Automatic Hierarchical Attention Neural Network for Detecting AD

Yilin Pan[1], *Bahman Mirheidari*[1], *Markus Reuber*[2,3], *Annalena Venneri*[4], *Daniel Blackburn*[4], *and Heidi Christensen*[1]

[1]Dept. of Computer Science, University of Sheffield, UK
[2]Academic Neurology Unit, University of Sheffield, UK
[3]Royal Hallamshire Hospital, UK
[4]Sheffield Institute for Translational Neuroscience, University of Sheffield, UK
{yilin.pan, heidi.christensen}@sheffield.ac.uk

Abstract

Picture description tasks are used for the detection of cognitive decline associated with Alzheimer's disease (AD). Recent years have seen work on automatic AD detection in picture descriptions based on acoustic and word-based analysis of the speech. These methods have shown some success but lack an ability to capture any higher level effects of cognitive decline on the patient's language. In this paper, we propose a novel model that encompasses both the hierarchical and sequential structure of the description and detect its informative units by attention mechanism. Automatic speech recognition (ASR) and punctuation restoration are used to transcribe and segment the data. Using the DementiaBank database of people with AD as well as healthy controls (HC), we obtain an F-score of 84.43% and 74.37% when using manual and automatic transcripts respectively. We further explore the effect of adding additional data (a total of 33 descriptions collected using a *'digital doctor'*) during model training, and increase the F-score when using ASR transcripts to 76.09%. This outperforms baseline models, including bidirectional LSTM and bidirectional hierarchical neural network without an attention mechanism, and demonstrate that the use of hierarchical models with attention mechanism improves the AD/HC discrimination performance.

Index Terms: Dementia detection, automatic diagnosis, hierarchical attention network, linguistic features

1. Introduction

Dementia is a type of neurodegenerative disease, and the most common cause is Alzheimer's Disease (AD) [1]. As a result of an aging society, the number of people with dementia is increasing rapidly all over the world [2]. In the absence of a cure, earlier and better diagnosis is critical for the timely treatment. It has been found that although memory impairment is the main early symptom for AD, language and speech abilities also decline, even in the very early stages [3].

Several tests, used routinely for diagnosis, have elements focusing on discourse and of those the *picture description* task is a constrained task that relies less on episodic memory, but requires more semantic knowledge and retrieval ability [4]. In the test, a picture is presented as a prompt, and the patient is asked to describe what they see in the picture. During this process, the answer is often recorded and this is subsequently used when the neuropsychologist scores the test. This is a relatively time consuming procedure, so investigating ways of automating the scoring from the recorded audio is of interest. The most commonly used picture prompt is a line drawing called the "Cookie Theft" picture originating from a test for aphasia [5].

It is well known that people with dementia shows signs of cognitive decline at both the word and sentence levels. At the word level this is evident in e.g., the number of pictorial themes, the repairing errors and the vocabulary richness. At the sentence level this is seen in things like sentence coherence, idea density (the rate at which ideas or elementary predictions are expressed in an utterance or text) [4] and in how the eye is guided across the picture prompt to elicit the description [6]. However, so far automatic scoring approaches have failed to take this *hierarchical structure* of the transcript directly into account. Inspired by that, this paper proposes a hierarchical bidirectional neural network, for extracting different level features. Furthermore, to distinguish the informative words like "mother", and "cookie jar" from more common words like "a" and "the" in the picture descriptions, an attention layer at both the word level and sentence level of the neural network is included.

To evaluate the model, the use of both manual and automatic transcripts (generated by using an automatic speech recogniser (ASR)) is explored. Cookie Theft picture descriptions from the DementiaBank database [7], as well as from an in-house database collected using our Intelligent Virtual Agent (IVA; 'digital doctor') system [8, 9] are used to test our model. To the best of our knowledge, this is the first use of a bidirectional hierarchical recurrent neural network combined with an attention mechanism (BHANN) for dementia detection.

The result shows that our model can achieve state-of-the-art results when applied to both manual and automatic transcripts. By comparing our model with two baseline models, we can conclude that the proposed hierarchical structure and attention layer both contribute to the final improvement in AD detection.

