Using text and acoustic features to diagnose Mild Cognitive Impairment and Alzheimer’s disease

Liu Ning (✉ 625406059@qq.com)  
Hangzhou normal university  https://orcid.org/0000-0002-9242-506X

Kexue Luo  
Tongde Hospital Of Zhejiang Province

Research article

**Keywords:** Machine learning, Mild cognitive impairment, Acoustic feature, Alzheimer’s disease

**DOI:** https://doi.org/10.21203/rs.3.rs-92702/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Spoken responses can provide diagnostic markers as language impairment maybe an important early performance for dementia patients. In this study, an automatic assessment system was proposed to discriminate MCI and AD patients from their speeches so as to achieve the aim of speeding up treatment and slowing down disease progression.

**Methods:** We integrated a group of acoustic, demographic, linguistic features and used machine learning algorithm to effectively predict MCI and AD patients. Additionally, in order to get the best result, comparison experiment is done effectively which includes three different feature extraction methods (e.g. acoustic, text and their combination) and four of the most popular algorithms, namely, Logistic Regression, SVM, Random Forest and LightGBM.

**Results:** According to Iflytek’s dataset “Alzheimer’s disease prediction challenge competition” in 2019, the performance of LightGBM was especially better than other algorithms, the state-of-the-art AUC value of which was between 0.75 and 0.89 in binary classification and across 0.57 in ternary classification. The result also revealed that age had a significant impact on all the proposed cognitive factors in the meanwhile.

**Conclusions:** The results indicate that our method is increasingly useful for assessing suspected AD and MCI by using multiple, complementary acoustic and linguistic measures.

1. **Background**

The cognitive abilities of human beings decline over time, which lead to a decline in their ability to remember and recall events, and difficulty in finding words properly to express themselves in the conversations. As a part of higher brain function, language function has an effective relationship with cognitive function [1]. The AD and MCI screening method based on linguistics has huge potential because linguistics is closely related to cognitive function. This noninvasive diagnostic method can reduce the burden of health care system if it can efficiently and accurately monitor the progression of AD and MCI disease over time, so it can be used as a reliable tool for detecting different stage of dementia simply and early as it can capture patients’ linguistic performance in a real-time manner. Picture description task [2] had been proven to be an effective method to detect AD and MCI by constructing an automatic evaluation model for speech acquisition, which has an important theoretical and practical significance.

Clinicians found that it was challenging to recognize cognitive impairment in any stage, up to 50% people can not be diagnosed timely even later dementia [3], the most important research in brain aging had focus on the earliest detectable phase as cognitive impairment is irreversible, the stage of which is commonly known as MCI and SCD [4]-[7], it goes undiagnosed and can not be found easily. It was also proved that some people will progress from MCI to dementia while others still stable for many years, but a minor proportion to return normal cognitive condition. Many studies [8]-[16] had applied similar
techniques to identify MCI and AD based on spontaneous speech and linguistic features, which mainly included picture description, story narration, map location identification, instruction execution and so on. In this paper, we used picture description task to achieve our aim. The acoustic and linguistic complexity measures were extracted from speech and transcripts in order to find the early markers of MCI and AD exactly, and then the feature sets were input into the classifier (computer algorithm) in different combination in order to discriminate AD and MCI individuals from healthy controls. At present, machine learning algorithms commonly used for MCI and AD recognition include SVM, random forest, decision tree, boosting algorithm, deep learning and so on. In this article, some common algorithms were adopted to realize our aim after the feature extraction and the results were compared in the metric index to find the best method.

2. Methods

2.1. Diagnostic Aphasia Examination

Picture description task comes from Diagnostic Aphasia Examination [17], the task is mainly to exhibit the picture on a subject and then describe what is happening in the picture, the participant can be hinted by the doctor if he is silent for a few seconds. In order to detect AD and MCI syndrome sufferings automatically, we build a dementia test data sets on mandarin, which includes the audio and text of picture description task over 500 recordings.

