Power of Words: A Review of Approaches Utilizing Speech and Text for Alzheimer's Disease Detection

Pranab Hazra[1], Sayantanny Podder[2], Soumi Das[3], Tanmoy Chakraborty[4], Rajashri Nandi[5], Ashis Kumar Dhara[6], Tushar Kanti Bera[7]

[1]PhD Scholar, Electrical Engineering, National Institute of Technology, Durgapur, India and Assistant Professor, Electronics and Communication Engineering, Narula Institute of Technology, India.
[2,3,4 and 5] Student, Electronics and Communication Engineering, Narula Institute of Technology, India
[6 and 7]Assistant Professor, Electrical Engineering, National Institute of Technology, Durgapur, India

ABSTRACT

Alzheimer's disease (AD) presents a profound challenge to healthcare systems worldwide, emphasizing the critical need for reliable and efficient diagnostic tools. In recent years, there has been a burgeoning interest in technological advancements, particularly in speech and text analysis, aimed at facilitating the early detection and monitoring of AD. This comprehensive review synthesizes the latest scientific findings and technological developments in the field of speech and text-based AD detection. The review delves into various methodologies employed in speech and text analysis, spanning from traditional neuropsychological assessments to cutting-edge machine learning algorithms and natural language processing techniques. Commencing with an overview of AD's neurobiological mechanisms and the urgency of early diagnosis, the review navigates through the diverse landscape of methodologies employed in speech and text analysis. It highlights potential biomarkers embedded within speech and language patterns that may serve as early indicators of cognitive impairment associated with AD. Furthermore, it examines the efficacy of different modalities, including spontaneous speech, conversational speech, and linguistic features extracted from written text, in distinguishing individuals with AD from healthy controls or those with other forms of dementia. The review also addresses challenges and limitations faced by existing methodologies, such as variability in linguistic expression and the need for large-scale validation studies.

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline, affecting memory, thinking skills, and eventually daily functioning. It represents the most common cause of dementia among older adults, affecting millions worldwide. Early detection of Alzheimer's disease is paramount for several reasons. Firstly, it allows for early intervention and treatment, potentially slowing disease progression and enhancing patients' quality of life. Secondly, it enables individuals and families to plan for the future and make informed decisions regarding caregiving and support services. Additionally, early detection facilitates patient inclusion in clinical trials aimed at developing new treatments and interventions [1]. However, traditional diagnostic methods, relying on clinical assessments, neuropsychological testing, and brain imaging techniques, can be time-consuming, costly, and inconclusive, particularly in the disease's early stages. Furthermore, there is a lack of standardization in speech and text analysis methods for Alzheimer's disease detection, hindering comparison across studies and limiting research reproducibility and generalizability. The manuscript aims to comprehensively review and synthesize existing literature on speech and text analysis techniques for Alzheimer's disease detection [2]. It seeks to provide a deep understanding of current methodologies and technologies in automated speech and text analysis, analyzing distinctive patterns exhibited by individuals with Alzheimer's compared to healthy individuals. Additionally, the manuscript aims to critically assess the diagnostic accuracy and reliability of various analysis approaches, while addressing challenges, limitations, and ethical considerations. It explores potential implications for early detection, monitoring, and management of Alzheimer's in clinical practice, highlighting opportunities for interdisciplinary collaboration. Lastly, it identifies gaps in current knowledge and proposes future research directions to advance diagnostic tools and interventions for Alzheimer's disease [3].In recent years, there has been growing interest in leveraging technological advancements, particularly in speech and text analysis, for early Alzheimer's disease detection. Speech and language impairments often precede other cognitive deficits, making analysis of speech and text patterns a promising approach for identifying at-risk individuals. Studies have shown distinct patterns in individuals with Alzheimer's disease, including changes in syntax, word-finding difficulties, and alterations in acoustic features. Advances in machine learning and natural language processing have enabled the development of automated tools capable of detecting subtle changes associated with Alzheimer's disease, revolutionizing early detection and monitoring.

The significance of this study extends beyond academia, impacting the broader Alzheimer's disease community and society. By advancing our understanding of speech and text analysis in Alzheimer's disease detection, it lays the foundation for more accurate, reliable, and accessible diagnostic tools and interventions [4]. Ultimately, this has the potential to improve patient outcomes, enhance quality of life, and reduce societal and economic burdens associated with Alzheimer's disease. Thus, the manuscript holds promise for driving positive change in Alzheimer's disease detection and beyond.

II. MATERIALS AND METHOD

There are many methods which are required for detecting or diagnosing the disease. The speech will pause when it is wrong or sometimes the wrong or meaningless words are replaced with the right ones. The wrong texts were also detected. Speech and text can also become a therapy for the patients. Through interaction in the early stage patients can recognize their memories. Till now specialists and engineers are trying to find out more paths for detecting the symptoms and therapies for their recoveries. Speech and text are responsible parameters for classifying Alzheimer’s and Non-Alzheimer’s patients as shown in Fig.1. There are two most important features in speech and text which is the indication of cognitive decline associated with Alzheimer's disease.

a) Linguistic and Acoustic Feature

Linguistic features play a vital role in indicating cognitive decline associated with Alzheimer's disease by reflecting changes in language abilities that occur as the disease progresses [5]. Individuals with Alzheimer's often exhibit alterations in vocabulary usage, syntactic complexity, and discourse coherence compared to healthy controls. These changes manifest as reduced lexical diversity, increased use of generic words, simplified grammatical structures, and impaired narrative coherence in both spoken and written language. Moreover, difficulties in word retrieval, semantic deficits, and pragmatic impairments further contribute to linguistic abnormalities observed in individuals with Alzheimer's disease.

Acoustic features are instrumental in detecting cognitive decline associated with Alzheimer's disease as they capture subtle changes in speech production characteristics that reflect underlying neurological changes. Individuals with Alzheimer's often exhibit alterations in prosody, including flattened intonation, reduced speech rhythm, and increased frequency of pauses and hesitations. These acoustic abnormalities are indicative of disruptions in speech planning, coordination, and execution, which are commonly observed in individuals with cognitive impairment. Moreover, changes in pitch variation, speech rate, and articulation precision further contribute to the acoustic profile of Alzheimer's disease [6]. By analyzing linguistic features, researchers can identify patterns that are characteristic of cognitive decline the acoustic features that are extracted from speech signals that can develop sensitive and specific algorithms for the early detection and monitoring of cognitive decline, facilitating timely interventions and improved patient outcomes.

