INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

An Attention-Based Hybrid Network for Automatic Detection of Alzheimer's Disease from Narrative Speech

Jun Chen, Ji Zhu, Jieping Ye

University of Michigan, Ann Arbor, MI 48109, USA
{junnchen,jizhu,jpye}@umich.edu

Abstract

Alzheimer's disease (AD) is one of the leading causes of death in the world and affects at least 50 million individuals. Currently, there is no cure for the disease. So a convenient and reliable early detection approach before irreversible brain damage and cognitive decline have occurred is of great importance. One prominent sign of AD is language dysfunction. Some aspects of language are affected at the same time or even before the memory problems emerge. Therefore, we propose an automatic speech analysis framework to identify AD subjects from short narrative speech transcript elicited with a picture description task. The proposed network is based on attention mechanism and is composed of a CNN and a GRU module. We obtained state-of-the-art cross-validation accuracy of 97 in distinguishing individuals with AD from elderly normal controls. The performance of our model makes it reasonable to conclude that our approach reveals a considerable part of the language deficits of AD patients and can help with the diagnosis of the disease from spontaneous speech.

Index Terms: Alzheimer's disease, dementia detection, speech analysis, attention mechanism

1. Introduction

Alzheimers disease (AD) is a progressive neurodegenerative disorder featured by loss of memory, dysfunction in language, psychological changes, and impairments in daily activities [1]. It is the most common cause of dementia and the fifth leading cause of death among the elders in the United States [2]. With the expanding of the older population, social and economic burdens caused by AD are ever-increasing.

Currently, the disease is in the absence of a cure or prevention method. However, early diagnosis of AD can offer affected individuals and their family the best chance for legal and financial planning, as well as the access to symptomatic treatment. It also helps the patients get involved in clinical trials and receive potential therapies when available [3].

One major challenge for AD is the lack of handy and reliable approach for early diagnosis [4]. Current diagnostic methods include medical history review, neuropsychological testing (such as Mini-Mental State Examination (MMSE)), standard biological tests (such as blood and cerebrospinal fluid tests), and neuroimaging examinations (such as computed tomography (CT) and magnetic resonance imaging (MRI)) [5]. Although changes in the brain may initiate years even decades before the onset of notable symptoms, most patients will only consider consulting an expert when significant mental and behavioral dysfunctions occur. On one hand, most of the current diagnostic practices have to be taken in clinical facilities, and many of them are time consuming and expensive. On the other hand, symptoms of AD are often confused with memory loss and cognitive decline caused by normal aging[6]. These facts discourage early-stage patients from consulting experts and receiving possible therapeutic intervention at the best time.

Besides the shrinkage of hippocampus that leads to memory loss, damage in cortex areas responsible for language and reasoning is also a prevalent anatomical sign of AD [7]. Therefore, deficits in spontaneous spoken language is potentially an informative hallmark of the disease. Significant correlation was reported between AD and deficits in all levels of language organization, including the phonetic [8], semantic [9][10], syntactic [8], and pragmatic dysfunctions [11][10].

In this paper, we aim to develop a reliable and easy-to-use screening method using narrative speech samples from a picture description task.

2. Related Works

A variety of studies have been carried out to detect AD from spontaneous speech. Some are directly performed on the voice recordings, with either manually selected acoustic features [12][13][14] or automatically learned representations from deep neural network [15]. Although acoustic markers can provide valuable supportive assessment of dementia, discarding information in higher levels of speech may limit model's ability to evaluate the language function comprehensively.

Some studies are devoted to build predictive models with hand-crafted features from transcripts. For example, Wankerl et al. (2017) [16] employed an n-gram based model on Pitt Corpus [17] and achieved an accuracy of 77. Bucks et al. (2000) [18] conducted a linear discriminant analysis (LDA) of spontaneous speech from 8 AD participants and 16 healthy controls. Eight linguistic features were extracted, including part-of-speech (POS) tag frequencies and measures of lexical diversity, and a cross-validation accuracy of 87.5 was obtained.

Some other studies incorporated both linguistic and acoustic features to detect AD. Fraser et al. (2016) [19] utilized logistic regression with nominal outputs, and selected 35 top-ranked features out of 370 using data from Pitt Corpus. The best cross-validation accuracy achieved is 87.5. Khodabakhsh et al. (2015) [20] used the recordings of Turkish conversational speech of 32 AD and 51 control subjects, obtaining an accuracy of over 80.