2.2. Dataset

We got the original speech samples, which were then transferred into transcripts by iFlytek Automatic Speech Recognition (ASR) platform [18]. Some data sets that were not suitable for the study had already been discarded, for example, poor recording effects, more dialects, test interruption and so on. In the meanwhile years of education less than 5 years and age under forty years old were also deleted in order to coincide with international research. Finally the number of the data sets that suit for our study were 111 CTRL, 144 MCI and 68 AD in all (60 females and 51 males in CTRL group, 84 females and 60 males in MCI group and 38 females and 30 males in AD group, respectively), as shown in Fig.1 and Table 1. The demographic feature and its statistical performance were illustrated in Table 1, from Table 1 we can find that AD group is obviously older than the other two groups, while there is not much difference between CTRL and MCI group, which means that AD cognitive status was aggravated especially with age went on. Years of education in AD group was the least, while the other two groups almost the same among ternary people. the Standard Deviation (SD) value among three groups in age and years of education make no difference. Age and years of education was statistically significant ($\alpha = 0.05$), which were marked with bold in the first column shown in Table 1.

Table 1. Demographic data (i.e. gender, age and education) of the three subject groups
### Table 4

|                          | Subject group | Statistics |
|--------------------------|---------------|------------|
| **Number of data (mean±SD)** |               |            |
| CTRL (n=111)             | 66.68±10.1    | 65.65±10.04| 66.68±10.1 |
| MCI (n=144)              | 65.65±10.04   | 73.66±10.09| 65.65±10.04|
| AD (n=68)                | 73.66±10.09   |            |            |
| **Age** (mean±SD)        | 66.68±10.1    | 65.65±10.04| 73.66±10.09| 16.97 | 0.000*** |
| **Years of education**   | 11.88±3.188   | 10.39±3.198| 9.93±3.185 | 10.71 | 0.000*** |
| (mean±SD)                |               |            |            |            |            |
| Gender                   | 54.05%        | 58.33%     | 55.88%     |
| (% woman)               |               |            |            |            |

***p<0.001, Statistically significant (α = 0.05) results are in bold

### 2.3 Feature extraction method

In order to select proper features to train the classifier, we separately trained acoustic features, demographic features, linguistic features and their different combinations, then selected the optimal feature portfolio by contrast experiments to find the best performance.

There are two sources of information for every participant, the vocal sample and its transcript. In the end 184 acoustic features from the audio file and 7 linguistic features were extracted, adding 3 demographic features, there were a totally of 194 possible feature parameters in all. The composition of characteristic parameters is listed in Table 4.

#### 2.3.1 Acoustic feature

In the experiment process, the questioner's audio was erased before feature extraction while the subject's audio was left in our experiment. The extracted features included two parts, the first one was a 88 extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), which is stated in detail in the study [19]. The second was based on 24 Low-level descriptors (LLD) [19], the parameter groups of which are listed as follows:

**Frequency-related parameters:**

- Pitch
- Jitter
- Formant 1, 2, and 3 frequency
- Formant 1
Energy/Amplitude-related parameters:

- Shimmer
- Loudness
- Harmonics-to-Noise Ratio (HNR)

Spectral parameters:

- Alpha Ratio
- Hammarberg Index
- Spectral Slope 0–500 Hz and 500–1500 Hz
- Formant 1, 2, and 3 relative energy
- Harmonic difference H1–H2
- Harmonic difference H1–A3

Spectral (balance/shape/dynamics) parameters:

- MFCC 1–4 Mel-Frequency Cepstral Coefficients 1–4.
- Spectral flux difference of the spectra of two consecutive frames.

Frequency related parameters:

- Formant 2–3 bandwidth added for completeness of Formant 1–3 parameters.

So the second feature sets were four statistical features of 24 LLD, which included mean, standard deviation, minimal and median, the dimension of this part was 96-dimension (24*4=96). Finally the total dimension of the acoustic parameters was 184 (96+88=184).