In the remainder of this paper Section 2 presents the background and related work, Section 3 presents our proposed system, Sections 4 and 5 describe the experimental setup and results, and finally, the conclusions are given in Section 6.

2. Related Work

With the outstanding performance of deep learning for many domains like speech processing and natural language processing (NLP), researchers have started to introduce this technology into automatic detection and classification of cognitive impairment like AD. A model proposed in [10] was designed to combine deep neural networks with deep language models for predicting Mild Cognitive Impairment (MCI). In [11], a Gated Convolutional Neural Network (GCNN) was used to capture the temporal information in audio paralinguistic features based on the features extracted by Opensmile [12]. Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Convo-

lutional Neural Network (CNN) and their combination model were applied to extract linguistic features from transcripts of a picture description task [13]. In [14], two n-gram neural network language models for healthy controls (HC) and patients with dementia were built respectively for detection. It has been proven that, for the same database, neural networks achieved a better result compared with traditional machine learning methods [11, 14]. The reason is, compared with deep learning methods, statistical models with handcrafted features are less robust and harder to design, as it requires more expert knowledge.

However, applying methods like deep learning, which rely on the availability of large amounts of data, to the healthcare domain is somewhat problematic as there is often a lack of sufficient data available. In the area of dementia detection, except one open source dataset named Dmentiabank, the other databases are mostly self-collected and not shareable due to ethical constraints. When working with deep learning models, the question of how to design a model to make good use of the information in limited data is therefore critical in our task.

In dementia detection, word embedding has been proposed to be used for converting spoken transcripts into vectors for detecting cognitive decline [9]. Two recent techniques, 'w2vec' [15] and 'GloVe' [16], which consider the context in the text, were proposed and achieved a better performance compared with traditional methods like bag-of-words (BOW) [17]. In contrast to w2vec, GloVe is designed to include the word-word co-occurrence counts information into the method instead of focusing only on the probabilities of words in the context.

A hierarchical model, which could extract both word-level and sentence-level information, has shown its efficiency in essay classification and scoring [18, 19]. For sequence modeling, bidirectional RNNs achieve outstanding performance by exploiting information both from the past and the future of the input sequence. Combined with attention mechanism [20] it can identify the more or less important content in transcripts. Furthermore, hierarchical models can also be applied for *spoken* transcripts [21] even though it was proposed for written essays originally. Working in the area of spoken language poses some challenges though. In particular, unlike written text, spoken transcripts obtained from an ASR system does not have any punctuation. Adding punctuation can improve the word stream's readability, not only for humans but also for natural language processing tools [22, 23]. In our task, automatic punctuation restoration is introduced to find proposed start and end positions of likely utterances in the ASR transcripts.

In this paper, a hierarchical bidirectional attention network was proposed for dementia detection. To exam the model's efficiency, we use both manual and automatic transcripts for the "Cookie Theft" picture description task. In addition, we show that results can be further improved by adding our own in-house collection of picture descriptions to the DementiaBank dataset.

3. System

A hierarchical model with attention layer is employed on spoken transcripts for dementia detection. The overall structure of our model is shown in Fig. 1. The dashed lines are used to represent dropout. This section describes how to represent a transcript in a vector and then estimate its diagnostic class.

3.1. Word Embedding

When applying deep learning methods on a text classification task, words first need to be transformed into high-dimension distribution vectors. This process is called word embedding and the output is called word vectors. The benefit of this operation is to capture the semantic information about the words in sentences. The efficiency of using word vectors for dementia detection has been demonstrated before [9].

