Measures of Lexical Diversity and Detection of Alzheimer’s Using Speech

Muskan Kothari, Darshil Vipul Shah, Moulya T., Swasthi P. Rao and Jayashree R.
Department of Computer Science and Engineering, PES University, Bangalore, India

Keywords: Alzheimer’s Disease, Speech, Feature Extraction, Brunet’s Measure, Sichel’s Measure, Lexical Diversity, MATTR, MTLD.

Abstract: Alzheimer's disease is the most common cause of dementia — a continuous decline in thinking, behavioral and social skills that affects a person's ability to function independently. Another area of concern is the overlap of symptoms with a similar disease of dementia - Frontotemporal Dementia(FTD). This paper aims to analyze the difference in linguistic features between control and dementia groups with respect to lexical diversity through measures like Brunet’s and Sichel’s measure, frequency rates of adverb, verb, and linguistic deterioration through repetition, disfluency, incomplete sentences, hesitation and long pauses through dataset obtained by DementiaBank. This is achieved through gauging the cognitive ability in speech, which is an inexpensive and non-invasive mode of analysis, qualifying as a screening test. The subjects are given certain description tasks such as the famous cookie theft picture, analyzed through conversations. The result displays the difference in lexical diversity which is a significant marker.

1 INTRODUCTION

Alzheimer’s disease is a progressive neurologic disorder that causes the brain to shrink (atrophy) and brain cells to die. Researchers across the world are constantly making efforts to find methods for the detection and treatment of this disorder in an effective, non-invasive and cost-efficient way. Speech is one of the most effective, inexpensive and non-invasive modes of testing.

AD affects one in ten adults over the age of 65 years in the United States (Alzheimer’s Association, 2015). Diagnosis is possibly more effective in the early stages of dementia. In low and middle income countries, diagnosis of AD frequently occurs several years after the onset of the disease. This leads to a treatment gap for early dementia sufferers (Alzheimer’s Disease International, 2011). This gap reduces the effectiveness of treatments, prolonging the patients’ state of reduced independence.

Sometimes AD might be misclassified into what’s known as Fronto Temporal Dementia, as the symptoms are very common to Alzheimer’s Disease and can jeopardize the appropriate diagnosis and medication for a patient with cognitive impairment since it is now considered to be as common as Alzheimer’s in middle aged patients. AD is often difficult to differentiate with FTD, especially in the early stage. Currently, there are no disease-modifying treatments for FTD. The acetylcholinesterase inhibitors widely used in patients with AD could lead to worsening of symptoms in those with FTD.

Therefore, accurate diagnosis from a differentiating perspective of FTD and AD and the reduction of misdiagnosis is of essential utility in clinical trials. FTD is also a highly heritable group of neurodegenerative disorders, with around 30% of patients having a strong family history. Diagnosing and confirming it early could be very helpful for the descendants of the patient as well.

Realizing the necessity, scope and potential of this area of research, the work described in this paper aims to resolve some core issues related to Alzheimer’s detection among patients, taking the first step in classifying it precisely through linguistic features extracted from transcribed files in CHAT (Codes for the Human Analysis of Transcripts) protocol (MacWhinney, 2000). It cannot be denied that examining linguistic features is one of the best and most inexpensive ways to detect Alzheimer’s, which is why this research will be using them in the unique model, inspired by the approaches we have explored.

Kothari, M., Shah, D., T., M., Rao, S. and R., J.
Measures of Lexical Diversity and Detection of Alzheimer’s Using Speech.
DOI: 10.5220/0011779000003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 806-812
ISBN: 978-989-758-623-1; ISSN: 2184-433X

2 RELATED WORK

The work done in (Renxuan Albert Li, 2020) mainly focuses on Mild Cognitive Impairment (MCI) using the Brain, Stress, Hypertension, and Aging Research Program (B-SHARP) dataset. Three speech tasks were given to the subjects and the recordings of 1-2 mins each were transcribed using Temi (Daniela Beltrami, 2018), a tool that automatically transcribed speech and linguistic features were further analyzed with a helpful tool called ELIT (Jacob Devlin, 2019). Three tasks involved speaking about picture description, room environment and daily activity. Task 2 out of 3 has highest accuracy which proved spatial descriptions to be most useful.