2.3.2 Demographic feature

The demographic feature in our system mainly included gender, age and education. Psychiatry and behavioral sciences in Vanderbilt University Medical Center found that two-third AD patients were female in America. Age is also another very important factor. It is universally acknowledged that the older a man is, the more likely he is to suffer from AD. Education is another important factor of MCI. numerous studies have shown that the higher educational level is, the lower AD incidence is. According to statistics, the incidence of AD was 20.4% for people with less than eight years of education and 15% for those with 9 to 11 years of education. The incidence for people having attended high school was 13.2% while that for those with college degree or above was 11.2%. Therefore, we extracted the important but easy-to-get three demographic factors as our feature parameters.

2.3.3 Linguistic feature

The transcripts are saved by TSV format. Table 2 presents transcript data set format.
Table 2. The format description of transcription

| Field name | Description |
|------------|-------------|
| no         | The number of the data set |
| start_time | The start time of the dialogue |
| end_time   | The end time of the dialogue |
| speaker    | The <A> stands for doctor,<B> stands for subject.sil,<DEAF>,<NOISE> all means silence |
| value      | The content of the dialogue |

The completeness of participants’ discourse reflects their cognitive level, the duration of the dialogue is an important index for cognitive status, the start time (start_time) and the end time (end_time) stands for the beginning time and end time of the dialogue, so the result of end_time - start_time is the duration of the dialogue in one sentence between doctor and subject, which is an important indicator for cognitive level, namely, the longer of the duration, the higher of the cognitive level relatively. Therefore, in this section, seven statistical features of the duration in every sentence is extracted as the linguistic feature, which are average (avg), standard deviation (std), minimal, median, maximum, skewness and shape. Table 3 manifests the mean and standard deviation of three groups for linguistic features, as well as statistical values of seven linguistic features. Of all the linguistic features, minimal and shape are significantly different at $\alpha = 0.03$ and $\alpha < 0.0001$ respectively. In the meanwhile we can also get some differences for three different groups. For example, avg, minimal, max, std, skew duration of all the sentences in the dialogue decreases with the cognitive impairment deterioration.

From the above analysis, the final input data is a vector of (323, 194) with all the feature parameters, which is then input into the classifier.

Table 3. Statistical values of linguistic features
### Table 4. The composition of characteristic parameters

| Feature | Speech feature | Transcript feature | Demographic feature |
|---------|----------------|--------------------|--------------------|
| Number of feature | 184 | 7 | 3 |
| ALL feature number | 194 | |

3. Results

3.1 Comparison Results

In order to provide better reference scores, we performed different classify experiments to classify AD, MCI and CTRL and compared the experimental results using Logistic Regression, Adaboost, SVM and LightGBM respectively. Firstly, different classification methods using LightGBM algorithm are shown in Table 5. To distinguish CTRL, MCI, and AD, we conducted five different classification experiments, including four binary classification and one ternary classification. The binary classification was to differentiate CTRL from AD and MCI, AD from CTRL and MCI, AD from CTRL and MCI from CTRL. And the
ternary classification was to distinguish AD, MCI and CTRL. Binary classification methods exhibited better performance than ternary classification. The first row in Table 5 was to differentiate patients who had cognitive problem from CTRL. And LightGBM algorithm achieved high scores. The second row was to separate AD patients from other patients, which also achieved good results. The third and fourth rows were to separate AD and MCI patients from CTRL, and the score in the third row was higher than that in the fourth row. After all, MCI can not be easily identified as its feature is not obvious compared with AD. Undoubtedly, the difference between AD patients and CTRL was larger than MCI patients and CTRL. Therefore, ternary classification was not easy to distinguish, and its score was lower than that of binary classification.

Secondly, the results of the different algorithms used for differentiating AD from CTRL, MCI from CTRL and ternary classification are shown in Tables 6, 7 and 8, respectively. As shown in these three tables, the performance of LightGBM was the best among these algorithms.

Thirdly, the results of LightGBM algorithm with five different feature methods are shown in Table 9, which included demographic feature, acoustic feature, linguistic feature, demographic +acoustic feature, and all of the three features. algorithm exhibited an excellent performance in demographic feature, demographic and acoustic + linguistic features, especially in AUC value.