In the intricate tapestry of human communication, the linguistic and acoustic features serve as delicate threads, weaving a narrative that reflects the intricate workings of the mind [7]. However, when Alzheimer's disease casts its shadow, this tapestry becomes frayed, its patterns distorted. Linguistic features, once vibrant and diverse, begin to lose their lustre, as individuals with Alzheimer's disease often exhibit a dwindling lexical repertoire, marked by repetitive use of generic terms and a decline in semantic richness. Syntactic structures, once robust and intricate, become simplified and fragmented, reflecting the underlying cognitive impairment. Additionally, discourse coherence wanes, with narratives losing their fluidity and logical progression, replaced by disjointed utterances and tangential digressions. Meanwhile, the acoustic landscape undergoes its transformation, as the rhythm of speech slows, punctuated by elongated pauses and hesitations. Intonation, once dynamic and expressive, becomes flattened, devoid of its former vitality. Variations in pitch and intensity lose their nuance, rendering speech monotonous and lacking in modulation. Articulation, once precise and clear, falters, giving way to slurred speech and phonemic distortions. These acoustic aberrations serve as poignant markers of cognitive decline, a sombre symphony of impairment echoing through the spoken word. In contrast, healthy controls paint a different picture, their linguistic tapestry vibrant, adorned with a kaleidoscope of words and syntactic structures. Conversations flow effortlessly, guided by a keen sense of coherence and cohesion. Acoustic melodies dance through the air, punctuated by the rhythmic cadence of speech and the melodious rise and fall of intonation. Variations in pitch and intensity add depth and texture to the spoken word, while articulation remains crisp and precise, reflecting the seamless coordination of cognitive processes.

b) Speech Task

speech tasks were performed over the phone by the four well researchers. All the participants are instructed to perform the speech task alone in a silent room so that no other interaction will happen. To conduct three speech tasks 10 mins and one phone call are needed. All the responses of the participants were recorded as sound files. The interview task, repetition task and recall task are taken [8]. In the low memory load condition an interview session was conducted, consisting of five different questions about daily life activities and personal information. The repetition task was administered the moderate memory loss condition took place. At first researcher read a sentence and the participants are instructed to remember this line and repeat this line precisely until they can. If this person is unable to say the sentence properly then the researcher will repeat these lines in the high memory load condition the recall tasks were repeated and the sentences will increase to a paragraph. As same as the previous way they were asked to recall the paragraph. And all have to recall about many details. The researcher can also describe a story and the participant has to remember the story as well as the main character of the story and the main event [9]. But if the details seem insufficient then the researcher encourages the person to elaborate the paragraph up to two times.

III. COMPREHENSIVE ANALYSIS BY PEERS

The table summarizes a range of research studies focused on utilizing speech and text analysis for Alzheimer's disease detection. Each study investigates different datasets and employs diverse feature extraction methods and classification techniques [10]. For instance, studies such as those by Davuluri et al. and Flavio Bertini et al. utilize Mel Spectrogram features from the ADReSS and Pitt Corpus datasets, achieving high accuracies of 95.83% and 93.30% respectively [21]. Randa Ben Ammar et al. explore the Dementia Bank dataset, examining temporal and verbal disfluencies, lexico-syntactic diversity, word, utterance rate, and MMSE score, achieving an accuracy of 91%. Other studies leverage advanced techniques like deep learning models such as CNN-LSTM and ResNet for feature analysis, achieving varying accuracies. Notably, these studies highlight the importance of both acoustic and linguistic feature sets, with some employing sophisticated classification methods like SVM, RF, and deep learning architectures to achieve accurate Alzheimer's disease detection rates, demonstrating the potential of speech and text analysis in early diagnosis and monitoring of cognitive decline [11].

Table I Recent Developments for Alzheimer's Detection Using Speech And Text

SL NO.	AUTHORS	DATASETS	FEATURES EXTRACTED		CLASSIFICATION METHOD	RESULT
			ACOUSTIC FEATURE SET	LINGUISTIC FEATURE SET		
1	Raghavendra Pappagari et al.[28]	ADReSS	Voxceleb, NIST SRE4, VGGish, eGeMAPS	BERT, IARPA,ASR	MMSE,ASR	Accuracy:81.3%
2	Liu et al.[19]	ADReSS	MFCC	-	RNN,CNN and Attention pooling	The model achieved an accuracy of 82.59%, with a precision of 85.29%, recall of 81.46%, and an F1 score of 82.94%
3	Rui He et al.[12]	Dementia detection			MCI,SCD	Highest score:9 Model performance was significantly different for linguistic domains (p < .001), and speech versus text (p = .043)
4	Davuluri et al.[10]	ADReSS	Mel Spectrogram	-	VGG-16	It achieved Accuracyof 95.83%
5	Jessica Robin et al.[35]	MMSE	CDR-SB ADAS-Cog13	-	MMSE,CDR	It achieved Accuracy of 95%

6	Flavio Bertini et al.[4]	Pitt Corpus	Mel Spectrogram	-	Auto encoder and MLP with data augmentation	The model's metrics are as follows: Accuracy - 93.30%, Precision - 90.7%, Recall - 86.5%, F1 Score - 88.5%.
7	Ulla Petti et al.[30]	Pitt Corpus	-	TTR, MATTR_20,LI verbs	Use of unique words and pronounce related matter, ESRC, HC participant	
9	Meghanani et al.[25]	ADReSS	-	Bigrams,Trigrams,4-grams,5-grams	fastText, CCN	The fastText model, incorporating both bigrams and trigrams, attained an initial accuracy of 81.4% alongside a score of 86.5%. Following fine-tuning, its accuracy further enhanced to 83.33%.
10	Ulla Petti et al.[31]	CCS	WER	QUADAS-2 PROBAST	ASR,AZTIAHO	64% of the studies reported the age and gender of the participants with 40% of the datasets being gender-balanced, and 45% of the studies reported the education level of the participants
11	Li et al.[17]	ADReSS	ComParE, X-vector	TF-IDF vector, Linguistic feature Sets, BERT embeddings	LDA. SVM and AT-LSTM	Accuracy:67%from X-vector by AT-LSTM &88% on BERT by SVM
12	Ziyun Cui et al.[9]	DAIC-WOZ Hu-BERT	WavLM	-	FI score, RMSE	score of 0.928 on the ADReSSo test set
13	Nasreen et al.[26]	CCC	Interactional features	Disfluency features	SVM,LR and MLP	On the SVM classifier, leveraging a combination of features, the following performance metrics were achieved: Accuracy - 90%, Recall - 90%, Precision - 90%, AUC - 89%, and F1 Score - 90%.
14	Mona Roxana Botezatu et al.[5]	MMSE Dementia severity	-	LCC, LSC	MDRS, BDS	LCC score significantly predicted MMSE score 44% of the variance, MMSE scores, 41% MDRS scores :61%

