



**HAL**  
open science

## Exploring the efficacy of text embeddings in early dementia diagnosis from speech

Khaoula Ajroudi, Mohamed Ibn Khedher, Olfa Jemai, Mounim A El-Yacoubi

► **To cite this version:**

Khaoula Ajroudi, Mohamed Ibn Khedher, Olfa Jemai, Mounim A El-Yacoubi. Exploring the efficacy of text embeddings in early dementia diagnosis from speech. 16th International Conference on Human System Interaction (HSI ), Jul 2024, Paris, France. 10.1109/HSI61632.2024.10613581 . hal-04596848

**HAL Id: hal-04596848**

**<https://hal.science/hal-04596848v1>**

Submitted on 31 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Exploring the Efficacy of Text Embeddings in Early Dementia Diagnosis from Speech

Khaoula Ajroudi\*, Mohamed Ibn Khedher†, Olfa Jemai\* and Mounim A. El-Yacoubi‡

\* Research Team in Intelligent Machines (RTIM), National Engineering School of Gabes (ENIG)

University of Gabes, Tunisia

khaoula.ajroudi@gmail.com, olfa.jemai@isimg.tn

†IRT - SystemX, 2 Bd Thomas Gobert, 91120 Palaiseau, France

mohamed.ibn-khedher@irt-systemx.fr

‡Samovar, Telecom SudParis, Institut Polytechnique de Paris, 19 place Marguerite Perey, 91120 Palaiseau, France

mounim.el\_yacoubi@telecom-sudparis.eu

**Abstract**—Language impairment is a key biomarker for neurodegenerative diseases such as Alzheimer’s disease (AD). With the rapid growth of Large Language Models, natural language processing (NLP) has become a preferred modality for the early prediction of AD from speech. In this work, we propose a two-stage process for early detection of AD from transcriptions of speech. The first step involves extracting a discriminative text embedding representation using public models from OpenAI. This embedding serves as input for a machine learning classifier in the second stage. In this paper, we investigate three text embedding models and eight machine learning classifiers, both deep learning (DL) based and non-DL based. The evaluation was conducted using the public ADReSSo dataset of 237 patients. The results show that models "ada-002" and "3-small" produce discriminative embeddings that lead to good performance when combined with a Deep Neural Network in classification, achieving accuracy rates of 83.10% and 84.51%, respectively.

**Index Terms**—Alzheimer’s Disease, Deep Learning, Large Language Model, Natural Language Processing, Speech Classification, Text Embedding.

## I. INTRODUCTION

Alzheimer’s Disease (AD) is the leading cause of dementia and one of the most common neurodegenerative disorders. AD prevalence in ageing population is surging, causing increasing charge on society as a whole [1], [2]. AD is characterised by memory impairment and cognitive decline. Many studies have suggested that language impairment is present in most AD patients and may be used to help diagnose and assess the severity of the disease [3].

A patient’s spoken language carries a lot of information on their cognition state and how much their brain has been damaged. Investigating language disorders using Natural Language Processing (NLP) plays an important part in the investigation of cognitive disorders in general.

However, human evaluation in clinical settings can be subjective, leading to inconsistent diagnoses even among experts [4]. This is why, in a Precision Medicine setting, the evolution of Artificial Intelligence including Machine Learning algorithms, is a potential step forward for providing an objective measurability.

The emergence of Large Language Models (LLMs) and Transformer architectures represents a significant advancement in the field of artificial intelligence, particularly in tasks such as text classification and text embedding extraction. These algorithms, exemplified by models like GPT (Generative Pre-trained Transformer), are capable of autonomously learning to understand and generate text from vast datasets. By utilizing transformer architectures, these new models, known as text embedding models, can capture complex semantic information and produce high-quality vector representations for text. Their emergence has revolutionized numerous domains, such as NLP [5], and more specifically, in our context, the early detection of Alzheimer’s Disease from spoken language.

This paper explores the potential of text embedding models to capture informative representations of speech from patients. These representations should be powerful enough to differentiate between the speech of healthy patients and those with Alzheimer’s disease.

Our approach takes as input an audio recording of a patient and delivers a decision, either AD or non-AD. To achieve this, our approach comprises three steps (Figure 1). The initial step involves transcribing the text using the Wav2Vec 2.0 pretrained model algorithm [6]. The second step involves extracting a discriminative representation using text-embedding models. This representation then passes through the third step, which involves a binary classifier to classify the patients.