The methodology proposed in (N. Wang, 2020) is highly personalized. It analyzes the hidden linguistic patterns of each subject separately using their own linguistic biomarkers over a duration. The main analysis done over here was a case study on President Reagan's speeches. Uses a lot of speech features such as pronoun-noun ratio, word frequency ratio, Honore's measure, Brunet's measure etc. Focuses on trying to predict in an automated manner rather than on trained data, and uses SVM for this approach, but it's observed that prediction using t-SNE is more accurate than the automated SVM approach.

The aim of the research done in (Haulcy, 2021) was to classify Alzheimer's using ADRess Dataset. The dataset consists of audio recordings along with the transcripts, and metadata for non-AD and AD patients. Feature sets were formed with LDA, and with PCA, and training of classifiers on feature sets to observe the effect of dimensionality reduction. One main advantage of using linguistic features is the usage of punctuation. The semantic and syntactic information is used by the model. The classifiers used are LDA, Decision Tree classifier, the k-nearest neighbors classifier, SVM and RF classifier.

So far, most of the work done was in English and no other language had been worked upon in detail. But in (Zhiqiang Guo, 2020), AD was detected in Mandarin. The dataset used here consists of transcriptions of the cookie theft picture in Mandarin. 208 transcriptions were recorded equally for both healthy and AD patients. The results of this experiment show that the contrastive learning method can achieve better accuracy than conventional CNN-based and BERT-based methods. The output was achieved by a model containing two pooling layers of english and mandarin and two auto-encoders of both the languages. The accuracy obtained here was 81.4%.

In (Chloé Pou-Prom, 2018), the researchers leverage the multiview nature of DementiaBank, to learn an embedding that captures different modes of cognitive impairment. Generalized canonical correlation analysis (GCCA) was applied to the dataset and the benefits of using multiview embeddings on identifying AD and predicting clinical scores were demonstrated. The short-coming of the research being that while GCCA allowed for an arbitrary number of views, it learnt only linear projections to the embedding space. In this case, DGCCA can be used which makes use of neural networks to learn non-linear mappings to the embedding space.

Semantic Verbal Fluency tests were used in (Felipe Paula, 2018) to detect certain clinical conditions like dementia The SVF dataset of a 100 patients was classified into groups of 25 controls each in classes like Amnestic Mild Cognitive Deficit (aMCD), Multi-domain Mild Cognitive Deficit (mMCD) and Alzheimer's Disease (AD). The SVF test uses a binary function called switch which operates on a sequence of N words. Three heuristics of the switch function were explored. These were the Detection based on global mean, detection based on local mean and hybrid detection.

An approach of using CNN and LSTM was seen in (Flavio Di Palo, 2019). The purpose of CNN and LSTM was to enable the learning of both implicitly learned features and targeted features to perform classification. A bi-directional LSTM was used instead, and an attention mechanism was applied on the hidden states of the LSTM. Class weights that were added to the loss function in this approach took the dataset imbalance into account.

Kathleen et al. in (Zhou, 2016) have devised ways to differentiate and identify between having AD and depression. To analyze further, textual and acoustic features were extracted from the patient's speech data. A subset of the extracted features were selected by using a correlation-based filter. A detailed analysis of correlation between depression and dementia was carried out by the authors. The selected features were then fed in ML classifiers like SVM and Logistic Regression (LR) models.

3 DATASET

From the review done in (Haulcy R, 2021), there are various datasets available for the study of Dementia in languages such as English, French, Greek, Hungarian, Italian, Mandarin, Portuguese, Spanish, Swedish and Turkish. While most of them are available upon request, the availability of the rest of the datasets is undefined. In English, there are 3 major datasets widely known to be available, namely DementiaBank, Pitt Corpus and WRAP. All of these datasets are available upon request.

For the purpose of this paper, the dataset chosen was Pitt corpus, available in English under non-protocol data where the media included audio files obtained from DementiaBank. This is an open-source repository of various corpora available on request. In DementiaBank, you have corpora available in 5 languages namely English, German, Spanish, Mandarin and Taiwanese, categorized under protocol data, non-protocol data and PPA non-protocol data. This corpus is maintained by Francois Boller and James Becker as part of a larger protocol administered at the University of Pittsburgh School of Medicine.