15	Matej Martine et al.[22]	ADReSS	eGeMAPS	GloVe embeddings	Combination of K-means clustering(with k=30) & RF	Accuracy:93.75%
16	Hongmin Cai et al.[8]	Dementia bank Pitt database	AD-GNN WavLM	CLIPPO TTS	CNN GPT	In comparison, the text modality achieved an accuracy of 93.75%, notably surpassing the audio modality's 85.04% accuracy. The text modality's accuracy of 0.8460 significantly outperformed the audio modality's 0.7714 accuracy
17	Arian Shamei et al.[37]	Pitt Corpus	LRT, LMER	-	VSA	AD group had a combined VSA of 5.4 vs 5.96 for HC.
18	Meghanani et al.[23]	ADReSS	log mel spectrogram MFCC	-	CNN-LSTM, ResNet 18-LSTM,Pblstm-CNN	Accuracy:64.58% by MFCC using CNN-LSTM &62.5% by log mel spectrogram using ResNet 18-LSTM
19	Aliment et al.[1]	ADReSS, ILSE interviews	Voice activity detection (VAD), ComParE, i-vectors, ECAPA-TDNN	LIWC,PoS, Perplexity , PoS Perplexity	GMM, LDA, SVM	Using SVM for ADReSS, the UAR for acoustic features stands at 66.7%, while for linguistic features, it achieves 77.1%. For ILSE, the UAR on acoustic features is 86%, while on linguistic features, it reaches 83.8%.
20	MumuneMerveParlak[29]	SLP	Cognitive communication disorder	-	Interpretation Criticism, Set of questions	82.1% had never given therapy . 89% of SLPs did not know about AD. 92.4% stated that patients in the mild stage, 89% of patients(intermediate stage), and 45.5% of patients advanced stage could benefit from therapy.
21	Perez et al,[32]	Pitt corpus	VAD duration features, pleasure arousal dominance(PAD), Hamonicity to Noise ratio	Phonemic features	eXtreme Gradient Boosting, ForestNet	ForestNet yielded a UAR of 79% when applied to a combination of duration, phonemic, and acoustic features.

a) Reduction of Vowel Space In Alzheimer's Disease

Alzheimer's disease is an insidious neurodegenerative disease with a loss of cognitive and bodily function [12]. Motor skills deteriorate as the disease progresses for example, the arm movements of education are accurate slower and discontinuous relative to those of controls. Assessments of fine motor skills, such as finger-tapping test, exhibit that AD results in impairments to the timing and motor execution of finger movements [13]. Research demonstrates that AD results in measurable motor storage during preclinical stages of the disease, therefore, assessments of speech motor skills via acoustic analysis serve as a non-invasive method for easy diagnosis and monitoring of disease progression, with previous research identifying AD-related changes to prosody, voice quality and segment timing [14]. Assessments of vowel space in DD patients have shown that vowel reduction effects are most salient for natural spontaneous speech, and less salient for isolated words or vowels. Therefore, the VSA of AD patients may be reduced outside of isolated word contexts. To evaluate the finding of XIU ET AL The availability of large public corpuses holding speeches which was allowed by AD speakers. for isolated words extends to continuous speech [15]. For this purpose, we compare the VSA(Vowel Speech Area) of AD speakers across large corpora spanning 100 AD speakers:
(1) the English language Pitt Corpus, which contains spontaneous speech as elicited via a picture description task, and (2) the Spanish language Ivanova Corpus, which contains read speech. They represent the largest publicly available data sets of continuous speech from AD speakers. With these data sets, we conduct one experiment to evaluate the effect of AD on VSA : (1) The Euclidean distance of vowel tokens is calculated in a vowel set or quadrilateral triangular from the specific speaker's centroids. As all the movement trajectories in AD are observed to be slower, reduced, and discontinuities for both fine motor skills and gross, we hypothesize that impairments to fine motor skills of AD should be reflected in the time of speech of AD speakers about the reduction of the VSA [16]. This should manifest as shortened movement trajectories for individual vowel targets that generalize to a smaller working VSA. We note that while did not observe VSA reduction in non-continuous speech, both datasets used in the present study employ some form of continuous speech. For individuals, All the shortened movements should be manifested as trajectories vowel targets so that they all can generalize to a smaller working VSA [17]. It is noted that after not observing the VSA reduction in non-continuous speech, both datasets used in the present study employ especially some of the forms of continuous speech.

