CANARY SPEECH

WHITEPAPER

Alzheimer's Model Performance

AUTHORS

Kevin Yang, MS Canary Speech, LLC

Henry O'Connell, CEO Canary Speech, LLC

`

A. Abstract

Background

An estimated 46.7 million people aged 65 and older have Alzheimer's or a related form of dementia. Because there is no cure, early detection and prediction of Alzheimer's is viewed as the best strategy for initiating early treatment and managing the disease.

Objectives

Demonstrate the potential of using both unstructured telephone conversations and analytics to detect voice features predictive of Alzheimer's.

Methods

Authors linked recordings of calls to a health insurer help line to an administrative claims database showing which callers had diagnostic histories of Alzheimer's disease and which did not. Speech analysis software extracted voice features from the calls and predicted using Multi-layered Perceptron (MLP) which callers had Alzheimer's, and which did not, based on features that were more common to each group.

Results

The model assessed differences in speech features to accurately identify more than 96% of the audio files of callers who had diagnostic histories of Alzheimer's. There was no increase in overall accuracy after male and female populations were separated.

Conclusions

Voice analysis of telephone conversations is an emerging tool to help predict and detect patients with Alzheimer's disease and related forms of dementia.

Keywords

Alzheimer's, Predictive Modeling, Multi-layered Perceptron (MLP), Voice, Claims

`

B. Background and Significance

Earlier detection of Alzheimer's disease and related forms of dementia and cognitive decline has been described as a "holy grail" in medicine because of the rising incidence of the disease and the physical and financial toll it takes.(1-2) An estimated 46.7 million people aged 65 and older have Alzheimer's or a related form of dementia, and the "huge majority" of them aren't diagnosed – even in high-income nation's such as the U.S.(3) Prevalence is expected to double as the global population ages. Since there is no cure, early detection and prediction of Alzheimer's is viewed as the best strategy for initiating early treatment and managing the disease. Early diagnosis increases the potential for medication therapy to slow the rate of cognitive decline in patients(4) and give them opportunities to obtain support services that could extend their independence and delay their moves to nursing homes(5). Research has targeted genetics(6), blood biomarkers(7), and diagnostic imaging(8) as ways to supplement existing clinical questionnaires such as the Mini Mental State Examination(9) for detecting Alzheimer's. Recently, studies have also focused on the potential of voice and linguistic patterns and changes as well. Khodabakhsh et al. found that analysis of prosodic elements of speech "such as stress, intonation, and emotion" were effective at identifying people in casual unstructured conversations who had Alzheimer's.(10) Fraser et al. reported 81 percent accuracy in distinguishing people with Alzheimer's by analyzing their linguistic descriptions of pictures and cognitive decline(11). Research has shown that Alzheimer's has an impact early on portions of the brain cortex responsible for speech and word phrasing,(12) making changes in speech a potential target for detection of the disorder before severe neurological symptoms emerge. Studies also have examined the potential for analysis of telephone conversations(13) or remote questionnaires(14) as lower-cost ways of assessing cognitive decline compared to in-person clinical visits and interviews. This study presents the potential of using both telephone conversations and analytics to detect voice features predictive of Alzheimer's. Validating such a method would be an important step as health care organizations try to get ahead of the emerging wave of Alzheimer's and dementia patients.

C. Objectives

Validate a method for predicting Alzheimer's using voice features, which includes:

- Linking call recordings to administrative claims data to obtain member demographics and Alzheimer's diagnoses for a study population.
- Describing and extracting voice features from Alzheimer's and non-Alzheimer's members' call recordings.
- Predicting Alzheimer's based on members' voice features.

D. Methods

`

Data Source and Sample

The data source for the study was an archive of calls placed in a six-month period to a help desk for a large U.S. health insurer. Members' calls are archived in a call recording system and call logs are generated for each call. For this study, the researchers unarchived the call logs and linked them to unique members in an administrative claims database processed between January 1, 2013 and July 31, 2016. The archive contained 20,949,183 call logs of conversations between January 1, 2016 and July 31, 2016. The logs consisted of calls from 6,698,602 unique members. Among those members, 1,078,318 were linked to the administrative claims, had complete claims data, and had at least 12 months of continuous health insurance enrollment. Of those 1,078,318 members, 12,219 had ICD-9 and ICD10 codes in their claims data indicating that they had been diagnosed with Alzheimer's and 1,066,099 did not.
A study population was drawn from this dataset, consisting of 651 Alzheimer's members and 1,018 non-Alzheimer's members. Call recordings for the study population were retrieved from the archive. Call recordings were selected for audio analysis only if they were placed by people calling for their own care needs and of sufficient quality to be used as the audio files for analysis. (Table I)