This paper focuses on the second and third steps. As several text embedding methods (primarily developed by OpenAI) and numerous AI classifiers exist, our goal is to explore the combination yielding the best performance. Our contributions are as follows:

- Regarding the text-embedding models, we have identified three models: ada-002, 3-small, and 3-large. Each model offers an embedding representation of different dimensions. As these models are pre-trained, our contribution involves investigating these models to determine which one is most suitable for the context of Alzheimer’s Disease.

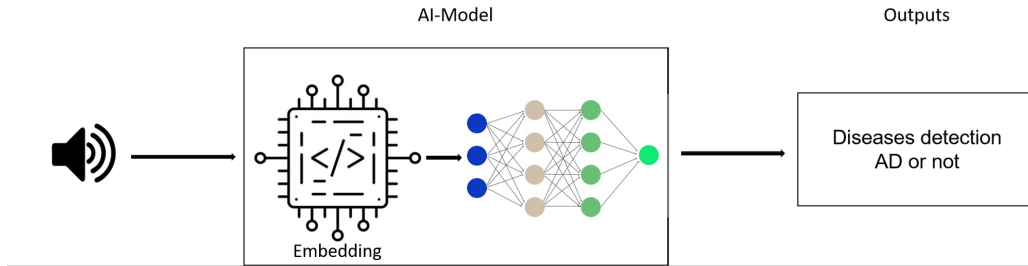


Fig. 1: Overview of the model for AD diagnosis and disease severity prediction from voice.

- Concerning the classification step, our contribution involves conducting a comparative study of AI models, DL-based models, and traditional models. This will enable us to select the most suitable classifier for our context, and to determine the most appropriate classifier for each type of embedding. Ultimately, we aim to identify the combination of text embedding and classifier that is best suited to our context.

The remainder of the paper is structured as follows. Section II provides a summary of the suggested methods for early Alzheimer’s disease detection. We describe our evaluation pipeline in Section III. In Section IV, we report our findings on the chosen dataset. Section V summarizes our findings and future prospects.

## II. STATE OF THE ART

The early detection of Alzheimer’s disease (AD) through a preclinical stage with a unified framework of speech parameter extraction and language models (LLMs) has attracted considerable interest in recent years [7]. Machine learning algorithms and computational approaches have been investigated by researchers and medical professionals to identify changes in speech. In this regard, researchers have explored various techniques for using speech as a biomarker for classifying AD [8], [9], [10].

A system for early detection of AD comprises two main stages: feature extraction and classification. In this state-of-the-art review, we focus on the extraction of features from speech in the form of audio. The techniques typically involve analyzing speech and extracting features directly from the audio or from their transcriptions.

In addition to analyzing speech data alone, researchers are exploring the fusion of multiple modalities for a more robust representation in early AD detection. By combining information from various sources such as acoustic and language features, fusion approaches aim to enhance the accuracy and reliability of detection systems [11]. Techniques like feature-level fusion and decision-level fusion are being investigated to effectively integrate data from different modalities.

In the past, acoustic and language features could only be extracted through handcrafted methods. However, recently, with the evolution of ML algorithms and especially

DL, the latter are utilized as feature extraction tools, known as DL-based embedding feature extraction.

In the following, we present an overview of the various approaches of each category proposed in the literature.

### A. Input features and embeddings

In this section, we explore feature extraction through acoustic and linguistic dimensions. Acoustic features encompass pitch and spectral content, while linguistic features include lexical and syntactic elements.

Dementia causes changes in people’s speech, observed through shifts in vocabulary and sentence construction. The linguistic aspects analyzed in this study focus on individual words and were extracted from manually transcribed databases. Individuals with AD tend to use shorter sentences and have a limited vocabulary w.r.t those without the condition. These linguistic features are expressed as proportions relative to the total number of words spoken in each instance, e.g., the percentage of adjectives used w.r.t the overall number of words uttered [12]. Words’ usage is important in understanding human language. Words can be represented as Bag-of-Words (BoW)[13] and Linguistic Inquiry and Word Count (LIWC)[13], encoding information about the words and their meanings. Other methods, like GloVe, Word2Vec [14], and TF-IDF [15], give us numerical representations of words based on how they are used and how often they appear.