(i) English Pitt Corpus:
The English language Pitt Corpus contains speech recordings of 178 A9 patients with up to five recordings obtained longitudinally over 20 years. from all available longitudinal sessions We used different audio files for all AD participants when they performed the cookie-theft picture description task, every audio file provides 1 minute. All audio files within the Corpus were in WAV format with a 44.1kHz sampling rate and a 16-bit depth encoding.
(ii) Spanish Ivanova Corpus:
The Spanish language Ivanova speech from 47 AD participants. Each participant performed a reading of the first paragraph of the classic novel Don Quixote, and it's an approximate reading time of 1 minute. Our analysis is limited to AD participants for which WAV files are available. And each file was provided with a 44.1kHz. sampling rate and 16-bit depth encoding. In this study, we found reduced VSA in English spontaneous speech, which provides support for the prediction that impairments to fine motor skills in AD should manifest in reductions to the movement trajectories of vowel articulation. The absence of VSA is deducted in Spanish read speech and Mandarin isolated words suggests that VSA reduction in AD may be influenced by task or language.

b) Language Tests And Depersonalization
Analysing changes in speech in AD is often done using different cognitive and language tests [18]. Communication is often linked to personhood, and non-standard communication can put an individual at risk of being treated as a non-person, and experiencing damaged self-esteem and discouragement [19]. It is assumed that the absence of communication means that the internal mental processes have stopped. Whilst some of the earlier literature suggested that the actions of individuals with AD are not guided by meaning, others argue that these individuals relate to their behaviour. Using insensitive psychometric tests highlights the communicative deficits and can create an image of a wandering, inappropriate, labile, and confused person, contributing to not seeing the disease as a separate entity from the person [20]. To avoid depersonalization, researchers have expressed treating individuals with AD as semiotic subjects, presuming their actions to be guided by meaningful reasons and looking at any communications deficits as similar to unfamiliar conventions from another culture, as well as focusing research on the aspects of communications that are preserved in addition to the ones that are lost. For decades Declaration of all the signs of AD has been a progressive debate. Some researchers state that the most pressing issue at the moment is disclosing the results of biomarker research in asymptomatic persons, and while researchers are not currently obligated to share the biomarker results due to the uncertainty of their clinical utility, the current consensus is increasingly learning towards informing the participants of their

diagnosis. Introducing AI-aided AD detection has raised additional questions about accuracy and predictive power which also affect the communications on research outcomes [22]. Consistently with the literature on using AI for detecting risks of psychiatric conditions publishing books, giving press conferences or TV interviews are aware of the recording taking place, and can reasonably expect these performances to be observed and made public, they may not be expecting their language data to be used for investigating and potentially making public early signs of AD and their language decline [23]. Additionally, storing and sharing voice data can pose or risk of identification, especially when the speaker is known to the listener. The study of speech and language Pathologists (SLP) is one of the fields of cognitive-communication disorders. Dementia is defined as a chronic and usually progressive syndrome characterised by a deteriorating in cognitive function that exceeds what is expected to occur during the typical again process which is a cognitive communication disorder [24]. Dementia is the disease which is occupying the seventh position as a cause of death among all the diseases. And one of the leading causes of disability and addiction among older people worldwide. Alzheimer's disease(AD) constitutes 60-70% of dementia cases. It is a progressive neurodegenerative disease, that is classically characterised by a gradual deterioration of memory, language, and other cognitive domains. SLPs(speech and language Pathologists) who work with AD should work from the earliest stage possible to support their cognitive communication. In focusing on sub-branches An SLP can provide all the ultimate therapy of cognition separately, or it will be aimed at all types of general cognitive communication. To monitor the effectiveness of therapies, patients should also complete pre and post-therapy evaluations for their quality of life depression, and activities of daily living [26]. Immediately starting therapy after a diagnosis, especially in patient groups with a progressive cognitive impairment such as, AD is of great importance, in terms of reducing the rate of progression and protecting cognitive functions. There are non-pharmacological therapy approaches that can be applied to AD, and the effects of these therapies on different functions have been demonstrated by many studies.

IV. MODEL USED FOR ALZHEIMER’S DISEASE DETECTION USING SPEECH AND TEXT

In the realm of detecting Alzheimer's disease through speech and text data, machine learning models serve as the compass guiding researchers through the intricate terrain of data analysis. Among these models, some stand out as stalwart guardians, adept at discerning the subtle nuances that betray cognitive decline. Support Vector Machines (SVMs), with their ability to carve out hyperplanes in high-dimensional space, offer a robust framework for classification tasks, effectively separating individuals with Alzheimer's disease from healthy controls based on extracted linguistic and acoustic features. Random Forests, with their ensemble of decision trees, thrive in the dense forest of data, capturing complex relationships between features and outcomes with remarkable accuracy. Meanwhile, Deep Learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), delve into the depths of raw speech and text data, extracting intricate patterns and temporal dependencies that elude traditional algorithms [25]. Like master artisans, these models sculpt insights from the raw clay of data, revealing hidden gems of knowledge that hold the key to early detection and intervention. As researchers navigate the vast expanse of machine learning algorithms, they harness the power of these models to illuminate the path toward a future where Alzheimer's disease is detected swiftly and accurately, offering hope and solace to those touched by its shadow [15].

In this paper, we have discussed four methods that are used to assess and compare the prediction performance for speech tasks to diagnose Alzheimer’s disease. These methods are (i) the GNN method(ii) the Data augmenter method(iii) the Multimodal method (iv) the CLIPPO-like method.

(i) GNN Method: In this method a lightweight globe best classification model for Alzheimer’s disease diagnosis using patients’ speech. In this method text embedding of the patient’s speech with a pre-trained language model, this model name is BERT, taking advantage of the pre-trained language model that has been trend on a massive text Corpus using a large-scale model architecture [26]. After this tape in the model constructs the graph representation of the embedding and utilizes a glass neutral network or GNN to learn discriminative features for the final disease classification.

(ii) Data Augmenter Method: Data augmentation is a method that can expand the variety used to train models by generating modified versions of existing data. In this method, we have a few processes.