Members were placed in the Alzheimer's study cohort if they had at least two Alzheimer's diagnosis codes in their administrative claims data in the same 12-month period. (Table II) The index date was defined as the first date of diagnosis. In addition, members must have had at least six months of pre-index health insurance continuous enrollment (baseline period) and six months of post-index enrollment (follow-up period). (Figure 1)

Descriptive Analysis - Study Population

The control population was identified by matching each Alzheimer's member randomly to a non-Alzheimer's members with the same gender and age group. This was done to create two roughly equivalent sample groups for audio analysis. Audio files of 1016 calls were analyzed from the Alzheimer's group (consisting of 651 members) and 1018 (consisting of 1018 members) were analyzed from the non-Alzheimer's group. There was no overlap of member across folds. (Table III & Table IV)

Voice Feature Extraction and Selection

Speech features were extracted from the members' audio files. Examples of feature groups include::

1. *Spectral features*: Mel-frequency cepstral coefficients, perceptual linear prediction
2. *Prosodic features:* Speaking rate, pitch, intensity
3. *Voice quality features*: Shimmer, jitter local, jitter DDP, harmonics-to-noise ratio
4. *Language features*: 1-gram word probability, pause filler ratio
5. *Others:* part-of-voicing feature

`

There were 760 features extracted per audio sample on 2036 voice samples. The number of features was substantially higher than the number of data samples. Feature filtering was performed to reduce overfitting. Features were included if the absolute value of the Pearson's correlation coefficient between the feature values and their class labels (0 for non-Alzheimer's or 1 for Alzheimer's) was greater than a threshold of .1. Features were otherwise excluded.

Predictive Modeling

Multi-layered Perceptron (MLP) with leave-one-fold-out cross-validation was used to compute a disease score. The disease score is the output of the sigmoid function on the top layer of the MLP model. MLP was chosen to map highly non-linear relationships from low-level speech features to Alzheimer's members. Five-fold cross-validation was used to analyze as many test outputs as possible. In addition, age and gender were balanced across folds for both the Alzheimer's and non-Alzheimer's groups to minimize their influence on the accuracy of the model. (Table V)

The output of the MLP model represents the disease score. The disease score has a possible range of [0,1]. A threshold of 0.5 was used for the cut-off for the binary classification. Calls were classified to have Alzheimer's if the disease score was greater than 0.5; otherwise they were classified as non-Alzheimer's.

In each iteration, k-fold analysis involved withholding one fold to act as the test set and using the other four as the training set. Cross validation (leave-one-fold-out) within the training set was used for tuning the MLP hyper-parameters and for choosing the optimal MLP structure. Dropout, early stopping, batch normalization and late fusion were used to reduce overfitting. After the model was tuned, classification accuracy was measured on both the training set and the test set. Recall, precision, and F1 score were computed on the test set.

Finally, results from the five separate analyses were accumulated to create composite test and training results.

E. Results

In both the training and test sets, the model assessed differences in speech features to accurately identify more than 96 percent of the audio files who had Alzheimer's, and those who did not. In addition, the differences in accuracy (.81%) between the training and test sets suggest that overfitting is not significant in the model training. As a measure of precision, meaning how often it was correct when it predicted Alzheimer's among the 2,034 audio files in the study and control groups, the result in the test set was 95.17 percent. The result in the test set was stronger for recall, meaning the share of the 1,016 audio files in the target group that correctly identified as having Alzheimer's, was 97.14 percent. (Table VI)

The resulting confusion matrix of the final set of all test folds showed more false positives than false negatives with a posterior threshold of 0.5. (Table VII)

The cross-validation experiment was repeated after the male and female populations were separated. This experiment did not affect accuracy. (Table VIII-XI)

`

F. Discussion

Altered speech and linguistics has been a known feature of Alzheimer's for many years. A 1995 translation of Aloysuis Alzheimer's 1907 clinical summary of one of his first patients noted that she "uses gap-fills and a few paraphrased expressions ("milk-pourer'' instead of cup)."(15) Today, speech analysis software and analytical methods make it possible to not only identify these differences in speech, but to assess how common those differences are to Alzheimer's patients. It appears that such tools are necessary for this purpose. Konig et al found in a French-speaking study that significant differences in speech "were not perceptible to the ear of a clinician most of the time," but that an automated analysis of those differences proved 87 percent accurate at separating out Alzheimer's patients from a healthy elderly control group.(16)

The Konig study similarly demonstrated that analysis of recorded speech could be successful, however, that study analyzed people carrying out controlled speech activities such as counting backward. This study demonstrated the possibility of identifying patients outside of controlled experiments – through analysis of their voice patterns in everyday phone conversations such as calls to a health insurance help line. The substantial number of false negative results may limit the current approach as a screening and diagnostic tool for clinicians. But follow-up studies could address some of the limitations of this approach and increase the precision of the methodology.