On the speech signal side, there are various acoustic features that can be derived from audio signals [16]. These features encompass different aspects of the sound, such as its temporal properties, frequency composition, and amplitude variations. Among these features, we mention Mel-Frequency Cepstral Coefficients (MFCCs)[17], Linear Predictive Cepstral Coefficients (LPCC) [18], and Linear Predictive Cosine Transform coefficients (LPCT) [19].

### B. Deep learning embeddings

In the previous section, we mentioned that features are hand-crafted, whether from acoustic signals in audio or linguistically from text. Recently, models based on Deep Learning architectures have been used to extract a discriminative representation of the input.

Deep learning embedding-based approaches consist of extracting discriminative representations of data obtained

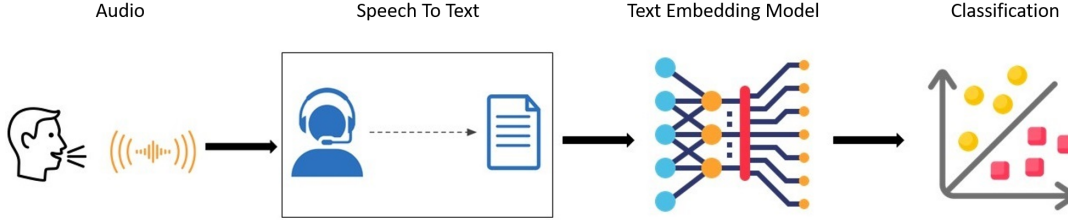


Fig. 2: Schematic showing the feature representation that are derived from speech.

through deep learning models. These embeddings effectively capture significant features or representations of the input data [20]. They are particularly useful for tasks like classification [4], clustering, and similarity computation. In various domains, different types of deep learning embeddings have been commonly employed, such as audio embedding and text embedding [21].

### C. Fusion based approaches

The fusion of acoustic and linguistic signals entails integrating acoustic and linguistic knowledge from the source data into a unified input description. In essence, this approach aims to leverage the complementary nature of audio and textual evidence of sounds to develop models capable of performing more effectively in tasks such as speech recognition, sentiment analysis, or text processing. The techniques employed for fusion may vary depending on the task and the quality of the source data [21][22]. A recently published survey [23] identified the most popular embeddings used in the AD context: BERT [24], Generative Pre-trained Transformer (GPT) [25], Transformer [26], DistilBERT [27], and RoBERTa [28]. Our approach lies in the category utilizing linguistic features while employing DL models for text embedding representation.

## III. METHODS

To investigate how speech can be used as an input to identify dementia, our approach includes two main steps: 1) Feature extraction and 2) Classification. The flowchart of our method is illustrated in Figure 2.

The input is a speech transcribed into text, followed by a feature extraction step based on a text-embedding model. Then, a ML classifier (both DL-based and non-DL-based) is applied. In this paper, our investigation focuses on finding suitable text-embedding-based features and classifiers for AD classification.

Text embeddings are numerical representations of text that capture the semantics, or meaning, of words, phrases, or entire documents. They are foundational in many NLP tasks, such as sentiment analysis and text classification. In our approach, we evaluate three text embedding models from OpenAI<sup>1</sup>: text-embedding-ada-002 (ada-002), text-

embedding-3-large (3-large), and text-embedding-3-small (3-small). Table I presents a comparison of the three text embedding models. The first model "ada-002", is a text embedding solution optimized for adaptive learning. The second model, "3-large", is an automatic large-scale embedding model capable of processing multiple languages. The third model, "3-small", is a small-scale model with good performance in resource efficiency.

### A. Classification

For the classification task, we evaluated classical machine learning models and deep neural networks using the text embedding features as input.

Regarding classical machine learning models, we considered the following classifiers: Logistic Regression (LR), Support Vector Classifier (SVM), Random Forest (RF), Bagging Classifier (Bagging), Extra Trees Classifier (ExtraTrees), XGBoost and Multilayer Perceptron (MLP).

Regarding DL-based models, we considered the Deep Neural Network (DNN) composed of sequential dense layers.

## IV. EXPERIMENTAL RESULTS

### A. Dataset

The ADReSSo Challenge offered benchmark datasets for three tasks: AD classification, MMSE score regression, and disease prognosis [29], [30]. The dataset for AD classification consisted of audio recordings from the "Cookie Theft" picture description task, taken from the Boston Diagnostic Aphasia Examination [31]. This task is commonly used to detect language disorders. Cognitively healthy individuals and AD patients recorded their descriptions of the "Cookie Theft" picture following the guidelines of the Boston Diagnostic Aphasia Examination. There were a total of 237 audio files in the dataset. In order to address potential bias, the files were split into training and test sets, with a 70/30 ratio and careful consideration of sex and age distributions. The training set had 166 instances, with 87 diagnosed as probable AD patients and 79 as healthy elderly controls. The test set consisted of 71 instances, with 35 AD patients and 36 cognitively normal individuals.