Fourthly, The ROC can not intuitively reflect the classification effect of a classifier. However, AUC, the area under the ROC curve, directly reflects the classification capability of ROC curve. Only the AUC values of all algorithms were compared in this work, as shown in Table 10. LightGBM achieved good results in all the feature methods in CTRL and AD datasets, which ranged from 0.81 to 0.87. There was no large difference in the score between LR and Adaboost. And the score of SVM was the lowest among these algorithms.

In addition, the bold in all the tables stands for the best score in a column. ‘+’ stands for combination in all the tables, ‘and’ stands for comparison between left and right.

**Table 5. LightGBM for different classify methods**

|                  | Accuracy | Precision | Recall score | F1 value | AUC  |
|------------------|----------|-----------|--------------|----------|------|
| CTRL and AD>MCI | 0.72     | 0.72      | 0.93         | 0.81     | 0.75 |
| AD and CTRL>MCI | 0.79     | 0.5       | 0.94         | 0.33     | 0.82 |
| AD and CTRL     | 0.64     | 0.67      | 0.80         | 0.64     | 0.89 |
| MCI and CTRL    | 0.63     | 0.62      | 0.59         | 0.58     | 0.80 |
| AD and MCI and CTRL | 0.6   | 0.60      | 0.60         | 0.55     | 0.57 |
Table 6. Different algorithm results for AD and CTRL

| Algorithm     | Accuracy | Precision | Recall rate | F1 score | AUC  |
|---------------|----------|-----------|-------------|----------|------|
| SVM           | 0.62     | 0.385     | 0.62        | 0.475    | 0.5  |
| Logistic Regression | **0.737** | **0.747** | **0.737** | **0.737** | 0.725 |
| Adaboost      | 0.721    | 0.723     | 0.721       | 0.72     | 0.703|
| LightGBM      | 0.676    | 0.667     | **0.796**   | 0.637    | **0.884615** |

Table 7. Different algorithm results for MCI and CTRL

| Algorithm     | Accuracy | Precision | Recall rate | F1 score | AUC  |
|---------------|----------|-----------|-------------|----------|------|
| SVM           | 0.62     | 0.385     | **0.62**    | 0.48     | 0.5  |
| Logistic Regression | 0.576   | 0.579     | 0.576       | 0.57     | 0.57 |
| Adaboost      | 0.573    | 0.57      | 0.573       | 0.57     | 0.56 |
| LightGBM      | **0.627**| **0.615** | 0.585       | **0.58** | **0.80** |

Table 8. Different algorithm results for AD,MCI and CTRL

| Algorithm     | Accuracy | Precision | Recall rate | F1 score | AUC  |
|---------------|----------|-----------|-------------|----------|------|
| SVM           | 0.565    | 0.319     | 0.565       | 0.41     | 0.5  |
| Logistic Regression | 0.501  | 0.507     | 0.501       | 0.5      | **0.599** |
| Adaboost      | 0.573    | 0.57      | 0.573       | **0.57** | 0.561|
| LightGBM      | **0.604**| **0.598** | **0.604**   | 0.55     | 0.566|

Table 9. The performance of LightGBM algorithm with different features in AD and CTRL datasets
Table 10. AUC value for all algorithms with different features in AD and CTRL datasets

| Features          | Logistic Regression | SVM | Adaboost | LightGBM |
|-------------------|---------------------|-----|----------|----------|
| demographic       | 0.654               | 0.654 | 0.706    | **0.85** |
| acoustic          | 0.704               | 0.5  | 0.618    | **0.81** |
| demographic+acoustic | 0.707            | 0.5  | 0.68     | **0.87** |
| demographic+acoustic+linguistic | 0.73      | 0.5  | 0.703    | **0.87** |