The dataset includes audio as well as downloadable transcripts which follow the CHAT protocol. The dataset includes the conversation between two participants playing two roles, one as the investigator (INV) and the other as the participant (PAR) who is the patient. The data includes responses for both control and dementia groups where control groups have elderly individuals and dementia groups include patients with probable and possible Alzheimer's disease. The group also includes a few patients from other dementia diseases. The conversations between the two roles is transcribed for 4 language tasks -

1. Cookie Theft - includes participants describing the cookie theft picture
2. Fluency - includes responses to the word fluency task for the dementia group only.
3. Recall - includes responses to story recall tasks for the dementia group only.
4. Sentence - includes responses to sentence construction task for dementia group only.

The focus for this paper is only for the cookie theft task since it includes both the groups. The reason for choosing the DementiaBank dataset over other available datasets in English is the fact that this dataset is balanced. It also includes other demographic information of the patients such as age, sex, diagnosis and MMSE score.

MMSE stands for Mini-Mental State Examination which is a set of 11 questions that a doctor asks the patient to assess the cognitive impairment. A total of 6 areas of mental abilities are checked through this examination which includes orientation to time and place, concentration, short-term memory recall which can be reasoned for the story recall task, language skills, visuospatial abilities which can be reasoned for the cookie theft task and finally, the ability to follow instructions. The maximum obtainable score for MMSE is 30. A score below 24 is usually indicative of possible cognitive impairment.

A total of 548 files are used for further analysis and research. 305 of the total files are from the dementia group, and 243 files are from the control group. To read the CHAT files in .cha format, pylangacq was used, which is a library to read conversational data represented in this format. It has various methods which allows to obtain information about the participants (in this case, it returns PAR and INV), the metadata stored in transcribed files (which usually start with the @ symbol), number of files, number of words, and number of utterances filtered by participants, through a reader object. It also gives information about tokens in each file which returns an object of tuples with 4 fields. Tokens give you word based annotations, and the fields include the word itself, the part-of-speech tag, morphological information and the grammatical relation. The grammatical relation is an object which tells the relation between two words, including 3 attributes which are the position of the dependent (the word itself), position of the head, and the relation between them.

The metadata transcribed in the files includes information like the encoding (in this case UTF8), language, participants, information about the participants like language, corpus, age, sex, role, group and education.The control files contain a total of 3896 utterances and 33931 words while the dementia files contain a total of 5585 utterances and 43471 words. A subset of the information obtained from one of the.cha files of the dementia group is detailed in table 1. The results of words, utterances, tokens and meta-data along with the method used from pylangacq is displayed.

4 METHODOLOGY

4.1 Data Preprocessing and Preparation

The first step for preparing the data was to analyze the different essential components that constitute the CHAT files. From previous methods explored, the utterances function posed to be very useful, along with the tokens methods. The dataset preparation started with extracting all the utterances by participants in each file. This means that using the utterances method, filtered by 'PAR', each file was

Table 1: Analysis of CHAT files.

Sentence	Information about participant conversations	
	Methods used	*Subhead*
"He's taking cookie jar. that's all."	.words()	["he's", 'taking', 'cookie', 'jar', '.', "that's", 'all', '.']
	.words(by_utterances=True)	[["he's", 'taking', 'cookie', 'jar', '.'], ["that's", 'all', '.']]
	.tokens()	Token(word='taking', pos='part', mor='take-PRESP', gra=Gra(dep=3, head=0, rel='ROOT'))
	.headers()	{'UTF8': '', 'PID': '11312/t-00002422-1', 'Languages': ['eng'], 'Participants': {'PAR': {'name': 'Participant', 'language': 'eng', 'corpus': 'Pitt', 'age': '56;', 'sex': 'male', 'group': 'ProbableAD', 'ses': '', 'role': 'Participant', 'education': '20', 'custom': ''}, 'INV': {'name': 'Investigator', 'language': 'eng', 'corpus': 'Pitt', 'age': '', 'sex': '', 'group': '', 'ses': '', 'role': 'Investigator', 'education': '', 'custom': ''}}, 'Media': '003-0, audio', 'G': 'Cookie'}

processed and the associated label was also prepared for the group that the 'PAR' belonged to. Control group was labeled 0 and the dementia group was labeled 1. The other important feature extraction was using POS tags. Parts of speech tagging have been proving to be essential to extract and learn some of the key features of speech. For patients with Alzheimer's, some of the POS tags are more frequent than normal patients. Using spaCy, an open source library highly suitable for tasks in Natural Language Processing and written in Python and Cython, deemed useful for POS tagging. Each utterance in each file was passed to a function that added the POS tag after the token in each row. Using this library, extraction or preparation tasks become easier because of the attributes that each token is embedded with.

