[J]. arXiv preprint arXiv:2009.04070, 2020.

[10] Sarawgi U, Zulfikar W, Soliman N, et al. Multimodal Inductive Transfer Learning for Detection of Alzheimer's Dementia and its Severity[J]. arXiv preprint arXiv:2009.00700, 2020.

[11] Yuan J, Bian Y, Cai X, et al. Disfluencies and Fine-Tuning Pre-trained Language Models for Detection of Alzheimer's Disease[J]. Proc. Interspeech 2020, 2020: 2162-2166.

[12] Liu Z, Guo Z, Ling Z, et al. Dementia Detection by Analyzing Spontaneous Mandarin Speech[C]//2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2019: 289-296.

[13] Feng W, Wang D, Tang L, et al. Effects of different cognitive Trainings on amnestic mild cognitive impairment in the elderly: a one-year longitudinal functional magnetic resonance imaging (MRI) study[J]. Medical science monitor: international medical journal of experimental and clinical research, 2018, 24: 5517.

[14] Griffis J C, Nenert R, Allendorfer J B, et al. Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke[J]. NeuroImage: Clinical, 2017, 14: 552-565.

[15] Gajardo-Vidal A, Lorca-Puls D L, Team P, et al. Damage to Broca's area does not contribute to long-term speech production outcome after stroke[J]. Brain, 2021, 144(3): 817-832.

[16] Li M, Meng Y, Wang M, et al. Cerebral gray matter volume reduction in subcortical vascular mild cognitive impairment patients and subcortical vascular dementia patients, and its relation with cognitive deficits[J]. Brain and behavior, 2017, 7(8): e00745.

[17] Goodglass H, Kaplan E. Barresi. Boston Diagnostic Aphasia Examination. 3rd Lippincott[J]. 2000.

[18] Boersma P. Praat, a system for doing phonetics by computer[J]. Glot. Int., 2001, 5(9): 341-345.

[19] Eyben F, Weninger F, Gross F, et al. Recent developments in opensmile, the munich open-source multimedia feature extractor[C]//Proceedings of the 21st ACM international conference on Multimedia. 2013: 835-838.

[20] Meyer J S, Huang J, Chowdhury M H. MRI confirms mild cognitive impairments prodromal for Alzheimer's, vascular and Parkinson-Lewy body dementias[J]. Journal of the neurological sciences, 2007, 257(1-2): 97-104.

[21] Styler W. Using Praat for linguistic research[J]. University of Colorado at Boulder Phonetics Lab, 2013.

[22] Zhang X, Wu P, Cai J, et al. A Contrastive Study of Chinese Text Segmentation Tools in Marketing Notification Texts[C]//Journal of Physics: Conference Series. IOP Publishing, 2019, 1302(2): 022010.

[23] Beltrami D, Gagliardi G, Rossini Favretti R, et al. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline?[J]. Frontiers in aging neuroscience, 2018, 10: 369.

[24] Kavé G, Dassa A. Severity of Alzheimer's disease and language features in picture descriptions[J]. Aphasiology, 2018, 32(1): 27-40.

[25] Jarrold W, Rofes A, Wilson S, et al. A" Verbal Thermometer" for Assessing Neurodegenerative Disease: Automated Measurement of Pronoun and Verb Ratio from Speech[C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020: 5831-5837.

[26] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python[J]. the Journal of machine Learning research, 2011, 12: 2825-2830.

[27] McKnight P E, Najab J. Mann - Whitney U Test[J]. The Corsini encyclopedia of psychology, 2010: 1-1.

[28] Koedam E L G E, Lehmann M, Van Der Flier W M, et al. Visual assessment of posterior atrophy development of a MRI rating scale[J]. European radiology, 2011, 21(12): 2618-2625.

[29] Jones G R D, Haeckel R, Loh T P, et al. Indirect methods for reference interval determination–review and recommendations[J]. Clinical Chemistry and Laboratory Medicine (CCLM), 2019, 57(1): 20-29.