3.2 Performance of different classification feature in AD and CTRL data sets

In order to provide reference scores, we performed classification experiments by using extracted demographics, acoustics, demographics plus acoustics, and demographic plus acoustic plus linguistic features as feature parameters to differentiate AD from CTRL. And we also compared the most commonly used algorithms such as Logistic Regression (LR), SVM, Adaboost and LightGBM algorithm. Firstly, we performed classification experiments by using only these three demographic attributes. The experimental results are shown in Fig.2. As can be seen from the figure, the AUC value of LR, SVM, Adaboost and LightGBM was 65.4%, 65%, 70.6% and 85.3%, respectively, indicating that the AUC value of
LightGBM was higher than those of other algorithms by about 15%, and there was no significant difference in the AUC value among LR, RF and SVM, whose results were all higher than 50% (The average probability of a random guess in a two class task).

The classification results in acoustic features are shown in Fig.3. Compared with other algorithms, the accuracy, precision, recall rate, F1 score and AUC value of LR algorithm were better, which were 71.5%, 72.2%, 71.5% and 71.7% respectively, while the AUC value of Logistic Regression was 70.4%, lower than that of LightGBM which was 81.1%. In addition, the performance of other algorithms was not largely different. The experimental results indicate that LR algorithm exhibited an excellent performance in acoustic features, and that of LightGBM was not bad, either.

The experiment results in demographic and acoustic features are shown in Fig.4. The performance of SVM was not better than that of other algorithms As shown in Fig.4, and there was no large difference in the performance between LR and RF. LightGBM algorithm had a much higher AUC value of 0.874126 and a worse F1 value of 0.4779. By comparing Fig. 4 and Fig.5, we found that the results were not necessarily valid if there were more features, and the performance in combined features might not be better than that in single feature. The feature possibly would have a negative impact on the results, so more experiments are required to find the optimal combination of features.

According to Fig.5, there was no great difference in the performance among LightGBM, LG and Adaboost, but SVM exhibited a poor performance. The AUC value of LightGBM was 0.87, more than 10% higher than that of other algorithms, so MCI can be used for selecting MCI from CTRL by using demographic, linguistic and acoustic features in the proposed LightGBM algorithm.

3.3. Important feature

In binary classification, LightGBM algorithm with the combination of all the features got the best classify result. We find the best important features with this method, the result of which is shown in Fig.6. Age was the most important parameter, with a relative importance of around 18.87%, which were calculated from the transcription of spontaneous utterances. However, we only obtained the optimal results by combining these three different features. Subsequently, a total of 30 most important feature parameters affecting consistent with the research results at home and abroad. Studies have shown that age is a key factor associated with the onset of AD. The older a man is, the more likely he or she suffers from AD.

As year of education increases, the risk of developing AD decreases. Among all the important parameters, 19 acoustic features were very important, with a relative importance of 67.92%. And the linguistic features, included ‘tsv_feats1’, ‘tsv_feats5’ and ‘tsv_feats6’ with a relative importance of about 13.20%, ranked fourth, ninth and eighteenth respectively, which was a smaller relative importance. Age played an important role in our method, and acoustics was the second key factor in differentiating cognitive results. And we also improved the final experiment by linguistic feature.
4. Conclusions

It is of great value for more accurately detecting the linguistic markers since there is not any good therapy for MCI and AD nowadays. We have demonstrated the probability of early detection for AD and MCI in a simple, convenient and relatively accurate method by extracting acoustic markers from spontaneous speech of the subjects and linguistic features from transcripts automatically. The proposed method of feature extraction still has a large space to improve. Firstly, linguistic features extracted in this paper only included four statistical characteristics of the duration in every sentence, which is rough and not exact relatively. There are some potential markers in the utterances, which maybe important signs of MCI and AD, so a better experimental result may got if textual feature is further explored, next we will further conduct experiments on elderly people’s textual information in order to extract the linguistic features more comprehensively and accurately. Secondly, in addition to three demographic factors that we used in our study, other factors, such as metabolic syndrome, psychological factors, social interaction, sleep habits, hobbies and physical exercise et al. May also influence cognitive status from clinical medicine major. Thirdly, 323 samples in our study are not enough to represent the whole group, and the performance will be better in machine learning classification with more samples.