The transcription files also included some of the key transcription symbols to signify the manner of speech or the verbal fluencies. Verbal utterances like repetitions, retractions, pauses of both types - short and long, incomplete words, incomplete sentences, assimilations, various errors, hesitations and disfluencies were captured through transcription symbols, which is elaborated in table 2.

Table 2: Transcription Symbols.

Sl. No.	*Symbol*	*Meaning*
1	[/]	Repetition
2	[//]	Retraction
3	[..]	Pause
4	[.]	Short pause
5	[...]	Long pause
6	[+sgram]	Grammatical error
7	&uh/&um/&mm/&hm	Hesitation
8	&w+	Disfluency

These transcription symbols were replaced with the expansions of what they represented. The concept of regular expressions was used to identify these symbols and each annotation was hereby replaced with the direct meaning.

At the end, we had a dataframe consisting of the label column, all utterances belonging to each file, POS tagged column consisting of the token followed by its POS tag after each, the expanded representation of the annotation in each utterance, and a final column without annotations to prevent skewing of POS tags.

4.2 Ratios and Measures

For the research pertaining to this paper, the linguistic features are divided into POS features and lexical diversity. For POS features, 3 values were computed, which are pronoun-noun ratio, adverb frequency rate and verb frequency rate. These measures are deemed important from the correlation result obtained in (N. Wang, 2020). Alzheimer's patients seemingly use more pronouns than nouns. The utterances of AD patients are also rich in adverbs and verbs compared to other POS tags. The results were consistent with the observations except for a slight variation in verb frequency rate. The P-N ratio obtained for AD patients was 0.6923 and for normal patients was

0.5181, which indicates that normal patients' speech included more nouns resulting in a P-N ratio less than AD patients who used more pronouns than nouns. For adverb frequency rate, the result obtained for AD patients and normal patients was 48.95 and 60.08 respectively. Our implementation computed the frequency by dividing the number of tokens by the number of adverbs. Thus, a higher number of adverbs per number of tokens would result in a lesser adverb frequency rate according to our implementation. This was consistent with the observation that AD patients use more adverbs. For verb frequency rate, using the same implementation as adverb frequency rate, the result obtained for AD patients and normal patients was 16.50 and 12.50 respectively. This implies that less number of verbs were used per number of tokens by AD patients compared to normal patients.

There are 4 measures computed for lexical diversity. From the case study in (Zhou, 2016) on President Reagan's speech, it was established that AD patients have a declined vocabulary richness in their speech. Here's where POS tags come to use once again. It proves that the speech including the vocabulary and the gaps can give a lot to infer. Three popular measures for vocabulary richness are the Honore's statistic (HS), Brunet's index (BM) and Sichel measure (SICH).

It is important to know what hapax legomena and hapax dislegomena mean. Hapax legomena are the word types that occur once in a text while dislegomena are those that occur twice in a text. By logic, hapax legomena is usually the indicator of lexical diversity. Honore's statistic which is usually denoted by R is based on the understanding that texts with rich vocabulary have larger proportions of words that are hapax legomena. But this measure is sensitive to sample size. Both Honore's and Sichel's result in a higher value when vocabulary is rich. In case of Brunet's (W), smaller the value, higher the vocabulary richness and is also not sensitive to the text length. The range of values is usually between 10 and 20. For the purpose of this study, Sichel's and Brunet's measure was chosen, which balances the results for lexical diversity since they are both inversely proportional.

The other two measures used were MTLD (Measure of Textual Lexical Diversity) and MATTR (Moving Average TTR), based on TTR (Type-Token Ratio) which is the number of different words in a sample of text. MTLD tells the average number of consecutive words that maintains a certain TTR before dropping. MATTR is simple enough, in that it calculates the TTR for a window of a certain size.