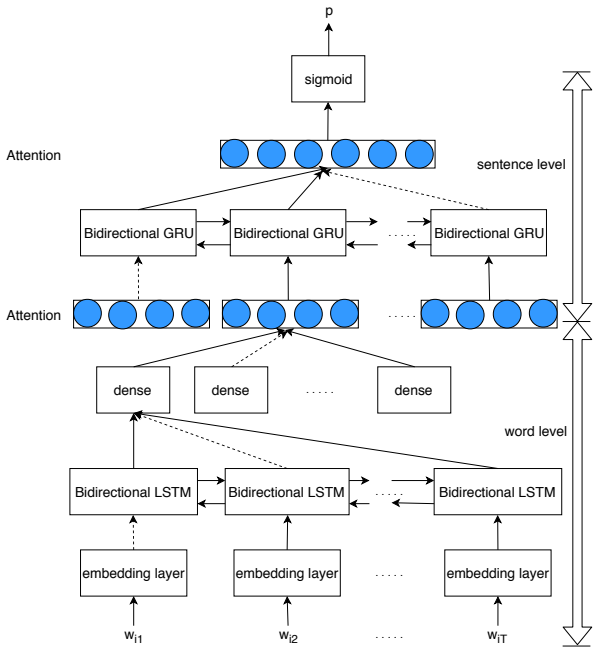

Figure 1: *Structure of proposed system: BHANN.*

In our model, w_{it} with $t \in [1, T]$ and $i \in [1, L]$ is used to represent the tth word in ith sentence. Each word w_{ij} is encoded into a fixed dimensional vector x_{ij} by a pre-trained embedding matrix W_e using the 'GloVe' algorithm at first. The word embedding matrix is trainable in the model.

3.2. Word Level Structure

It has been shown that patients with dementia tend to phrase things using 'vague' or ineffective information, repeat words and phrases more frequently and identify pictorial themes inaccurately. To extract such word-level characteristic patterns from the variable length sequence, a bidirectional LSTM is applied on the word vectors.

The bidirectional LSTM can get the representation of words and their surrounding information from forward and backward directions. In our model, the final representation h_{it} is achieved by adding the vector from forward LSTM $\overleftarrow{h_{it}}$ and backward $\overrightarrow{h_{it}}$.

$$\begin{aligned} \overleftarrow{h_{it}} &= \overleftarrow{LSTM}(W_e w_{it}, h_{it-1}) \\ \overrightarrow{h_{it}} &= \overrightarrow{LSTM}(W_e w_{it}, h_{it-1}) \\ h_{it} &= \overrightarrow{h_{it}} + \overleftarrow{h_{it}} \end{aligned} \tag{1}$$

A dense layer with ReLU activation function is applied in the following:

$$d_{it} = ReLU(W_d h_{it} + b_d) \tag{2}$$

In addition, in order to model that the importance of each word will differ, an attention mechanism is used followed by the dense layer. Specifically,

$$u_{it} = tanh(W_w d_{it} + b_w)$$
$$\alpha_{it} = \frac{exp(u_{it}^T u_w)}{\sum_t exp(u_{it}^T u_w)} \quad (3)$$
$$s_i = \sum_t \alpha_{it} h_{it}$$

u_{it} is the hidden representation of the one-layer Multi-layer Perception (MLP). Then we measure the word importance by calculating the similarity of u_{it} with a word-level vector u_w, which is initialised randomly and used as a high-level representation of a fixed query over the words like in a memory network [24]. Finally, a sentence representation s_i is calculated by a weighted sum of the words in the ith sentence.

3.3. Sentence Level Structure

After obtaining the sentence representation, a bidirectional GRU layer is applied to each sentence for sentence level information extraction. According to [25], the choice of whether to use an LSTM or an RNN mostly depends on the dataset and corresponding tasks. Likewise, we base our decision on the experimental performance. Given the sentence vector s_i, we can get the transcript representation by using a similar structure as for the word level model, but replacing the LSTM with GRU.

$$\overleftarrow{h_i} = \overleftarrow{GRU}(s_i, h_{i-1})$$
$$\overrightarrow{h_i} = \overrightarrow{GRU}(w_i, h_{i-1}) \quad (4)$$
$$h_i = \overrightarrow{h_i} + \overleftarrow{h_i}$$

Then an attention layer is applied:

$$u_i = \tanh(W_s h_i + b_s)$$
$$\alpha_i = \frac{exp(u_i^T u_s)}{\sum_t exp(u_i^T u_s)} \quad (5)$$
$$v = \sum_t \alpha_i h_i$$

where v is the representation of the whole transcript. Similarly, u_s is initialized randomly.