The results present some interesting questions for future research, including the more predictive results in the audio analysis of women with and without Alzheimer's compared to men. Differences between men and women in the symptoms and neurological impacts of Alzheimer's have been documented,(17) along with the impact gender can have on the course of the disease.(18) So gender-specific differences in speech pattern might not be surprising, but they have not been studied extensively.

G. Limitations

While the data source for this study was novel in that in linked claims data to audio files, there were limitations. Audio files were compressed and decompressed as part of the process for this study. This may have affected the quality of the audio in a way that distorted the results. The relatively small study population available for research also might have limited the ability of the analysis to detect certain vocal patterns that are common to Alzheimer's patients. The selection process also could have affected the results by excluding certain population groups. It was necessary for audio analysis, for example, to only include people in the study whose help center calls were clearly audible. It is possible that being inaudible could be a distinctive feature in predicting cognitive decline.

`

The analysis did not separate out predictive results by diagnostic codes, but future research could assess whether linguistic software is more likely to differentiate patients with late-onset versus early-onset Alzheimer's disease. Given the value in earlier identification and diagnosis of Alzheimer's and cognitive decline, a method that identities the early-onset form of the disease would seem particularly important.

The method in this analysis also did not allow for a determination of the features of speech and linguistics that were most distinct between men and women – and allowed for a better prediction rate of Alzheimer's in women. Future research could parse out those features to focus on ones that are most meaningful.

H. Conclusion

This study contributes to a growing body of research on methods to analyze call recordings or telephone conversations, to extract voice features, and to identify patients with Alzheimer's or other forms of dementia and cognitive decline by their voice and linguistic patterns and changes. A follow-up study with a larger sample size would allow for deep learning model development and the ability to conduct subset analyses by gender or disease severity to continue progress toward the use of speech for early identification of Alzheimer's disease.

I. Clinical Relevance Statement

Alzheimer's disease and related forms of dementia are a growing problem in America's aging society. Early detection could help people with the disease pursue treatment that can delay symptoms and prolong independence. Voice analysis could assist with this diagnostic goal if it is reliably able to identify patients with Alzheimer's.

J. Multiple Choice Questions

What voice feature was noted by Aloysuis Alzheimer in his 1907 clinical summary of one of his first patients with characteristics of Alzheimer's disease?

a) Jumbled phrases.
b) Gap fill noises.
c) Uncontrolled cursing.
d) Voice trail off.

This study used different statistical measurements to assess the ability of voice analysis to identify people with Alzheimer's disease or related forms of dementia. What did the recall measurement assess?

a) Share of patients already known to have Alzheimer's to be correctly identified by voice analysis.

`

b) Share of patients out of study and control group to be predicted to have Alzheimer's.
c) Share of patients identified by voice analysis to have lapses in memory.
d) Share of patients incorrectly predicted by voice analysis to have Alzheimer's.

K. Conflicts of Interest

The authors declare that they have no conflicts of interest in the research.

L. Protection of Human and Animal Subjects

Human and/or animal subjects were not included in the project.

M. References

1. The search for the holy grail of Alzheimer's Disease [Internet]. Spectator Health: The Spectator; 2015 [updated 07/29/2015; cited 12/13/2017]. Available from: https://health.spectator.co.uk/the-search-for-the-holy-grail-of-alzheimers-disease/.

2. The 'holy grail' test that gives early warning of Alzheimer's [Internet]. The Daily Mail: Associated Newspapers; 2010 [updated 07/06/2010; cited 12/13/2017]. Available from: http://www.dailymail.co.uk/health/article-1292386/The-holy-grail-test-gives-early-warning-Alzheimers.html#ixzz51AGYRQeY.

3. World Alzheimer Report 2016 [Internet]: Alzheimer's Disease International; 2016 [updated 09/30/2016; cited 12/13/2017]. Available from: https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf.