<sup>1</sup><https://platform.openai.com/docs/guides/embeddings/>

TABLE I: Classification results for the AD detection task for the three models embedding

Text embedding model	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
text-embedding-ada-002	SVM	<b>85.92</b>	<b>84.21</b>	88.89	<b>86.49</b>
	LR	77.46	76.32	80.56	78.38
	RF	74.65	70.46	86.11	77.50
	Bagging	78.88	73.33	<b>91.67</b>	81.48
	ExtraTrees	76.06	72.09	86.11	78.48
	xgboost	77.46	76.32	80.56	78.38
	MLP	76.06	74.35	80.55	77.33
	DNN	83.10	81.58	86.11	83.78
text-embedding-3-large	SVM	80.28	83.00	77.78	80.00
	LR	73.24	72.98	75.00	73.97
	RF	71.83	70.00	77.78	73.68
	Bagging	69.01	65.91	80.56	72.50
	ExtraTrees	71.83	69.05	80.56	74.36
	xgboost	73.24	72.97	75.00	73.97
	MLP	<b>84.51</b>	<b>83.78</b>	<b>86.11</b>	<b>84.93</b>
	DNN	76.06	74.35	80.55	77.33
text-embedding-3-small	SVM	77.46	75.00	<b>83.33</b>	78.95
	LR	83.10	85.29	80.56	82.86
	RF	74.65	72.50	80.56	76.31
	Bagging	78.87	76.92	<b>83.33</b>	80.00
	ExtraTrees	76.06	75.68	77.78	76.71
	xgboost	78.87	78.37	80.55	79.45
	MLP	78.87	78.38	80.56	79.45
	DNN	<b>84.51</b>	<b>90.32</b>	77.78	<b>83.80</b>

### B. Evaluation Protocol

The evaluation protocol concerns both the training and testing stages. During training, we used the grid search technique to find the best combination of hyperparameters. For each hyperparameter, a set of values is suggested, and then all combinations are evaluated.

To optimize the hyperparameters, cross-validation, consisting of dividing the training set into  $n$  folds (5 in our experiments)  $n$  was used. Each fold takes turns playing the role of the validation set, while the remaining  $(n - 1)$  folds are used for training. This process is repeated  $n$  times, performing  $n$  evaluations. The combination performance is the average of all evaluations. For each classifier, the combination of hyperparameter values yielding the best performance was retained. In the test stage, after fixing the best hyperparameters, we used, for evaluation, standard metrics: accuracy, precision, recall, and F1-Score.

### C. Results

In this section, we discuss the results of our approach on the test dataset, which involves two main steps: feature extraction using text embedding models and classification using machine learning algorithms. Three text embedding models are evaluated to extract a discriminative represen-

tation of the text. This representation then serves as the input for the ML classifiers.

The binary classification results, summarized in Table I, presents the classifications using three different embedding models: "ada-002", "3-large", and "3-small". The evaluated classifiers include Support Vector Machine, Logistic Regression, Random Forest, Bagging, ExtraTrees, XGBoost, Multi-Layer Perceptron, and Deep Neural Network.

Our DNN model consists of four dense layers with 512, 256, 128, and 2 units respectively. Each layer uses ReLU activation and is followed by a dropout layer with a dropout rate of 0.1. The model concludes with a final softmax layer.

To compare classifier performances, we consider usual metrics, namely accuracy, precision, recall, and F1 score. To visually demonstrate which embedding method performs best, the results are grouped by classifier, showcasing the three embedding methods in Figure 3.

### D. Discussion

Based on the results we have obtained, we can draw two observations. The first pertains to the ML classifier, and the second concerns the text embedding method.