Finally, one dense layer with a sigmoid function is applied for classification:

$$p = sigmoid(W_c v + b_c) \quad (6)$$

where W_c and b_c are the weight vector and bias vectors respectively and p is used to represent the possibility of the classes the transcript belongs to.

4. Experiment Setup

4.1. Datasets

In our paper, the DementiaBank and an in-house database of Cookie Theft picture descriptions prompted by an IVA [8] are used for examining our model. As our system is designed for binary classification, in DementiaBank, the diagnostic class for some of the participants changed during their longitudinal follow-up (mostly with MCI turning into AD), and those participants detected as MCI have been excluded for this study. In total, 222 samples from 89 HC and 255 from 168 patients with AD were selected from the original 551 files. For the IVA dataset, 33 out of 76 files for participants with a diagnosis of 17 HCs and 16 ADs were selected. The dataset details are shown in Table 1. It can be found that a mismatch exits on the average utterance length between the two datasets.

Table 1: *Dataset information*

Dataset(No)	Len.	Utts.	Spks.	Avg. Utts.
DemBank(477)	7h40m	6124	257	4.50s
IVA(33)	40m	264	33	9.04s

4.2. Automatic Transcript Generation

To get the automatic transcriptions, the Kaldi [26] toolkit hybrid TDDN-LSTM recipe was used for training the ASRs. For language models we followed the in-domain 3/4 grams with the KN/Turing smoothing. Note that we added an additional 64 hours worth of conversational data (see [9] for more details) to boost the acoustic model of the ASRs. Finally, we got the transcripts with a word error rate (WER) of 41.6% on DementiaBank and 33.8% on IVA.

To add punctuation in the ASR transcripts, the pre-trained text-only model shared in github[1] is used. It predicts placement and type of punctuation by using a bidirectional LSTM network with attention layer. Further details can be found in [22].

4.3. Evaluation Setting

We used 10-fold cross-validation (CV) to segment DementiaBank and ensure that the particular speaker is not included in either train (8 folds), test (1 fold) or development (1 fold) set at the same time (speaker independent). To ensure it, a randomly ordered speaker list is generated at first and then a list for all the files is generated. This file list is split (train, test and development), and subsequently the sets are adjusted so that all files from each speaker is in one set only. Our method can not only ensure that files from the same speaker are found in the same set, but it also keeps the number of files in the three sets in each fold as balanced as possible. The IVA data was only used for the training set. When combining with the IVA data, we kept the 10-fold segmentation lists unchanged except adding the IVA data into training set, so test folds have stayed the same and are directly comparable.

4.4. Baseline Models

In order to prove the efficiency of the proposed system, two baseline models are designed: a bidirectional LSTM (BLSTM) model and a hierarchical bidirectional recurrent neural network (HBRNN). BLSTM treats the spoken transcript as a sequence of words rather than a sequence of sentences. The input is the words in the joint set of all sentences s_i with $i \in [1, L]$ in the transcript, like $[s_1, s_2, ..., s_L]$. Following the bi-LSTM layer is a dense layer with a sigmoid activation function. This model is designed to test the benefit of having the hierarchical mechanism. HBRNN has the same structure as the hierarchical bidirectional recurrent neural network we described in Section 3.3, but without the attention mechanism for testing the benefit of having the attention mechanism.

4.5. Model Configuration and Structure

In the word embedding part, Stanford's public available 100 dimensional vectors trained using GloVe on Wikipedia 2014 and

[1] https://github.com/ottokart/punctuator2

'Gigaword 5' is taken as our pre-trained embedding matrix. The files are tokenized with NLTK[2] tokenizer.