4. Allan CL, Behrman S, Ebmeier KP. Early diagnosis beneficial in Alzheimer's disease. Practitioner. 2013 Jan;257(1757):15-8, 2.

5. Public Health Alzheimer's Resource Center [Internet]: Alzheimer's Association; 2017 [cited 12/13/2017]. Available from: https://www.alz.org/publichealth/early-detection.asp.

6. Your Guide to Understanding Genetic Conditions: Alzheimer disease [Internet]. Genetics Home Reference: U.S. Department of Health and Human Services; 2013 [updated May 2013; cited 12/13/2017]. Available from: https://ghr.nlm.nih.gov/condition/alzheimer-disease#genes.

7. Keshavan A, Heslegrave A, Zetterberg H, Schott JM. Blood Biomarkers for Alzheimer's Disease: Much Promise, Cautious Progress. Mol Diagn Ther. 2017 Feb;21(1):13-22.

8. Dickerson BC, Stoub TR, Shah RC, Sperling RA, Killiany RJ, Albert MS, et al. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. Neurology. 2011 Apr 19;76(16):1395-402.

9. The MMSE Test [Internet]: Alzheimer's Society; 2017 [cited 12/13/2017]. Available from: https://www.alzheimers.org.uk/info/20071/diagnosis/97/the_mmse_test.

`

10. Khodabakhsh A, Yesil F, Guner E, Demiroglu C. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. EURASIP Journal on Audio, Speech, and Music Processing. 2015 Mar 25;9.

11. Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. J Alzheimers Dis. 2016;49(2):407-22.

12. Bucks, RS; Singh, S.; Cuerden, JM.; Wilcock, GK. Analysis of spontaneous, conversational speech in dementia of Alzheimer type evaluation of an objective technique for analysing lexical performance. Aphasiology. 2000;14(1):71-91.

13. Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. Cognitive and Behavioral Neurology. 1988;1.

14. Yu B, Quatieri T, Williamson J, Mundt J. Cognitive impairment prediction in the elderly based on vocal biomarkers. Interspeech 2015; Sept. 6-10, 2015; Dresden, Germany. ICSA; 2015.

15. Stelzmann R, Schnitzlein H, Murtagh F. An English Translation of Alzheimer's 1907 Paper, "über eine eigenartige erkankung der hirnrinde". Clin Anat. 1995;8:429-431.

16. Konig A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst). 2015 Mar 29;1(1):112-24.

17. Mazure C, Swendsen J. Sex differences in Alzheimer's disease and other dementias. Lancet Neurol. 2016 Apr;15(5):451-452.

18. Mielke MM, Vemuri P, Rocca WA. Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. Clin Epidemiol. 2014 Jan 8;6:37-48.

`

Table I: Attrition Table

Calls	n
# of call logs (1/1/2016-7/31/2016)	20,949,183
Members	**n**
Unique members in call logs.	6,698,602
# of members that link to claims	1,501,372
Members with complete claims data	1,486,878
Members who have at least 12 months of continuous enrollment	1,078,318
Study Population	**n**
# members with Alzheimer's	12,219
# members without alzheimer's	1,066,099
Refined Study Population (Final)	**n**
# of Alzheimer's verified as calling on their own behalf	651
# of members without Alzheimer's	1,018

Table II: Alzheimer's Diagnosis Codes

Diagnosis	Code Type	Description
331.0	ICD-9	Alzheimer's Disease
G30	ICD-10	Alzheimer's Disease
G30.0	ICD-10	Alzheimer's Disease with Early Onset
G30.1	ICD-10	Alzheimer's Disease with Late Onset
G30.8	ICD-10	Other Alzheimer's Disease
G30.9	ICD-10	Alzheimer's Disease Unspecified