One common challenge with the methods above is that handcrafted feature based approaches heavily rely on the linguistic and medical expertise of researchers. Some features may be too nuanced to be detected, especially at the early stages of dementia. Moreover, a bunch of studies have the drawback of small sample size, which corrupts the reliability of their conclusions.

In order to address these issues, Karlekar et al. (2018) [21] applied neural models on utterances from four speech tasks in Pitt corpus and obtained the best accuracy of 91.1. However, they broke each speech transcript into individual utterances and treated those utterances as independent samples. This makes the

http://dx.doi.org/10.21437/Interspeech.2019-2872

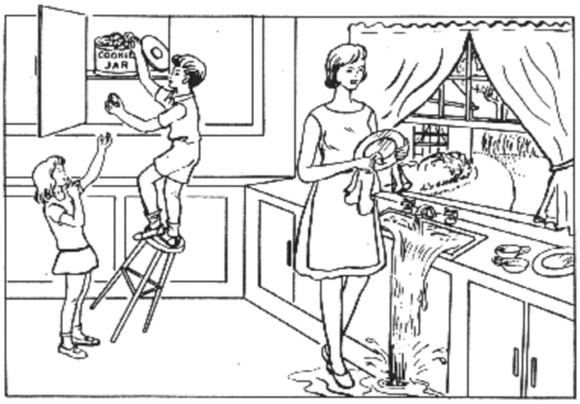

Figure 1: *Boston cookie theft stimulus photo. Participants were asked to describe all events in the image.*

model lose the ability to check thematic coherence and mention of key concepts in the speech. Besides, training on imbalanced data set (11458 utterances from AD and 2904 from control interviewees) may limit the performance of model on real application where the majority of population are AD-free.

Our study differs from previous works in several ways. First, like Karlekar et al., we choose neural models to avoid extracting feature manually. Second, we consider a larger sample size than most previous works. What's more, to have a comprehensive view of language function at different levels, we incorporate two neural architectures to analyze both local speech patterns and global macro-linguistic functions.

3. Data

The data we utilize are derived from the DementiaBank (DB) Pitt corpus[1], the largest publicly available corpus of narrative speech from AD and elderly healthy subjects [17]. The data were collected between 1983 and 1988. The English speaking participants were asked to perform four tasks. Specifically, we use the data from Boston Cookie Theft task (because data from control group are absent from other three tasks). Boston Cookie Theft is a picture description task from the Boston Diagnostic Aphasia Examination [22]. Participants were shown an image and asked to describe the scene they see (See Figure 1). Narrative speech was recorded and then manually transcribed at the word level following the TalkBank CHAT (Codes for the Human Analysis of Transcripts) protocol [23]. For this study, we use the transcripts of the recordings and keep the word-level transcription. Filler words (like um, uhh), punctuations, pauses, and repetitions were kept to obtain an accurate reflection of the actual speech.

To be eligible for study enrollment, all participants were required to be at least 44 years old and have an initial MMSE score above 10, have 7 or more years of education and no history of neurological disorders [19]. MMSE is composed of a series of questions designed to assess different cognitive functions. With a maximum score of 30, an MMSE score of 27 and above is suggestive of not having a dementia related disease [24], while a score below 24 suggests dementia in some stage [25]. Regarding to the dementia group, we narrow our selection to participants with a diagnosis of possible AD or probable AD, resulting in 256 samples. Together with the 242 recordings from the elderly normal control group, a total of 498 recordings were used in our study. Demographic details of the data set are given in Table 1.

Table 1: *Demographics of Pitt corpus data*

	AD ($n = 256$)	**Control** ($n = 242$)
Age (years)	$71.8\ (8.4)$	$64.9\ (7.7)$
Gender (male/female)	$87/169$	$88/154$
Mini-Mental State Exam	$18.6\ (5.1)$	$29.1\ (1.1)$

4. Methods

4.1. Framework Overview

As shown in Figure 2, the proposed attention-based hybrid network contains following five components:

1. Embedding layer: map each word in input transcript into a continuous vector;
2. GRU layer: apply bidirectional GRU on the top of word embedding layer to get contextual word representations;
3. CNN layer: perform one-dimensional convolution on embedded word sequence to extract local n-gram patterns from transcript;
4. Attention layer: take as input the output of previous layer (feature map from CNN or contextual word representations from biGRU), calculate weight vector that measure the importance of each feature and compute the transcript-level vector as a weighted sum of features;
5. Output layer: concatenate transcript-level vectors from CNN and GRU branches, and feed the resulting vector to a fully-connected layer for softmax classification.