Regarding the text embedding models, out of the 8 classifiers tested, in 6 cases, the "3-large" model yields the lowest performance, and only in the case of MLP does it provide the highest performance. This may be justified by

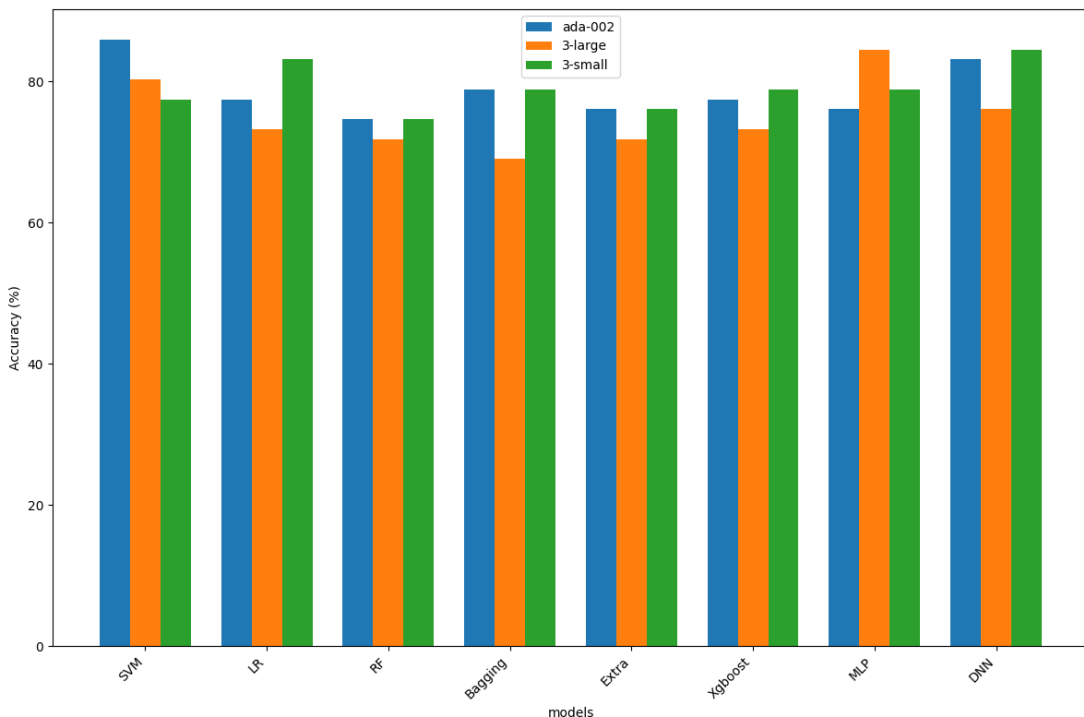


Fig. 3: Evaluation of Classifier Accuracy by Embedding Model

the fact that this model generates an embedding representation of a very high dimension (3072), and given that the dataset is small, the ML algorithms do not find enough data to learn their hyperparameters. For the other two models, "ada-002" and "3-small," the results are close in the case of Bagging, RF, XGBoost, and DNN. With the remaining models, the performance alternates.

When averaging the accuracy across all classifiers by embedding method, we find 79.04% for "3-small", and 78.6% for "ada-002". Whereas for "3-large," it is 73.84%.

A representation in a very high dimension remains a rich one, but it is essential, especially in our case with a small dataset, to apply certain techniques such as data augmentation and transfer learning.

The second observation concerns the ML classifiers, for the two best embedding models, "ada-002" and "3-small". The DNN is among the top two best models. It is the best model for "3-small", with an accuracy of 84.51%, and the second-best model for "ada-002" with 83.10%, following SVM which 85.92%. This is somewhat expected given the capabilities of DNNs demonstrated in the literature compared to traditional ML. However, it is unexpected that there is not a larger gap with traditional ML models. This can be justified by the fact that DNNs require a large quantity of data to effectively learn their parameters.

It is important to note that the last two points are based on the accuracy metric, but this analysis is almost identical for the other metrics, precision, recall, and F1-score.

## V. CONCLUSIONS

In this paper, we have investigated the capacity of text embedding models based on Large Language Models to aid in the early diagnosis of dementia. Our main approach was to use text embedding models to generate a discriminant text embedding, a vector representation of transcribed speech text, that captures the semantic meaning of the input. Subsequently, an ML classifier is applied to classify patients as either having AD or not. Three embedding models from OpenAI were investigated, along with a set of eight classifiers. The results have shown that the models "ada-002" and "3-small" are the most effective, and the DNN classifier outperforms the traditional ML classifiers.