The proposed model is trained on a fixed number of epochs (20) and evaluated on the development set at each epoch. Batch size is set to 20 and the best model is selected on the F-score on the development set. The number of RNN units, including LSTM and GRU is set to 100 and the dense layer dimension in word level is set as 50. For the attention layers' dimension, both the sentence and word level is set to 30. The sentence length is set to 30 and we zero-pad shorter sentences. The sentence numbers in a transcript is set to 30 and we do zero-padding on shorter transcript. An adam optimizer with 0.001 learning rate is used. To avoid overfitting, we apply dropout to the output of all the functional layers. For all dropout layers, the dropout rate is set to 0.3. The final criteria are calculated by averaging the results on 10 fold CV. For the BLSTM model, we set the word number in a transcript as 30×30. All the parameters we mentioned above are selected on BHANN and then applied to the other models.

5. Result

To test the efficiency of our model, both manual and automatic transcripts are tested. First, the manual transcript is used and the result is presented in Table 2. The experiment is composed of three parts to verify the efficiency of the hierarchical structure, the attention mechanism, and the punctuation restoration. By comparing the results from our proposed system BHANN and the two baselines, we can draw the following conclusions:

- Effect of hierarchical mechanism: By comparing the result of BLSTM and HBRNN, we find that, after including hierarchical mechanism, the F-score can be improved from 75.02% to 78.26%. It proves that the hierarchical neural network is reasonable for our task.
- Effect of attention mechanism: After including attention mechanism to hierarchical model, the F-score was further increased compared with that on HBRNN, which proves the efficiency of attention mechanism for hierarchical model.
- The Influence of automatic punctuation: To evaluate the affect of automatic punctuation on our experiment, we add punctuation by the automatic punctuation restoration method on punctuation removed manual transcripts. The result shows that automatic punctuation restoration can cause about 3% decline on F-score.

Table 2: *The detection results of BHANN and the baselines on manual transcripts of DementiaBank.*

Punctuation	System	precision	recall	F-score
Manual	BLSTM	75.02%	73.73%	73.45%
Manual	HBRNN	78.26%	77.77%	75.68%
Manual	BHANN	**84.02**%	**84.97**%	**84.43**%
Automatic	BHANN	81.17%	81.23%	79.77%

Then, the automatic transcripts with automatic punctuation restoration is tested. Rather than only use the data from DementiaBank for training, we also included our in-house IVA data. We can draw the following conclusions from Table 3:

[2]http://www.nltk.org/

- Manual and automatic transcript: The performances of BLSTM, HBRNN and BHANN remained consistent on manual and automatic transcripts, which further proved the efficiency of our proposed system. In addition, we found that a gap existed between the F-score values of manual and automatic transcripts, which originates from two sources: WER in ASR system and punctuation restoration error.
- Effect of additional training data: After including 33 transcripts of our in-house IVA data into the training set, a better result can be achieved (last row) on recall and F-score. It proves that our model still has space to be improved if more training data can be included, even from a non-homogeneous source dataset.

Table 3: *The detection results of BHANN and the baselines on automatic transcripts of DementiaBank and IVA.*

Training Set	System	precision	recall	F-score
Dembank	BLSTM	68.18%	67.74%	66.44%
Dembank	HBRNN	74.03%	74.80%	72.11%
Dembank	BHANN	**79.22**%	76.33%	74.37%
Dembank+IVA	BHANN	78.83%	**77.73** %	**76.09**%

Our system compares favourably with previous methods working on *manual* transcripts of DementiaBank. In [3] and [27], an accuracy of 81.92% and a F-score of 77.50% was achieved, compared with 84.02% for our system. Even though [13] got a comparable result of 84.9%, it did not used 10-fold CV and a speaker independent way to evaluate the system. In addition, it used the whole DementiaBank with 551 files in the experiment.

Our result on *automatic* transcripts is also state-of-art. Compared with the automatic transcripts based dementia detection accuracy (62.3%) on DementiaBank presented in [9], we got an F-score of 76.09%. In [28], an almost similar precision (79%) was achieved by selecting features extracted from audio and transcripts at the same time. In addition to requiring more time and medical knowledge in comparison to our methods, it also used extra acoustic features to improve the result.