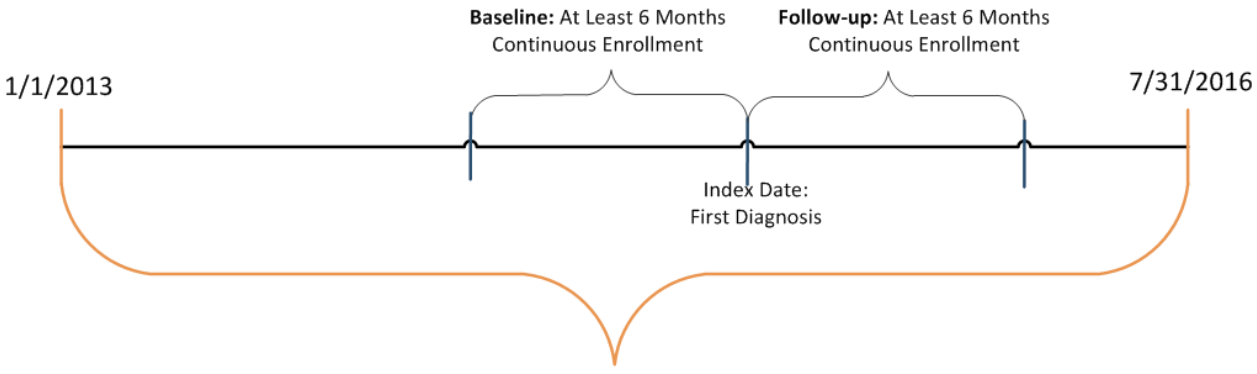

Figure 1: Population Identification Period

Table III: # Audio Files in Target & Control Groups

Age, Gender	Target Group/ Alzheimer's Members (# Audio Files)	Control Group/ Non-Alzheimer's Members (# Audio Files)
50 to 59, Female	10	9
50 to 59, Male	3	3
60 to 64, Female	24	24
60 to 64, Male	21	20
65 to 74, Female	282	281
65 to 74, Male	95	96
75 plus, Female	401	403
75 plus, Male	180	182
Total	1016	1018

Table IV: # Audio Files Across Folds

	Control Data Fold (Non-Alzheimer's)						Target Data Fold (Alzheimer's)					
	0	1	2	3	4	Total	0	1	2	3	4	Total
Age 50 to 59	3	3	2	1	4	13	3	3	2	0	4	12
Age 60 to 64	8	12	8	9	8	45	8	12	7	9	8	44
Age 65 to 74	79	75	76	73	74	377	79	74	75	74	75	377
Age 75+	117	119	115	116	114	581	117	120	117	116	115	585
Total	207	209	201	199	200	1016	207	209	201	199	202	1018
Female	147	145	141	141	143	717	147	144	142	140	144	717
Male	60	64	60	58	57	299	60	65	59	59	58	301
Total	207	209	201	199	200	1016	207	209	201	199	202	1018

`

Table V: # Members Across Folds

	Target Data Fold (Alzheimer's)						Control Data Fold (Non-Alzheimer's)					
	0	1	2	3	4	Total	0	1	2	3	4	Total
Age 50 to 59	3	3	2	1	1	10	3	3	2	0	4	12
Age 60 to 64	8	3	6	6	6	29	8	12	7	9	8	44
Age 65 to 74	39	48	46	39	43	215	79	74	75	74	75	377
Age 75+	78	81	85	79	74	397	117	120	117	116	115	585
Total	128	135	139	125	124	651	207	209	201	199	202	1018
Female	88	90	99	82	88	447	147	144	142	140	144	717
Male	40	45	40	43	36	204	60	65	59	59	58	301
Total	128	135	139	125	124	651	207	209	201	199	202	1018

Table VI: Entire Population Metrics

Metric	%
Accuracy (Train)	96.92
Accuracy (Test)	96.11
Precision	95.17
Recall	97.14
F1-score	96.14

Table VII: Entire Population Confusion Matrix

# Calls		PREDICTED		
		Negative (No Diagnosis)	Positive (Alzheimer's Diagnosis)	Totals
TRUE	Negative (No Diagnosis)	968	50	1018
	Positive (Alzheimer's Diagnosis)	29	987	1016

Table VIII: Male Population Metrics

Metric	%
Accuracy (Train)	93.6
Accuracy (Test)	92.33
Precision	90.67
Recall	94.31
F1-score	92.45

`

Table: IX: Male Population Confusion Matrix

# Calls		PREDICTED		
		Negative (No Diagnosis)	Positive (Alzheimer's Diagnosis)	Totals
TRUE	Negative (No Diagnosis)	272	29	301
	Positive (Alzheimer's Diagnosis)	17	282	299

Table X: Female Population Metrics

Metric	%
Accuracy (Train)	96.56
Accuracy (Test)	95.67
Precision	96.98
Recall	94.28
F1-score	95.61

Table XI: Female Population Confusion Matrix

# Calls		PREDICTED		
		Negative (No Diagnosis)	Positive (Alzheimer's Diagnosis)	Totals
TRUE	Negative (No Diagnosis)	696	21	717
	Positive (Alzheimer's Diagnosis)	41	676	717