Future work includes improving DNN performance by using techniques such as transfer learning and data augmentation. Moreover, the used text embedding models are pretrained and do not account for the specificity of our data. We plan to create a second embedding specific to our context by using DL-based models, similar to the existing models. We will also explore a multimodal approach by considering other modalities such as handwriting analysis-based AD assessment [32], [33], [34], [35], [36], in addition to speech's [10], [37].

## REFERENCES

- [1] Karim Haddada, Mohamed Ibn Khedher, and Olfa Jemai. Comparative study of deep learning architectures for early alzheimer detection. In *International Conference on Cyberworlds, CW 2023, Sousse, Tunisia, October 3-5, 2023*, pages 185–192. IEEE, 2023.

- [2] Alzheimer's Association. 2019 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 15(3):321–387, 2019.
- [3] S. Zargarbashi and B. Babaali. A multi-modal feature embedding approach to diagnose alzheimer disease from spoken language. *arXiv preprint arXiv:1910.00330*, 2019.
- [4] Rahul Pappagari, Jaeyoung Cho, Siddharth Joshi, Luis Moro-Velázquez, Peter Zelasko, Jesús Villalba, and Najim Dehak. Automatic detection and assessment of alzheimer disease using speech and language technologies in low-resource scenarios. In *Interspeech*, volume 2021, pages 3825–3829, August 2021.
- [5] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 2023.
- [6] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33:12449–12460, 2020.
- [7] Mengke Shi, Gary Cheung, and Seyed Reza Shahamiri. Speech and language processing with deep learning for dementia diagnosis: A systematic review. *Psychiatry Research*, 329:115538, November 2023.
- [8] A Balagopalan and J Novikova. Comparing acoustic-based approaches for alzheimer's disease detection. *arxiv 2021*. *arXiv preprint arXiv:2106.01555*.
- [9] Jun Chen, Jieping Ye, Fengyi Tang, and Jiayu Zhou. Automatic detection of alzheimer's disease using spontaneous speech only. In *Interspeech*, volume 2021, page 3830. NIH Public Access, 2021.
- [10] S. Mirzaei, M. El Yacoubi, S. Garcia-Salicetti, J. Boudy, C. Kahindo, V. Cristancho-Lacroix, H. Kerhervé, and A.-S. Rigaud. Two-stage feature selection of voice parameters for early alzheimer's disease prediction. *IRBM*, 39(6):430–435, December 2018.
- [11] Felix Agbavor and Hualou Liang. Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health*, 1(12):e0000168, December 2022.
- [12] M. González Atienza, J. A. González López, and A. M. Peinado. An automatic system for dementia detection using acoustic and linguistic features. In *ISCA*, January 2021.
- [13] Ahmed H Alkenani, Yuefeng Li, Yue Xu, and Qing Zhang. Predicting alzheimer's disease from spoken and written language using fusion-based stacked generalization. *Journal of Biomedical Informatics*, 118:103803, 2021.
- [14] S Joshua Johnson, M Ramakrishna Murty, and I Navakanth. A detailed review on word embedding techniques with emphasis on word2vec. *Multimedia Tools and Applications*, pages 1–29, 2023.
- [15] Pranav Mahajan and Veeky Baths. Acoustic and language based deep learning approaches for alzheimer's dementia detection from spontaneous speech. *Frontiers in Aging Neuroscience*, 13:623607, 2021.
- [16] James Glass et al. Classifying alzheimer's disease using audio and text-based representations of speech. *Frontiers in Psychology*, 11:624137, 2021.
- [17] Suprava Patnaik. Speech emotion recognition by using complex mfcc and deep sequential model. *Multimedia Tools and Applications*, 82(8):11897–11922, 2023.
- [18] Baveet Singh Hora, S Uthiraa, and Hemant A Patil. Linear frequency residual cepstral coefficients for speech emotion recognition. In *International Conference on Speech and Computer*, pages 116–129. Springer, 2023.
- [19] J Ancilin and A Milton. Improved speech emotion recognition with mel frequency magnitude coefficient. *Applied Acoustics*, 179:108046, 2021.