With the development of AI technology, more and more researcher join in this field, now the accuracy of deep learning in this area has been about 90% [20]-[23] in recent two years, the result of which is better than traditional machine learning algorithm, the automatic feature extraction algorithm, such as CNN (convolutional neural network), RNN(Recurrent Neural Network), Transformer, Bert and so on, can capture more subtle linguistic markers especially for MCI. In the future we hope a better classify performance by combining automatic feature extraction method with deep learning algorithm and manual feature extraction with clinical discourse analysis based on the theoretical framework of systemic functional linguistics. Future more, the interpretability is also of great value as deep learning, though high accuracy, lacks interpretability, so future we will explore an interpretable model which may coincide with the need of people.

Language is a window for detecting cognitive impairment, which can uncover the relationship between sufferer's mental activities and utterance, recognize signs of AD and MCI, detect subtle language patterns that are overlooked previously. Except picture description task, AI system can diagnose cognitive deection from text of personal e-mail or speech in social media, so many different types of text produced by people can all be used to train machine learning algorithm. In the meanwhile, many other neurological diseases, such as tristimania, stroke, aphasia and brain trauma, may also influence the way that language is used, so disease detection early has great prospects as long as the patient's verbal data can be collected effectively.

The vocal and linguistic features in our study are a useful reference in the future research. Linguistics is an important latent marker for some disease, which can decrease detect time and fee then improve doctors’ efficiency. This paper just casts a brick to attract jade hoping that more researches can join us to solve the global issue about MCI and AD.
Abbreviations

AD
Alzheimer’s disease
MCI
Mild cognitive impairment
CTRL
Normal people
ASR
Automatic speech recognition
SD
Standard deviation
LLD
Low-level descriptors
eGeMAPS
Extended geneva minimalistic acoustic parameter set
LR
Logistic Regression
SVM
Support Vector machine
CNN
Convolutional neural network
RNN
Recurrent Neural Network

Declarations

Conflict of interest
The authors have no conflicts of interest to disclose.

Competing interests
The authors declare that they have no competing interests.

Ethics approval and consent to participate
This study obtained ethics approval from Research Ethics Committee of Quanzhou normal university.

Consent for publication
Not applicable.
Funding

This research was supported by Fujian Province Young Teacher Education Research Project (No. JAT170480). School of Health Management in Hangzhou normal university supported the study by providing the Data Science Laboratory and facilities for data management.

Authors’ contributions

Liu developed the idea and designed the research. Luo developed the search strategy and analyzed the data. Liu interpreted the analysis. All authors read and approved the final manuscript.

Availability of data and materials

The original data sets got from “Alzheimer’s disease prediction challenge competition” in 2019 (http://challenge.xfyun.cn/2019/gamedetail?blockId=978). The data sets can be visited the website in Github: https://github.com/lzy1012/Alzheimer-s-disease-datasets.

Acknowledgements

We thank Doctor Luo who provide some medical support for AD and MCI disease. Thank you for the data sets which provide by Alzheimer’s disease prediction challenge competition in 2019.

References

1. 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. 1975;12(3):189–98.
2. Forbes-McKay K, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer’s disease with a picture description task. Int. J. Geriatric Psychiatry 2005 (26), 243–256.2005.
3. Fraser K, Rudzicz F, Graham N, Rochon EA. Automatic speech recognition in the diagnosis of primary progressive aphasia. In: Proceedings of SLPAT. Grenoble, France, pp. 47–54. 2013.
4. Ritchie K, Touchon J. Mild cognitive impairment: Conceptual basis and current nosological status. Lancet. 2000;355:225–8. [PubMed: 10675135].
5. Dong G, Yang L, Li CSR, et al. Dynamic network connectivity predicts subjective cognitive decline: the Sino-Longitudinal Cognitive impairment and dementia study[J]. 2020.
6. Chen G, Liu C, Yang K, et al. Beneficial Effects of Brain Reserve on Cognition in Subjective Cognitive Decline from the SILCODE Study[J]. Journal of Alzheimer’s disease: JAD, 2020:1–8.
7. Casanova R, Saldana S, Lutz MW, et al. Investigating Predictors of Cognitive Decline Using Machine Learning[J]. Journals of Gerontology: Series B, 2020, 75.
8. Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE. Rochon, E. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex. 2014;55:43–60.
9. Fraser KC, Rudzicz F, Rochon, E. Using text and acoustic features to diagnose progressive aphasia and its subtypes. In: Proceedings of Interspeech. Lyon, France, pp. 25–29. 2013.