(1) Synonym representation: In this method, we use a synonym dictionary such as what is said to replace words in the original text with their synonyms.
(2) Counterfitting embedding replacement: In the embedding refers to replacing words in a sentence with other words that are close to them in the embedding space. In the counterfitting embedding method distance between synonyms increases the distance between antonyms.
(3) Random sentence Deletion: In this method, it removes random sentences from the original text. Throughout this process from the incomplete data to make accurate predictions.
(4) Augmentation with GPT model: by using this method we need to get help from chat GPT. By using appropriate prompt chat GPT can rephrase the patient speech transcript data as instructed and does effectively increase the size of the training set [27].
(iii) Multimodal Method: it is the method where audio data is used to treat Alzheimer's disease patients. For instance changes in speech patterns such as pace tone and rhythm especially unsmooth speech such as stuttering and pauses are often early indicators of cognitive decline in Alzheimer's disease patients. These features can be extracted from audio data but our lost in text transcriptions [28]. This work has been done by wav LM a universal pretrained speech model developed by Microsoft. The model initially applies a random transformation to the input speed signal such as mixing and adding noise to enhance the models' robustness.
(iv) CLIPPO-LIKE Method: CLIPPO or clip pixels only is a method where pictures or images are used to treat Alzheimer's patients. It is a pixel-best multimodal model that can understand both images and alt-text simultaneously without requiring text encoders or tokenizers. Its approach is to render alt text as images and then encode both images and text using the same architecture [29]. It uses a contrastive learning laws function to make embeddings of matching images and all text as close as possible and all other image and all text embeddings as far apart as possible. It is a very useful method for the treatment of Alzheimer's dementia [29].
The most common signs of AD are disorientation, memory decline, confusion and behavioural changes. This symptom is to loss of impedance and having a clear impact on patients are their close ones.
Two typical signs of the disease are, memory and cognitive decline, as well as language impairment,which is also common, as this is related to cognitive and memory-related problems and neurodegenerative processes. In this respect, technologies can deliver new precision and medical tools that will give objective quantitative analysis and reliable proof and circular for a faster diagnosis [30].
The literature suggests some common signs in this patient which are related to particular aspects such as communication and any word retrieval deficits such as progressive logo panic as anomic aphasia [31].

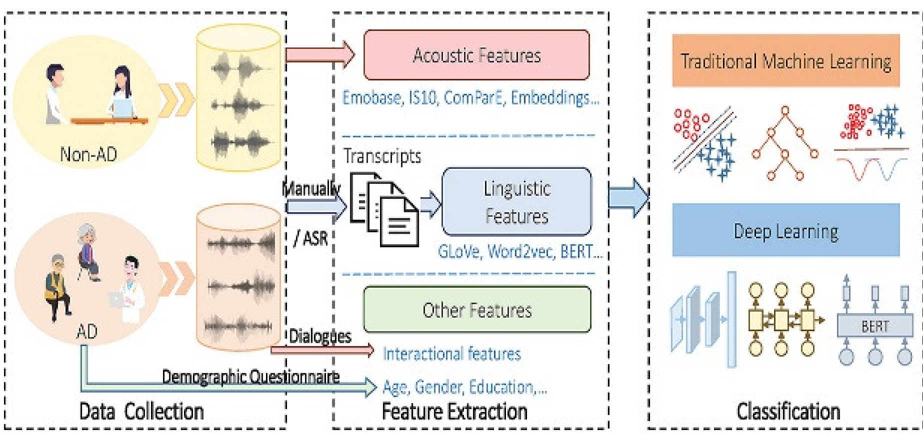

Fig.1. Process flow for classifying Alzheimer's and Non-Alzheimer's patients

On the other hand, in characterizing the loss of verbal fluency pause and silence-related features are also allowed.
The use of verbal filters such as - /um:/, or /eh:/ or by using the description of one word is also common. In more advanced stages, suffering, repetition of ideas and questions or difficulty forming simple sentences become frequent.

However, after much research, numerous teams proposed varieties of approaches to detect this AD. Automatically predict mini mental status evaluation (MME)^2 in a datasheet containing speech and manual transcription from 78 AD patients and 78 sex/age-matched control.

- Corpus 1: the final data set consisted of the transcripts of 135 public that were recorded over a period starting from 37 years before the diagnosis to 2 years after the diagnosis.
- Corpus 2: A total of 405 manually transcripted interviews were recorded over a period starting from 46 years before diagnosis and up to 13 years after the diagnosis.

In the field of natural language processing (NLP), certain models emerge as celestial navigators, guiding researchers through the cosmos of linguistic data in the quest to detect Alzheimer's disease. Among these models, recurrent neural networks (RNNs) stand as beacons of temporal understanding, their neural pathways tracing the intricate rhythms and patterns woven into the fabric of speech and text [32]. Long Short-Term Memory (LSTM) networks, a species of RNNs, possess a memory as deep as the ocean, capable of retaining contextual information over extended sequences, unravelling the chronological tapestry of language with finesse. Meanwhile, transformers, with their attention mechanisms akin to a magnifying glass on a treasure map, focus on the salient features of text, uncovering subtle semantic nuances and linguistic anomalies indicative of cognitive decline. As researchers navigate the celestial seas of NLP, they harness the power of these models to chart a course toward a future where Alzheimer's disease is detected with sensitivity and precision, illuminating the darkest corners of the human condition with the light of understanding and compassion.

V. ETHICAL IMPLICATIONS OF USING SPEECH AND TEXT DATA FOR ALZHEIMER'S DETECTION

Collecting and analyzing sensitive personal data, such as speech samples and textual information, raises concerns regarding privacy and informed consent. Researchers must ensure that participants fully understand the purpose of data collection, how their information will be used, and the potential risks and benefits involved. Safeguarding data security and confidentiality is paramount to protect participants' privacy from unauthorized access or misuse. Additionally, there are concerns related to bias and fairness in algorithm development, as detection models should be inclusive and avoid reinforcing existing biases based on demographic or cultural factors [33]. Transparency in data processing and interpretation is crucial for building trust with participants and stakeholders. Furthermore, researchers must address potential stigmatization and discrimination against individuals identified as at risk or diagnosed with Alzheimer's disease, promoting supportive and non-judgmental approaches to diagnosis and care. Ultimately, research in this area should prioritize ethical principles of beneficence, non-maleficence, equitable access, and regulatory compliance to ensure responsible and equitable use of speech and text data for advancing Alzheimer's detection and healthcare practices [34].

From numerous studies, it is observed that language changes can act as a biomarker of cognitive decline. They can able to detect all the changes that can contribute to diagnosis as soon as possible and two track the progression of the disease at an early stage. For early AD detection, language-based tools have the potential to provide a fast, non-invasive accessible cheap tool as there are approved no specific treatments. Early detection is needed because-

(1) Early detection can slow that disease progression [35].
(2) The patient should be empowered to make his decision on his own.
(3) Preventive lifestyle changes could be introduced.

Natural language processing is a machine learning-based process that has matured to a top point where it can offer a lot of support for clinical practice. Representation of text computationally by converting written text to interpretable data. Common NTP tasks in the clinical domain include text classification named entity recognition and relation extraction. In the clinical domain, there are numerous benefits and the studies about it have shown the early signs of AD and it can be detected by speech using those techniques.