- [20] Bidisha Sharma, Maulik Madhavi, and Haizhou Li. Leveraging acoustic and linguistic embeddings from pretrained speech and language models for intent classification. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, June 2021.
- [21] Morteza Rohanian, Julian Hough, and Matthew Purver. Multi-modal fusion with gating using audio, lexical and disfluency features for alzheimer's dementia recognition from spontaneous speech. *arXiv preprint arXiv:2106.09668*, October 2021.
- [22] Edward L Campbell, Laura Docío-Fernández, Javier Jiménez Raboso, and Carmen García-Mateo. Alzheimer's dementia detection from audio and text modalities. *arXiv preprint arXiv:2008.04617*, 2020.
- [23] Mengke Shi, Gary Cheung, and Seyed Reza Shahamiri. Speech and language processing with deep learning for dementia diagnosis: A systematic review. *Psychiatry Research*, page 115538, 2023.
- [24] Aparna Balagopalan, Benjamin Eyre, Jessica Robin, Frank Rudzicz, and Jekaterina Novikova. Comparing pre-trained and feature-based models for prediction of alzheimer's disease based on speech. *Frontiers in Aging Neuroscience*, 13, April 2021.
- [25] Junghyun Koo, Jie Hwan Lee, Jaewoo Pyo, Yujin Jo, and Kyogu Lee. Exploiting multi-modal features from pre-trained networks for alzheimer's dementia recognition. In *Interspeech 2020*. ISCA, October 2020.
- [26] Ning Liu, Zhenming Yuan, and Qingfeng Tang. Improving alzheimer's disease detection for speech based on feature purification network. *Frontiers in Public Health*, 9, March 2022.
- [27] Sandeep Kumar Pandey, Hanumant Singh Shekhawat, Shalendar Bhasin, Ravi Jasuja, and SRM Prasanna. Alzheimer's dementia recognition using multimodal fusion of speech and text embeddings. In *International conference on intelligent human computer interaction*, pages 718–728. Springer, 2021.
- [28] Erik Edwards, Charles Dognin, Bajjibabu Bollepalli, and Manesh Singh. Multiscale system for alzheimer's dementia recognition through spontaneous speech. In *Interspeech 2020*. ISCA, October.
- [29] Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney. Detecting cognitive decline using speech only: The addresso challenge. *arXiv preprint arXiv:2104.09356*, 2021.
- [30] Zafi Sherhan Syed, Muhammad Shehram Shah Syed, Margaret Lech, and Elena Pirogova. Tackling the addresso challenge 2021: The muet-rmit system for alzheimer's dementia recognition from spontaneous speech. In *Interspeech*, pages 3815–3819, 2021.
- [31] Mandy WM Fong, Ryan Van Patten, and Robert P Fucetola. The factor structure of the boston diagnostic aphasia examination. *Journal of the International Neuropsychological Society*, 25(7):772–776, 2019.
- [32] Christian Kahindo, Mounim El Yacoubi, Sonia Garcia-Salicetti, Anne-Sophie Rigaud, and Victoria Cristancho-Lacroix. Characterizing early-stage alzheimer through spatiotemporal dynamics of handwriting. *IEEE Signal Processing Letters*, PP:1–1, 01 2018.
- [33] Mounim A. El-Yacoubi, Sonia Garcia-Salicetti, Christian Kahindo, Anne-Sophie Rigaud, and Victoria Cristancho-Lacroix. From aging to early-stage alzheimer's: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning. *Pattern Recognition*, 86, 2019.
- [34] Nickson Mwamsojo, Frederic Lehmann, Mounim A. El-Yacoubi, Kamel Merghem, Yann Frignac, Badr-Eddine Benkelfat, and Anne-Sophie Rigaud. Reservoir computing for early stage alzheimer's disease detection. *IEEE Access*, 10:59821–59831, 2022.
- [35] Quang Dao, Mounim A. El-Yacoubi, and Anne-Sophie Rigaud. Detection of alzheimer disease on online handwriting using 1d convolutional neural network. *IEEE Access*, 11:2148–2155, 2023.
- [36] Jana Sweidan, Mounim A. El-Yacoubi, and Anne-Sophie Rigaud. Explainability of cnn-based alzheimer's disease detection from online handwriting. April 2024.
- [37] Holger Fröhlich, Noémi Bontridder, Dijana Petrovska-Delacràta, Enrico Glaab, Felix Kluge, Mounim El Yacoubi, Mayca Marán Valero, Jean-Christophe Corvol, Bjoern Eskofier, Jean-Marc Van Gyseghem, Stéphane Lehericy, Jürgen Winkler, and Jochen Klucken. Leveraging the potential of digital technology for better individualized treatment of parkinson's disease. *Frontiers in Neurology*, 13, February 2022.