6. Conclusions

We present our study on an automatic approach for detecting dementia by utilising only the spoken transcripts for picture description. The employment of BHANN on manual and automatic transcript both achieved results improving on current state-of-art. By including our in-house dataset into the training set, the result was further improved.

Distinguishing between HC and AD is a much simplified task compared to the range of conditions doctors are facing in memory clinics, and we plan to explore a more realistic multi-classification tasks with adding diagnostic classes, including MCI and Functional Memory Disorder (FMD).

7. Acknowledgements

This work is supported under the European Unions H2020 Marie Skodowska-Curie programme TAPAS (Training Network for PAthological Speech processing; Grant Agreement No. 766287).

8. References

[1] A. Association *et al.*, "2017 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 13, no. 4, pp. 325–373, 2017.

[2] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang *et al.*, "Global prevalence of dementia: a delphi consensus study," *The lancet*, vol. 366, no. 9503, pp. 2112–2117, 2005.

[3] 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.

[4] K. D. Mueller, B. Hermann, J. Mecollari, and L. S. Turkstra, "Connected speech and language in mild cognitive impairment and alzheimers disease: a review of picture description tasks," *Journal of clinical and experimental neuropsychology*, vol. 40, no. 9, pp. 917–939, 2018.

[5] H. Goodglass and E. Kaplan, *The assessment of aphasia and related disorders*. Lea & Febiger, 1972.

[6] B. Mirheidari, Y. Pan, T. Walker, R. Markus, A. Venneri, D. Blackburn, and H. Christensen, "Detecting alzheimer's disease by estimating attention and elicitation path through the alignment of spoken picture descriptions with the picture prompt," 2019, unpublished.

[7] 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.

[8] B. Mirheidari, D. Blackburn, K. Harkness, A. Venneri, M. Reuber, T. Walker, and H. Christensen, "An avatar-based system for identifying individuals likely to develop dementia," in *Proc INTERSPEEH*. ISCA, 2017.

[9] B. Mirheidari, D. Blackburn, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Detecting signs of dementia using word vector representations," *Proc. Interspeech 2018*, pp. 1893–1897, 2018.

[10] S. O. Orimaye, J. S.-M. Wong, and J. S. G. Fernandez, "Deep-deep neural network language models for predicting mild cognitive impairment." in *BAI@ IJCAI*, 2016, pp. 14–20.

[11] 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.

[12] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

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

[14] J. Fritsch, C. Bergler, S. Wankerl, and E. Nöth, "Automatic diagnosis of alzheimers disease using neural networks language models," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, p." to appear", ISCA*, 2018.

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

[16] 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.

[17] Z. S. Harris, "Distributional structure," *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.

[18] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480–1489.

[19] D. S. McNamara, S. A. Crossley, R. D. Roscoe, L. K. Allen, and J. Dai, "A hierarchical classification approach to automated essay scoring," *Assessing Writing*, vol. 23, pp. 35–59, 2015.

[20] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[21] B.-H. Tseng, S.-S. Shen, H.-Y. Lee, and L.-S. Lee, "Towards machine comprehension of spoken content: Initial toefl listening comprehension test by machine," *arXiv preprint arXiv:1608.06378*, 2016.

[22] O. Tilk and T. Alumäe, "Bidirectional recurrent neural network with attention mechanism for punctuation restoration." in *Interspeech*, 2016, pp. 3047–3051.

[23] O. Tilk and T. Alumae, "Lstm for punctuation restoration in speech transcripts," in *Sixteenth annual conference of the international speech communication association*, 2015.

[24] S. Sukhbaatar, J. Weston, R. Fergus *et al.*, "End-to-end memory networks," in *Advances in neural information processing systems*, 2015, pp. 2440–2448.

[25] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[26] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," IEEE Signal Processing Society, Tech. Rep., 2011.

[27] A. Budhkar and F. Rudzicz, "Augmenting word2vec with latent dirichlet allocation within a clinical application," *arXiv preprint arXiv:1808.03967*, 2018.

[28] R. B. Ammar and Y. B. Ayed, "Speech processing for early alzheimer disease diagnosis: Machine learning based approach," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2018, pp. 1–8.