In AD issues of autonomy can arise in day-to-day affairs, consenting to treatment, legal representation and participating in research.

VI. IMPACT OF FEATURES ON DETECTION ACCURACY AND RELIABILITY

The impact of different combinations of features and algorithms on the accuracy and reliability of Alzheimer's disease detection is profound and multifaceted. When exploring various feature sets, such as linguistic, acoustic, and semantic features extracted from speech and text data, researchers uncover unique insights into the underlying cognitive changes associated with the disease. The selection and fusion of these features are pivotal, as they directly influence the discriminative power of machine learning and deep learning algorithms [36]. For instance, combining acoustic features like pitch and speech rate with linguistic features such as vocabulary richness and syntactic complexity can provide a more comprehensive representation of cognitive decline. Similarly, leveraging diverse algorithms like Support Vector Machines (SVMs), Random Forests, or advanced neural networks enables the exploitation of different learning paradigms and model architectures. The synergy between feature selection and algorithm choice can enhance detection accuracy by capturing subtle variations in language and speech indicative of Alzheimer's pathology. Furthermore, the interpretability and generalizability of the detection system hinge on the harmonious integration of these components, paving the way for robust diagnostic tools that can aid in early intervention and personalized care.

VII. LONGITUDINAL ANALYSIS

a) Changes in Speech and Language Patterns Over Time

Changes in speech and language patterns over time can indeed serve as early indicators of Alzheimer's disease progression. Research suggests that fine alterations in linguistic abilities, including vocabulary usage, syntax, discourse coherence, and speech characteristics, may manifest in the early stages of cognitive decline associated with Alzheimer's disease [37]. For instance, individuals at risk or in the prodromal stage of Alzheimer's may demonstrate a decline in lexical diversity, characterized by a reduced variety of words used in conversation or writing. They may also exhibit difficulties in word finding, leading to increased hesitations or pauses during speech. Changes in syntactic complexity, such as simpler sentence structures, and impaired discourse coherence, with disjointed or tangential narratives, can also be early signs of cognitive decline.Moreover, alterations in speech characteristics, including reduced speech rate, flattened intonation, and decreased fluency, may become more pronounced as Alzheimer's disease progresses. These changes reflect underlying neurocognitive impairments affecting language processing and production. By monitoring speech and language patterns longitudinally, researchers and clinicians can track subtle shifts over time, allowing for early detection and intervention before a significant cognitive decline occurs [38]. Therefore, speech and language analysis holds promise as a non-invasive and potentially sensitive tool for identifying individuals at risk of Alzheimer's disease progression, facilitating timely clinical assessments and interventions.

b) Correlation of Speech Changes and Clinical Measures of Cognitive Decline

The changes observed in speech and language patterns in individuals at risk of or diagnosed with Alzheimer's disease often correlate with clinical measures of cognitive decline. As cognitive impairment progresses, linguistic and acoustic alterations become more pronounced and may align with standardized clinical assessments used to evaluate cognitive function. For example, reduced lexical diversity, increased word-finding difficulties, and simplified syntactic structures in speech and writing can correlate with lower scores on cognitive tests assessing verbal fluency, language comprehension, and executive function. Furthermore, changes in speech characteristics such as slower speech rate, disrupted prosody, and increased hesitations may correlate with clinical assessments of attention, processing speed, and working memory [39]. Clinical measures such as the Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), or various neuropsychological tests can capture these cognitive changes objectively, providing a quantitative framework for assessing disease progression.

By establishing correlations between linguistic/acoustic changes and clinical measures of cognitive decline, researchers can validate the utility of speech and language analysis as complementary tools for monitoring disease progression in Alzheimer's and related disorders. This integration enhances the sensitivity and specificity of diagnostic approaches, facilitating earlier detection, targeted interventions, and personalized care for individuals affected by cognitive decline.

VIII. GENERALIZATION OF SPEECH AND TEXT-BASED DETECTION METHODS

Speech and text-based detection methods across different populations, demographics, and clinical settings are a critical consideration in their application for Alzheimer's disease detection. Differences in language usage, dialects, and cultural norms across populations can influence the performance of detection models trained on specific linguistic features [39]. Speech and text patterns may vary significantly among individuals from different regions or cultural backgrounds, impacting the robustness of algorithms developed using homogeneous datasets.

Secondly, demographic factors such as age, education level, and socio-economic status can also affect the manifestation of linguistic and cognitive changes associated with Alzheimer's disease. Detection methods need to be evaluated across a range of demographic groups to ensure they maintain accuracy and reliability across diverse populations. Moreover, variations in clinical settings, including differences in diagnostic protocols, access to healthcare, and availability of resources, can pose challenges for deploying speech and text-based detection tools universally [40]. Models trained and validated in controlled research settings may encounter practical limitations when applied in real-world clinical environments with varying levels of expertise and infrastructure.

To address these challenges, researchers must conduct rigorous validation studies across diverse populations, demographics, and clinical settings. This involves collecting representative datasets encompassing different languages, cultural contexts, and demographic profiles [41]. Additionally, model adaptation techniques such as transfer learning, domain adaptation, and cross-validation can enhance the generalizability of detection methods beyond their initial training context. Speech and text-based detection methods can be refined and tailored for broader clinical applicability, ultimately enhancing their utility in early detection and management of Alzheimer's disease on a global scale.

IX. INTEGRATION OF SPEECH AND TEXT INTO HEALTHCARE FOR ALZHEIMER’S DETECTION

Integrating speech and text-based detection tools into healthcare settings for routine screening and monitoring of Alzheimer's disease presents promising opportunities to enhance early detection and personalized care. These tools can be integrated into clinical workflows through several key strategies. First, establishing standardized protocols for administering speech and text-based assessments as part of routine check-ups can facilitate systematic screening for cognitive changes associated with Alzheimer's [41]. Healthcare providers can incorporate digital platforms or specialized software applications to capture and analyse speech and text data during patient interactions. Additionally, implementing machine learning algorithms that automate the analysis of linguistic and acoustic features can assist healthcare professionals in generating objective diagnostic insights efficiently [43]. Moreover, integrating these tools with electronic health records (EHRs) enables seamless data exchange and longitudinal tracking of patients' linguistic profiles over time, facilitating continuous monitoring and early intervention. Collaborative efforts between clinicians, researchers, and technology developers are essential to ensure the ethical use, validation, and adoption of speech and text-based detection tools within healthcare settings, ultimately advancing precision medicine approaches for Alzheimer's disease detection and management.