These components are presented in detail in this section.

4.2. CNN

Convolutional neural networks (CNNs) are capable of extracting n-gram embeddings from the input sequence and composing an informative latent representation of the text. Building classical n-gram models is computationally expensive as the time needed is increased exponentially to the order of the vocabulary[26]. In contrast, CNNs are more efficient in terms of representation. One-dimensional convolution considers windows of n consecutive words. For each window size, a set of filters are applied to generate corresponding feature maps. After training, randomly initialized filters turn into specific n-gram detectors. By sliding filters with varying sizes, unigram through pentagram embeddings are generated accordingly. Afterwards, ReLU function is applied to introduce non-linearity and max-pooling layer is added to aggregate the most salient semantic features. The resulting feature maps are concatenated and passed to the following layer.

Although good at detecting local connectivity and extracting n-gram features, CNN has inherent limitation of being difficult to model long distance dependencies and global information.[27]

4.3. GRU

Recursive models are more suitable for tasks with sequential input. Recurrent neural networks (RNNs) [28] have memory over

[1] http://dementia.talkbank.org. Downloaded September 7, 2018.

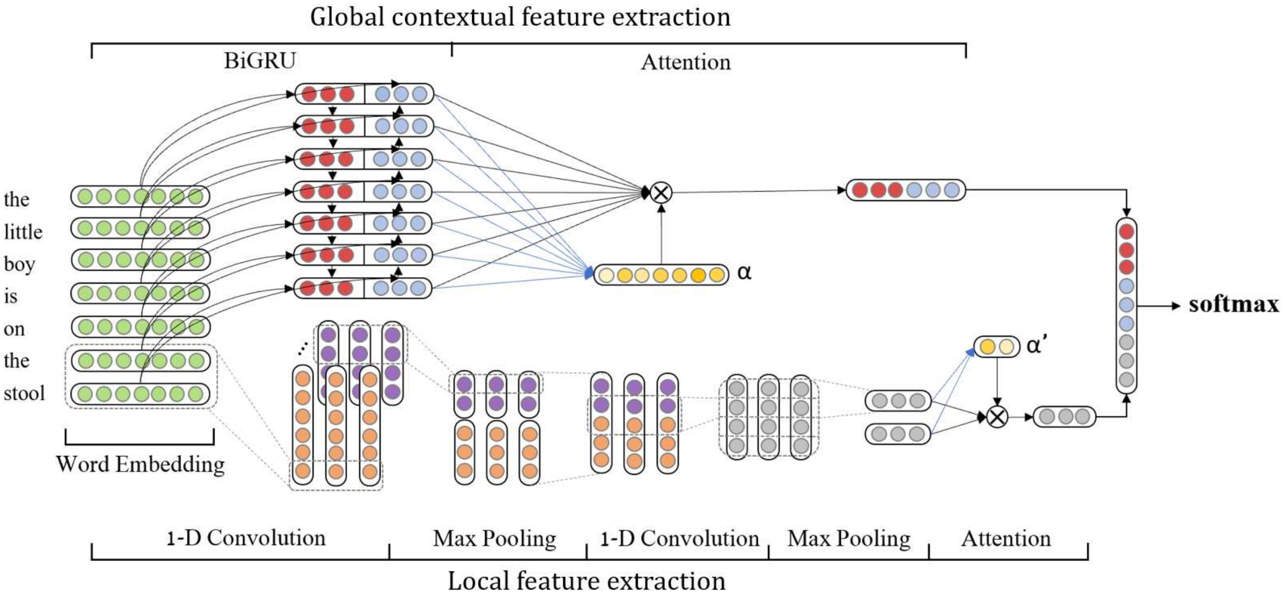

Figure 2: *Architectures of our proposed model of attention-based hybrid networks. Input transcripts are embedded and go through attentive BiGRU and attentive CNN branches. The resulting representations are concatenated and then fed to the following classification layer.*

previous timestamps and use this memory in computing current status. Unlike CNNs, RNNs have flexible computational steps, which provides better capability of modeling word sequence of arbitrary length. As vanilla RNN suffers from the vanishing gradient problem, we use a gated variant, Gated Recurrent Units(GRU) [29] in our model. GRU has two gates, one update gate controlling the information flows into memory and one reset gate controlling the information flows out of memory. Compared to another RNN variant long short-term memory (LSTM), GRU is of less complexity and needs less training time.