4.3 Equations and Measures

Brunet's measure was implemented using (1)

$$W = N^{V^{-a}} \quad (1)$$

where -a is a scaling constant, usually equals -0.172. N denotes the length of text and V denotes the number of different words. Sichel's measure is as simple as computing hapax dislegomena on the text.

4.4 Training Models

The training of models started with the preparation of transcribed speeches of AD patients. As explained earlier, the CHAT protocol and its meanings were thoroughly analyzed and POS tagging was applied.

In addition to the POS tag and preprocessed utterances, 4 measures of lexical diversity and 3 ratios of linguistic features were included in the dataset. The gaps in utterances are of equal importance to differentiate a control patient from an AD patient. It is observed that the speech of AD patients shows higher occurrences of repetitions, retractions, disfluency, long pauses, hesitation, grammatical errors and incomplete sentences.

To conclude, all the features mentioned and described thus far have been used to prepare and store the dataset. From the research of existing work, CNNs, SVMs and LSTMs give the best results. For this research, a total of 7 models were trained on the features scaled appropriately. The top three models to give the highest accuracy were MultinomialNB, SVC and Random Forest Classifier.

A comparison of the mean values obtained for each feature in both groups were also compared and the results were consistent with the existing work except for verb frequency ratio which deviated from the existing inferences. All the linguistic features and gaps denoting retraction, repetition, disfluency, hesitation and more showed a higher mean in values for AD patients compared to the control group.

5 RESULTS AND DISCUSSION

The highest accuracy as seen in table 4 obtained was 88.92% by KNearestNeighbors classifier followed by SVC and MultinomialNB. From the preparation and analysis of all the measures, it was clear that the AD patients have a degraded linguistic sense of speech which is seen in poor lexical diversity, higher use of adverbs and pronouns, less use of nouns and we have also identified through verb frequency rate that despite the observation in (Zhou, 2016), verbs are not that frequent in AD patients. The results obtained after computing the mean values for other measures like MATTR, MTLD showed that MATTR and MTLD for AD patients was less than normal patients which is an indication of reduced lexical diversity in the speech of AD patients, and the number of occurrences of repetition, retraction, hesitations, grammatical errors etc, were higher than normal patients.

Reported in table 3 are the pair of values obtained for lexical diversity measures and number of occurrences in the utterances of AD and control group.

Table 3: Comparison of Linguistic Measures.

Measure	**Group**	
	Control	***AD***
MATTR	0.597633	0.566128
MTLD	34.004573	32.048859
Repetition	0.711934	1.780328
Retraction	1.300412	2.101639
Long pause	0.069959	0.098361
Disfluency	0.732510	1.655738
Hesitation	3.419753	3.603279
Grammatical error	1.234568	1.436066
Incomplete sentence	0.172840	0.518033

It can be inferred that the values for repetition, retraction, disfluency and incomplete sentence were significantly higher for AD than control and could pose as a useful measure for training the model and detection purpose.

Finally, the most significant accuracies obtained are tabulated below for the top 3 models. The test size was set to 0.15 and a random state of 61 was applied. Decision tree resulted in the lowest accuracy of 60.24%. The confusion matrix was plotted along with the computation of F1 score, precision and recall for each model of the 7 models.

Table 4: Results from Top 3 Models.

Model	**Results**			
	Accuracy	***Precision***	***Recall***	***F1 score***
KNN	88.92	0.8592	0.8537	0.8563
SVC	84.33	0.8133	0.8128	0.8197
MultinomialNB	84.33	0.8164	0.8216	0.8225

6 CONCLUSION

This research highlights a significant marker in analyzing speech of AD patients. From a medical perspective, using speech is an inexpensive and a non-invasive process which qualifies as screening tests. Capable of quick and reliable results, the inferences from this work include the degradation of lexical diversity in the speech of AD patients, where measures like Brunet's and Sichel's gave differentiable mean values for the two control groups. MATTR and MTLD are another pair of measures where the mean values for AD patients were less than the control group. In terms of utterances and manner of speech, the top 4 significant markers were repetition, retraction, disfluency and incomplete sentences; the mean number of occurrences was ~ 78-201% higher in AD group.