CONCLUSION

The review underscores the transformative potential of harnessing speech and text data for Alzheimer's disease detection. The diverse approaches discussed, ranging from linguistic analysis to machine learning models, highlight the intricate interplay between language and cognitive decline. By examining linguistic and acoustic features, researchers can unveil delicate indicators of Alzheimer's pathology, paving the way for early detection and intervention. Ethical considerations, including privacy safeguards and bias mitigation, underscore the importance of responsible data usage in this field. Looking ahead, integrating speech and text-based detection tools into healthcare settings holds promise for routine screening and personalized monitoring, facilitating timely diagnosis and tailored interventions. Collaborative efforts between researchers, clinicians, and technology developers are essential for translating these advancements into impactful clinical practice. Ultimately, harnessing the power of words offers a compelling avenue toward enhancing Alzheimer's disease detection and improving patient outcomes.

References

1. Ablimit, A., Botelho, C., Abad, A., Schultz, T., &Trancoso, I. (2022, May). Exploring dementia detection from speech: Cross corpus analysis. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6472-6476). IEEE.
2. Adarkhi, K. R., Patel, H. A., & Patel, R. K. (2024). A Comprehensive Review of Advancement in Speech-Based Approach for Alzheimer Disease Detection in India. *International Research Journal on Advanced Engineering and Management (IRJAEM)*, *2*(04), 1039-1047.
3. Bang, J. U., Han, S. H., & Kang, B. O. (2024). Alzheimer's disease recognition from spontaneous speech using large language models. *ETRI Journal*.
4. Bertini, F., Allevi, D., Lutero, G., Calzà, L., &Montesi, D. (2022). An automatic Alzheimer’s disease classifier based on spontaneous spoken English. *Computer Speech & Language*, *72*, 101298.
5. Botezatu, M. R., Miller, E., &Kiselica, A. M. (2023). Limited connectedness of spontaneous speech may be a marker of dementia due to Alzheimer’s disease. *Frontiers in Aging Neuroscience*, *15*.
6. BT, B., & Chen, J. M. (2024). Performance Assessment of ChatGPT versus Bard in Detecting Alzheimer’s Dementia. *Diagnostics*, *14*(8), 817.
7. Burke, E., Gunstad, J., Pavlenko, O., & Hamrick, P. (2024). Distinguishable features of spontaneous speech in Alzheimer’s clinical syndrome and healthy controls. *Aging, Neuropsychology, and Cognition*, *31*(3), 575-586.
8. Cai, H., Huang, X., Liu, Z., Liao, W., Dai, H., Wu, Z., ... & Li, X. (2023). Exploring Multimodal Approaches for Alzheimer's Disease Detection Using Patient Speech Transcript and Audio Data. *arXiv preprint arXiv:2307.02514*.
9. Cui, Z., Wu, W., Zhang, W. Q., Wu, J., & Zhang, C. (2023, December). Transferring Speech-Generic and Depression-Specific Knowledge for Alzheimer’s Disease Detection. In *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 1-8). IEEE.
10. Davuluri, R., &Rengaswamy, R. (2022). A Pre-trained Neural Network to Predict Alzheimer's Disease at an Early Stage. *International Journal of Advanced Computer Science and Applications*, *13*(5).
11. García-Gutiérrez, F., Alegret, M., Marquié, M., Muñoz, N., Ortega, G., Cano, A., ... & Valero, S. (2024). Unveiling the sound of the cognitive status: Machine Learning-based speech analysis in the Alzheimer’s disease spectrum. *Alzheimer's Research & Therapy*, *16*(1), 26.
12. He, R., Chapin, K., Al-Tamimi, J., Bel, N., Marquié, M., Rosende-Roca, M., ... &Hinzen, W. (2023). Automated classification of cognitive decline and probable alzheimer's dementia across multiple speech and language domains. *American Journal of Speech-Language Pathology*, *32*(5), 2075-2086.
13. He, R., Chapin, K., Al-Tamimi, J., Bel, N., Marquié, M., Rosende-Roca, M., ... &Hinzen, W. (2023). Automated classification of cognitive decline and probable alzheimer's dementia across multiple speech and language domains. *American Journal of Speech-Language Pathology*, *32*(5), 2075-2086.
14. Ivanova, O., Martínez-Nicolás, I., &Meilán, J. J. G. (2024). Speech changes in old age: methodological considerations for speech-based discrimination of healthy ageing and Alzheimer's disease. *International Journal of Language & Communication Disorders*, *59*(1), 13-37.
15. Jahan, Z., Khan, S. B., &Saraee, M. (2024). Early dementia detection with speech analysis and machine learning techniques. *Discover Sustainability*, *5*(1), 1-18.
16. Kheirkhahzadeh, M. (2023). Speech Classification using Acoustic Embedding and Large Language Models Applied on Alzheimer’s Disease Prediction Task.
17. Li, J., Yu, J., Ye, Z., Wong, S., Mak, M., Mak, B., ... & Meng, H. (2021, June). A comparative study of acoustic and linguistic features classification for alzheimer's disease detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6423-6427). IEEE.
18. Li, C., Xu, W., Cohen, T., &Pakhomov, S. (2024). Useful blunders: Can automated speech recognition errors improve downstream dementia classification? *Journal of Biomedical Informatics*, *150*, 104598.
19. Liu, Z., Guo, Z., Ling, Z., & Li, Y. (2021, June). Detecting Alzheimer’s disease from speech using neural networks with bottleneck features and data augmentation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7323-7327). IEEE.
20. Li, J., Yu, J., Ye, Z., Wong, S., Mak, M., Mak, B., ... & Meng, H. (2021, June). A comparative study of acoustic and linguistic features classification for alzheimer's disease detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6423-6427). IEEE.
21. Luz, S., Haider, F., Fromm, D., Lazarou, I., Kompatsiaris, I., &MacWhinney, B. (2024). An overview of the ADReSS-M Signal Processing Grand Challenge on Multilingual Alzheimer's Dementia Recognition through Spontaneous Speech. *IEEE Open Journal of Signal Processing*.
22. Martinc, M., Haider, F., Pollak, S., & Luz, S. (2021). Temporal integration of text transcripts and acoustic features for Alzheimer's diagnosis based on spontaneous speech. *Frontiers in Aging Neuroscience*, *13*, 642647.