In order to consider both previous and future context during training, our model utilize bidirectional GRU (BiGRU) structure. BiGRU consists of a forward and a backward GRU. Hidden states of two GRUs are concatenated at each location, resulting in a sequence of hidden states containing information from both directions.

4.4. Attention Mechanism

Attention is most commonly used in sequence-to-sequence models. Attentive neural networks automatically focus on the decisive sub-parts of the sentence and capture the most informative semantics in the text.

Attention mechanism has demonstrated great success in a variety of natural language processing tasks, including question answering [30], machine translations [31] and speech recognition [32]. More recently, Transformer, a sequence-to-sequence model based purely on attention mechanism, and Bert [33], a language representation model build upon Transformer Encoder, redefined the state of the art for over ten NLP tasks and even surpassed the performance of human in several challenges [34].

Inspired by the success of these works, we deploy attention mechanism in our model. Specifically, let $X \in \mathbb{R}^{d\times T}$ be the input matrix to the attention layer. For BiGRU branch, T is the number of hidden states and d is the dimension of the contextual word embedding. For CNN branch, T is the length of generated feature map and d is the number of filters. The transcript-level representation r is computed by an weighted sum of the input features, similar to feed-forward attention mechanism in [35] and [36]:

$$M = \tanh(w^T X) \tag{1}$$

$$\alpha = softmax(M) \tag{2}$$

$$r = X\alpha^T \tag{3}$$

where M is the output of a feed-forward function and w is the corresponding learnable parameter. α is the weight reflects the relative importance of each input feature. The dimension of w, α, r is d, T, d separately.

By applying attention mechanism, we enable the model to pay more attention to the n-gram patterns and contextual word vectors that are of greater importance in distinguishing transcripts of AD subjects from those of normal elders.

4.5. Hybrid Network

CNNs and RNNs have different objectives when modeling a text. While RNNs try to create a semantic composition of an arbitrarily long sentence, CNNs try to extract salient n-gram patterns. Observing both models achieve results comparable to the best performance in previous work, and considering they each have their own complementary strengths, we propose an attention-based hybrid network to integrate these two modules. Our hybrid framework includes two parts, a word-level attention module(BiGRU branch) to summarize global semantics, and a feature-level(CNN branch) attention module to capture local patterns. Two branches have the same input and are trained jointly. The output vectors of them are concatenated and the final vector is fed to a fully-connected layer followed by a softmax classifier.

5. Results

To evaluate the proposed method, a standard 10-fold cross-validation procedure is performed. Data are prepared as described in Section 3. Table 2 reports the average accuracy with standard deviation across the folds. Neural models composed of one or two component branches, with or without attention are tested.

For BiGRU, both forward and backward GRUs have 128 hidden units. For CNN, three 1-D convolution layers are stacked, followed by a 1-D max pooling layer. Filter sizes of [1,2,3,4,5] are used for the first Convolution layer and the resulting feature maps are concatenated. Filter size of 3 is used for the other two convolution layers. Number of filters is 128 for each filter size in all the convolution layers. All models had a vocabulary size of 1751, and use an Adam optimizer with a learning rate of $1e^{-3}$ and a decay rate of $2e^{-5}$. Best cross-validation accuracy of 97.42 is achieved by our proposed hybrid attentive model, fine-tuned with 100-dimensional GloVe word embedding.

From the table, we first notice that incorporating attention mechanism benefits both CNN and BiGRU baselines substantially. Another observation is that hybrid network makes a considerable improvement of performance over either of its component branches. This result shows that two branches make complementary contributions to the final output. Through hybrid attentive architecture, our model learns which features from different language levels to focus on.