7 FUTURE WORK

This paper talks about the validation of existing inferences with a deviation in verb frequency ratio and also contributes by implementing 4 lexical diversity ratios. There is some potential to include the demographic information from the transcripts and analyze the differences in the onset and changes in cognitive impairments between male and female. To contribute to the work described in this paper in future in order to make it more complete, we want to implement Conditional Random Fields (CRF) to predict the relation between consecutive POS tags and analyze useful inferences obtained, if any. Another addition would be to train models like t-SNE and hybrid CNN-LSTM, like in (Sweta Karlekar, 2018) on the prepared dataset.

ACKNOWLEDGEMENTS

Expressing profound gratitude to Dr. Jayashree R for encouraging and guiding us along the way and the Dept. of Computer Science and Engineering at PES University, for providing this opportunity to expand our potential of impact, for conducting frequent research and inculcating problem-solving disciplines. This opportunity would not be possible without the grant support in the research conducted by the maintainers and researchers of DementiaBank and Pitt Corpus. We are thankful to Carnegie Mellon University, for facilitating resources and granting access.

REFERENCES

MacWhinney B. 2000. *The CHILDES Project: Tools for analyzing talk*, 3rd edition. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Renxuan Albert Li, Ihab Hajjar, Felicia Goldstein, and Jinho D. Choi. 2020. Analysis of Hierarchical Multi-Content Text Classification Model on B-SHARP Dataset for Early Detection of Alzheimer's Disease. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 358–365, Suzhou, China. Association for Computational Linguistics.

Daniela Beltrami, Gloria Gagliardi, Rema Rossini Favretti, Enrico Ghidoni, Fabio Tamburini, and Laura Calzà. 2018. Speech Analysis by Natural Language Processing Techniques: A Possible Tool for Very Early Detection of Cognitive Decline? *Frontiers in Aging Neuroscience*, 10(369):1–13.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

N. Wang, F. Luo, P. Vishal, K. Subbalakshmi, and R. Chandramouli. 2020. "Personalized early stage Alzheimer's disease detection: a case study of president Reagan's speeches." In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing*, pages 133–139, Online. Association for Computational Linguistics.

Haulcy R and Glass J (2021) Classifying Alzheimer's disease using audio and text-based representations of speech. Front. Psychol. 11:624137. doi: 10.3389/fpsyg.2020.624137

Zhiqiang Guo, Zhaoci Liu, Zhenhua Ling, Shijin Wang, Lingjing Jin, and Yunxia Li. 2020. Text Classification by Contrastive Learning and Cross-lingual Data Augmentation for Alzheimer's Disease Detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6161–6171, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Chloé Pou-Prom and Frank Rudzicz. 2018. Learning multiview embeddings for assessing dementia. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2812–2817, Brussels, Belgium. Association for Computational Linguistics.

Felipe Paula, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio. 2018. Similarity Measures for the Detection of Clinical Conditions with Verbal Fluency Tasks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 231–235, New Orleans, Louisiana. Association for Computational Linguistics.

Flavio Di Palo and Natalie Parde. 2019. Enriching Neural Models with Targeted Features for Dementia Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 302–308, Florence, Italy. Association for Computational Linguistics.

Zhou, Luke & Fraser, Kathleen & Rudzicz, Frank. (2016). Speech Recognition in Alzheimer's Disease and in its Assessment. 1948-1952. 10.21437/Interspeech.2016-1228.

Yamada Y, Shinkawa K, Kobayashi M, Nishimura M, Nemoto M, Tsukada E, (2021), "Tablet-based automatic assessment for early detection of Alzheimer's disease using speech responses to daily life questions." Front. Digit. Health 3:653904. doi: 10.3389/fdgth.2021.653904

Chen L, Dodge HH, Asgari M. Topic-Based Measures of Conversation for Detecting Mild Cognitive Impairment. Proc Conf Assoc Comput Linguist Meet. 2020 Jul;2020:63-67. PMID: 33642674; PMCID: PMC7909094.

Sweta Karlekar, Tong Niu, and Mohit Bansal. 2018. Detecting Linguistic Characteristics of Alzheimer's Dementia by Interpreting Neural Models. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 701–707, New Orleans, Louisiana. Association for Computational Linguistics.