23. Meghanani, A., Anoop, C. S., & Ramakrishnan, A. G. (2021, January). An exploration of log-mel spectrogram and MFCC features for Alzheimer's dementia recognition from spontaneous speech. In *2021 IEEE spoken language technology workshop (SLT)* (pp. 670-677). IEEE.
24. Mestach, M., Hartsuiker, R. J., &Pistono, A. (2024). Can we track the progression of Alzheimer's Disease via lexical-semantic variables in connected speech?. *Journal of Neurolinguistics*, *70*, 101189.
Mittal, A., Sahoo, S., Datar, A., Kadiwala, J., Shalu, H., & Mathew, J. (2020). Multi-modal detection of alzheimer's disease from speech and text. *arXiv preprint arXiv:2012.00096*.
25. Meghanani, A., Anoop, C. S., & Ramakrishnan, A. G. (2021). Recognition of alzheimer's dementia from the transcriptions of spontaneous speech using fasttext and cnn models. *Frontiers in Computer Science*, *3*, 624558.
26. Nasreen, S., Rohanian, M., Hough, J., &Purver, M. (2021). Alzheimer's dementia recognition from spontaneous speech using disfluency and interactional features. *Frontiers in Computer Science*, *3*, 640669.
27. Park, C. Y., Kim, M., Shim, Y., Ryoo, N., Choi, H., Jeong, H. T., ... &Youn, Y. C. (2024). Harnessing the Power of Voice: A Deep Neural Network Model for Alzheimer's Disease Detection. *Dementia and Neurocognitive Disorders*, *23*(1), 1.
28. Pappagari, R., Cho, J., Joshi, S., Moro-Velázquez, L., Zelasko, P., Villalba, J., &Dehak, N. (2021, August). Automatic Detection and Assessment of Alzheimer Disease Using Speech and Language Technologies in Low-Resource Scenarios. In *Interspeech* (Vol. 2021, pp. 3825-3829).
29. Parlak, M. M., &Köse, A. (2023). Investigation of the knowledge, experiences, and opinions of Speech and Language Pathologists on assessments and therapies for cognitive communication disorders in people with Alzheimer's disease-A cross-sectional survey in Turkey. *Hacettepe University Faculty of Health Sciences Journal*, *10*(1), 45-57.
30. Petti, U., Nyrup, R., Skopek, J. M., & Korhonen, A. (2023, June). Ethical considerations in the early detection of Alzheimer's disease using speech and AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1062-1075).
31. Petti, U., Baker, S., Korhonen, A., & Robin, J. (2023). The generalizability of longitudinal changes in speech before Alzheimer's disease diagnosis. *Journal of Alzheimer's Disease*, *92*(2), 547-564.
32. Pérez-Toro, P. A., Rodríguez-Salas, D., Arias-Vergara, T., Klumpp, P., Schuster, M., Nöth, E., ... & Maier, A. K. (2022). Interpreting acoustic features for the assessment of Alzheimer's disease using ForestNet. *Smart Health*, *26*, 100347.
33. Priyadarshinee, P., Clarke, C. J., Melechovsky, J., Lin, C. M. Y., BT, B., & Chen, J. M. (2023). Alzheimer's dementia speech (audio vs. text): Multi-modal machine learning at high vs. low resolution. *Applied Sciences*, *13*(7), 4244.
34. Pulido, M. L. B., Hernández, J. B. A., Ballester, M. Á. F., González, C. M. T., Mekyska, J., &Smékal, Z. (2020). Alzheimer's disease and automatic speech analysis: a review. *Expert systems with applications*, *150*, 113213.
35. Robin, J., Xu, M., Balagopalan, A., Novikova, J., Kahn, L., Oday, A., ... & Teng, E. (2023). Automated detection of progressive speech changes in early Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *15*(2), e12445.
36. Runde, B. S., Alapati, A., & Bazan, N. G. (2024). The Optimization of a Natural Language Processing Approach for the Automatic Detection of Alzheimer's Disease Using GPT Embeddings. *Brain Sciences*, *14*(3), 211.
37. Shamei, A., Liu, Y., &Gick, B. (2023). Reduction of vowel space in Alzheimer's disease. *JASA Express Letters*, *3*(3).
38. Syed, M. S. S., Syed, Z. S., Lech, M., &Pirogova, E. (2024). Automated Recognition of Alzheimer's Dementia: A Review of Recent Developments in the Context of InterspeechADReSS Challenges. *Biomedical Signal Processing*, 43-61.
39. TaghiBeyglou, B., &Rudzicz, F. (2024). Context is not key: Detecting Alzheimer's disease with both classical and transformer-based neural language models. *Natural Language Processing Journal*, *6*, 100046.
40. Vigo, I., Coelho, L., & Reis, S. (2022). Speech-and language-based classification of Alzheimer's disease: a systematic review. *Bioengineering*, *9*(1), 27.
41. Vrindha, M. K., Geethu, V., Anurenjan, P. R., Deepak, S., &Sreeni, K. G. (2023, May). A Review of Alzheimer's Disease Detection from Spontaneous Speech and Text. In *2023 International Conference on Control, Communication and Computing (ICCC)* (pp. 1-5). IEEE.
42. Yamada, Y., Shinkawa, K., Nemoto, M., Nemoto, K., & Arai, T. (2023). A mobile application using automatic speech analysis for classifying Alzheimer's disease and mild cognitive impairment. *Computer Speech & Language*, *81*, 101514.
43. Ying, Y., Yang, T., & Zhou, H. (2023). Multimodal fusion for alzheimer's disease recognition. *Applied Intelligence*, *53*(12), 16029-16040.