Notably, different experiment settings of embedding layer significantly affect model's ability of identifying dementia. The pre-trained word embeddings we tested include GloVe [37] based on word co-occurrence, and word2vec based on skip-gram with negative sampling [38]. When fixing the embedding layer during training, the accuracy of hybrid attention model with pre-trained 100-dimensional GloVe (identical to the model with best accuracy, except for whether or not fine-tuning of the embedding layer is allowed) drops from 97.42 to 79.26. While using different word embeddings, 100 and 300-dimentional GloVe, 300-dimentional word2vec, and even 100-dimentional randomly initialized (noted as rand-init in the table) embedding, does not make much difference, as long as the embedding layer is trainable.

Word embeddings are typically pre-trained in a large external corpus. The learned word vectors can capture general syntactical and semantic information and are proven to be efficient in downstream tasks [38]. In our case, however, randomly initialized embedding shows a comparable performance with pre-trained embedding. One possible reason is that the topic of the Cookie Theft task is fixed and the number of unique words found in the recordings is relatively small, so the word embedding can be learned during the training phase from the transcript themselves. Another explanation may be the fact that assessing mental status and language ability of a subject is a process different from performing conventional NLP tasks. In our situation, fine-tuning with respect to the specific objective is necessary for obtaining good performance. In spite of that, using pre-trained word embedding is still a good practice since it gives a good starting point for embedding matrix and speeds up the training process.

6. Conclusions

In this paper, we proposed an attention-based hybrid network, and obtained a new benchmark cross-validation accuracy of 0.97 on "Boston Cookie Theft" data set. Although the method is not supposed to replace doctors examinations, it demonstrates the potency of automated speech analysis as a useful tool for early screening of AD. By identifying people at high risk of the disease, we help them get access to professional examinations and possible treatment in time.

One of the two key ideas of our model is incorporating attention mechanism to allow the network to focus on the decisive features. The other is combining CNN and GRU branches to have a more comprehensive assessment of language ability from different aspects.

For future work, a speech recognition module could be incorporated to make the approach fully automated. Starting from audio recordings can introduce acoustic features not reflected by transcripts (like the speed of the speech and the duration of the pauses), as well as avoid labor-intensive hand-transcription. Another potential work is adapting our model to a more challenging work of identifying subjects with mild cognitive impairment (MCI). MCI is the intermediate stage between normal aging and AD, and is regarded as a suitable stage for early intervention and potential treatment. Hopefully, this work will lead us one step closer to a quick and reliable early diagnosis of Alzheimer's disease.

Table 2: *Accuracy comparison between different models (%).*

Model	Embedding details	Accuracy
CNN	Glove100, trainable	92.84 (5.68)
Att-CNN	Glove100, trainable	94.83 (4.14)
BiGRU	Glove100, trainable	94.63 (4.13)
Att-BiGRU	Glove100, trainable	95.45 (5.44)
Att-CNN+Att-BiGRU	Glove100, trainable	**97.42 (3.09)**
Att-CNN+Att-BiGRU	Glove100, static	79.26 (5.60)
Att-CNN+Att-BiGRU	rand-init, trainable	97.03 (3.64)
Att-CNN+Att-BiGRU	Glove300, trainable	96.23 (4.33)
Att-CNN+Att-BiGRU	Word2vec, trainable	97.01 (2.00)

7. References

[1] A. Burns and S. Iliffe, "Clinical review: Alzheimers disease," *British Medical Journal*, vol. 338, p. b158, 2009.

[2] A. Association *et al.*, "2013 alzheimer's disease facts and figures," *Alzheimer's & dementia*, vol. 9, no. 2, pp. 208–245, 2013.

[3] L. Herman, A. Atri, and S. Salloway, "Alzheimer's disease in primary care: The significance of early detection, diagnosis, and intervention," *The American journal of medicine*, vol. 130, no. 6, p. 756, 2017.

[4] J. Wang, B. J. Gu, C. L. Masters, and Y.-J. Wang, "A systemic view of alzheimer diseaseinsights from amyloid-β metabolism beyond the brain," *Nature Reviews Neurology*, vol. 13, no. 10, p. 612, 2017.

[5] L. The, "Alzheimer's disease: expedition into the unknown." *Lancet (London, England)*, vol. 388, no. 10061, p. 2713, 2016.

[6] R. Y. Lo, "The borderland between normal aging and dementia," *Tzu-Chi Medical Journal*, vol. 29, no. 2, p. 65, 2017.

[7] "What happens to the brain in alzheimer's disease," https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease, accessed: 2019-03-30.