10. Freedman M, Leach L, Kaplan E, Winocur G, Shulman K, Delis D, 1994. Clock Drawing: A Neuropsychological Analysis. New York: Oxford University Press. Galvin JE, Sadowsky CH. Practical guidelines for the recognition and diagnosis of dementia. J. Am. Board Family Med. 25 (3), 367–382. 2012.

11. Garrard P, Maloney LM, Hodges JR, Patterson K. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. Brain. 2005;128(2):250–60.

12. Garrard P, Rentoumi V, Gesierich B, Miller B, Gorno-Tempini ML.. Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. Cortex. 2014;55:122–9.

13. Kathleen C, Fraser JA, Meltzer, and Frank Rudzicz. 2016. Linguistic features identify Alzheimer’s disease in narrative speech. Journal of Alzheimer's Disease 49, 2 (2016), 407–422.

14. Gosztolya G, Toth L, Grosz T, Vincze V, Hoffmann I, Szatloczki G, Pakaski M, Kalman J. 2016. Detecting Mild Cognitive Impairment from spontaneous speech by correlation-based phonetic feature selection. In: Proceedings of Interspeech. San Francisco, CA, USA, pp. 107–111.

15. Gosztolya G, Vincze V, Toth L, et al. Identifying Mild Cognitive Impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features[J]. Computer Speech & Language, 2019, 53(JAN.):181–197.

16. Fraser KC, Meltzer JA, Rudzicz F, et al. Linguistic Features Identify Alzheimer's Disease in Narrative Speech[J]. Journal of Alzheimer's Disease. 2015;49(2):407–22.

17. Goodglass H, Kaplan E, Barresi B, Boston Diagnostic Aphasia Examination Record Booklet. Lippincott Williams & Wilkins, 2000.

18. Xu Y, Song Y, Long YH, et al. The description of iFlyTek Speech Lab system for NIST2009 Language Recognition Evaluation[C]// International Symposium on Chinese Spoken Language Processing. IEEE, 2010.

19. Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, Devillers´LY, Epps J, Laukka P, Narayanan SS, et al. The Geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. IEEE Trans Affect Comput. 2016;7(2):190–202.

20. Wang N, Chen M, Subbalakshmi KP. Explainable CNN-attention Networks (C-Attention Network) for Automated Detection of Alzheimer's Disease[J]. 2020.

21. Flavio Di Palo and Natalie Parde. 2019. Enriching Neural Models with Targeted Features for Dementia Detection. arXiv preprint arXiv:1906.05483 (2019).

22. Sweta Karlekar T, Niu, Bansal M. 2018. Detecting linguistic characteristics of Alzheimer’s dementia by interpreting neural models. arXiv preprint arXiv:1804.06440 (2018).

23. Oh K, Chung Y-C, Kim KW, Kim W-S, Il-Seok Oh. 2019. Classification and Visualization of Alzheimer’s Disease using Volumetric Convolutional Neural Network and Transfer Learning. Scientific Reports 9, 1 (2019), 1–16.
Figures

Figure 1

Sample/label distribution table
Figure 2

Results in demographic feature
Figure 3

Results in acoustic feature
Figure 4

Results in Demographic+Acoustic feature
Figure 5

Demographic+Acoustic+linguistic feature results
Figure 6

The proportion of feature importance