[8] D. F. Tang-Wai and N. L. Graham, "Assessment of language function in dementia," *Geriatrics*, vol. 11, no. 2, pp. 103–110, 2008.

[9] E. Catricalà, P. A. Della Rosa, V. Plebani, D. Perani, P. Garrard, and S. F. Cappa, "Semantic feature degradation and naming performance. evidence from neurodegenerative disorders," *Brain and language*, vol. 147, pp. 58–65, 2015.

[10] M. Y. Savundranayagam, M. L. Hummert, and R. J. Montgomery, "Investigating the effects of communication problems on caregiver burden," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 60, no. 1, pp. S48–S55, 2005.

[11] S. B. Chapman, J. Zientz, M. Weiner, R. Rosenberg, W. Frawley, and M. H. Burns, "Discourse changes in early alzheimer disease, mild cognitive impairment, and normal aging," *Alzheimer Disease & Associated Disorders*, vol. 16, no. 3, pp. 177–186, 2002.

[12] J. J. G. Meilán, F. Martínez-Sánchez, J. Carro, D. E. López, L. Millian-Morell, and J. M. Arana, "Speech in alzheimer's disease: Can temporal and acoustic parameters discriminate dementia?" *Dementia and Geriatric Cognitive Disorders*, vol. 37, no. 5-6, pp. 327–334, 2014.

[13] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert *et al.*, "Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112–124, 2015.

[14] J. Weiner, C. Herff, and T. Schultz, "Speech-based detection of alzheimer's disease in conversational german." in *INTERSPEECH*, 2016, pp. 1938–1942.

[15] T. Warnita, N. Inoue, and K. Shinoda, "Detecting alzheimer's disease using gated convolutional neural network from audio data," *arXiv preprint arXiv:1803.11344*, 2018.

[16] S. Wankerl, E. Nöth, and S. Evert, "An n-gram based approach to the automatic diagnosis of alzheimer's disease from spoken language." in *INTERSPEECH*, 2017, pp. 3162–3166.

[17] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis," *Archives of Neurology*, vol. 51, no. 6, pp. 585–594, 1994.

[18] R. S. Bucks, S. Singh, J. M. Cuerden, and G. K. Wilcock, "Analysis of spontaneous, conversational speech in dementia of alzheimer type: Evaluation of an objective technique for analysing lexical performance," *Aphasiology*, vol. 14, no. 1, pp. 71–91, 2000.

[19] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify alzheimers disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[20] A. Khodabakhsh, F. Yesil, E. Guner, and C. Demiroglu, "Evaluation of linguistic and prosodic features for detection of alzheimers disease in turkish conversational speech," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, p. 9, 2015.

[21] S. Karlekar, T. Niu, and M. Bansal, "Detecting linguistic characteristics of alzheimer's dementia by interpreting neural models," *arXiv preprint arXiv:1804.06440*, 2018.

[22] H. Goodglass, E. Kaplan, and B. Barresi, *Boston Diagnostic Aphasia Examination Record Booklet*. Lippincott Williams & Wilkins, 2000.

[23] B. MacWhinney, *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press, 2014.

[24] S. O. Orimaye, J. S. Wong, K. J. Golden, C. P. Wong, and I. N. Soyiri, "Predicting probable alzheimers disease using linguistic deficits and biomarkers," *BMC bioinformatics*, vol. 18, no. 1, p. 34, 2017.

[25] "medical tests," https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests, accessed: 2019-06-20.

[26] F. C. Pereira, Y. Singer, and N. Tishby, "Beyond word n-grams," in *Natural Language Processing Using Very Large Corpora*. Springer, 1999, pp. 121–136.

[27] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *ieee Computational intelligenCe magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[28] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[29] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[30] T. Rocktäschel, E. Grefenstette, K. M. Hermann, T. Kočiskỳ, and P. Blunsom, "Reasoning about entailment with neural attention," *arXiv preprint arXiv:1509.06664*, 2015.

[31] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[32] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Advances in neural information processing systems*, 2015, pp. 577–585.

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[35] C. Raffel and D. P. Ellis, "Feed-forward networks with attention can solve some long-term memory problems," *arXiv preprint arXiv:1512.08756*, 2015.

[36] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 207–212.

[37] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[38] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.