A Systematic Review of NLP for Dementia: Tasks, Datasets, and Opportunities

Lotem Peled-Cohen and **Roi Reichart**
Technion - Israel Institute of Technology, Israel
lotemi.peled@gmail.com, roireichart@gmail.com

Abstract

The close link between cognitive decline and language has fostered long-standing collaboration between the NLP and medical communities in dementia research. To examine this, we reviewed over 240 papers applying NLP to dementia-related efforts, drawing from medical, technological, and NLP-focused literature. We identify key research areas, including dementia detection, linguistic biomarker extraction, caregiver support, and patient assistance, showing that half of all papers focus solely on dementia detection using clinical data. Yet, many directions remain unexplored, such as artificially degraded language models, synthetic data, digital twins, and more. We highlight gaps and opportunities around trust, scientific rigor, applicability, and cross-community collaboration. We raise ethical dilemmas in the field, and highlight the diverse datasets encountered throughout our review including recorded, written, structured, spontaneous, synthetic, clinical, social media–based, and more. This review aims to inspire more creative, impactful, and rigorous research on NLP for dementia.

1 Introduction

Dementia is a broad term for a decline in cognitive function caused by various underlying pathologies. It is a progressive, irreversible condition that worsens over time, with no known treatment or cure. The global impact of the disease is staggering. According to the World Health Organization (WHO),[1] dementia is currently *the seventh leading cause of death globally*. As of 2023, approximately 55 million people worldwide are living with dementia, and this number is expected to nearly double every 20 years (Prince et al., 2015). Global dementia costs are estimated in the trillions of US dollars, with approximately half of these costs attributed to care provided by informal carers (e.g., family members and close friends), as hospitals and care facilities are overcrowded, and healthcare professionals require specialized training to diagnose patients and address their complex needs.

There are multiple types of dementia, including Alzheimer's (accounting for 60%–70% of cases[1]), vascular dementia, Lewy body dementia, and others. Although these differ in pathology, with some manifesting as abnormal protein deposits and others as reduced blood flow to the brain, they often share common symptoms such as memory loss, confusion, behavioral changes, and language deterioration (McKhann et al., 2011). In this review, we use the general term 'dementia' to encompass all forms and levels of severity of the illness.

Dementia is currently diagnosed through pathological tests (e.g., brain imaging, blood tests) and cognitive assessments, such as *Naming and Verbal Fluency tests*, which assess the ability to identify objects or generate words in specific categories (Kaplan et al., 2001), and *Picture Description tests*, evaluating language and narrative skills through descriptions of complex images (Mueller et al., 2018). In these face-to-face interviews, clinicians seek linguistic markers of cognitive decline, such as repetitive language, empty or disorganized speech, word-retrieval difficulties, and excessive descriptions (''the thing you write with'' instead of ''pen'').

Given the prominent role of language in our understanding and diagnosis of the disease, it is no surprise that NLP methodologies are widely used in both the NLP and medical communities. Cognitive assessments and similar diagnostics often result in recorded and transcribed data, offering a wealth of textual content that is well-suited for NLP analysis. This, combined with relatively new sources of data (such as social media posts or conversations with LLMs), enhance the potential

[1] https://www.who.int/news-room/fact-sheets/detail/dementia.

Transactions of the Association for Computational Linguistics, vol. 13, pp. 1204–1244, 2025. https://doi.org/10.1162/tacl.a.35
Action Editor: Dilek Hakkani-Tur. Submission batch: 2/2025; Revision batch: 5/2025; Published 9/2025.

of NLP to advance dementia research in various directions.

Previous literature reviews on NLP for dementia focus on detection methods (Clarke et al., 2020; Petti et al., 2020; Saleem et al., 2022; Parsapoor, 2023; Qi et al., 2023; Hiremath and Biradar, 2023; Vrindha et al., 2023), with some specifically covering deep learning approaches to detection (Shi et al., 2023; Javeed et al., 2023). Other reviews extend beyond detection, addressing tasks like linguistic biomarker extraction (Gagliardi, 2024) or exploring the role of technology in patients' lives (D'Onofrio et al., 2017; Shahapure et al., 2022; Saragih et al., 2023; Peres and Campos, 2024), though these are not fully focused on NLP methods. A few reviews are even more specific, covering dementia-related data (Mueller et al., 2018; De la Fuente Garcia et al., 2020; Yang et al., 2022).

Our review stands out by covering the full spectrum of dementia-related NLP efforts, rather than focusing on specific aspects like tasks (e.g., detection), technologies (e.g., deep learning), data types (e.g., clinical), or communities (NLP vs. medical). Moreover, we created our review with NLP readers in mind, unlike prior reviews, which were primarily published in medical literature and aimed at clinicians (Figure 3). Additionally, our work is distinguished by its scope and coverage. We review *242 papers* from four distinct scientific communities, categorized by publication venue: the medical community (e.g., Nature Medicine); the NLP community (e.g., ACL); the Speech community (e.g., Interspeech); and a broader technological community that is not necessarily NLP- or Speech-specific (e.g., Frontiers in Computer Science). In Section 2, we detail the construction of our cohort.

In Section 3, we describe the main task families identified in our reviewed papers: *linguistic biomarker extraction, caregiver support, patient assistance*, and, most prominently, *dementia detection*, which accounts for over 56% of the cohort's focus. For each task family, we review motivations, current approaches, and potential future directions.

Section 4 examines gaps and opportunities. We highlight, for instance, that the vast majority of studies, across all four task families, rely on a handful of well-known datasets, despite the existence of many unique datasets varying in size, type, and purpose (a summary of 17 dementia-specific datasets can be found in Table 2). We then dive into the field's scientific rigor and explores NLP's potential to influence medical research-provided trust between the two communities is established.

To provide readers with a concrete takeaway on future research directions, Section 5 outlines open challenges such as personalized LLMs for patients and caregivers, artificially degraded language models, and many more. We conclude with the unique ethical considerations of the field (Section 6) and the limitations of our review (Section 7).

Our work aims to inspire researchers from various fields to tap into the vast potential of NLP in dementia research. We hope to provide a fresh perspective on the domain, emphasizing that the opportunities extend far beyond detection. Whether developing clinical applications, analyzing disease progression, or alleviating caregiver burden, this review serves as a valuable resource for those looking to contribute to the fight against dementia.

2 Methodology

This review was conducted according to the PRISMA guidelines (Moher et al., 2009). We searched for NLP- and dementia-related papers in titles, abstracts, and keywords across ACL Anthology, PubMed, DBLP, IEEE Xplore, Springer, and Wiley (full query details in Appendix A). An automatic screening ensured papers were full academic studies (rather than posters or theses), peer-reviewed, and in English. We then manually screened for eligibility, selecting studies that (1) focus on text as a primary modality, (2) use datasets that are at least partially in English (to narrow our scope while still addressing multilingual studies), and (3) were relevant to this review. For instance, while the paper of Botros et al. (2020) mentions dementia and language models in its abstract and keywords, it focuses on smart home sensors, making it ineligible. Appendix A provides further examples of papers that do not meet our relevance criteria, alongside a fully detailed overview of our query, screening process, inclusion criteria, and a PRISMA flowchart.

Our screening resulted in **242 relevant papers**, which we then manually annotated with their main contributions, datasets used, algorithmic methods and whether statistical significance was reported

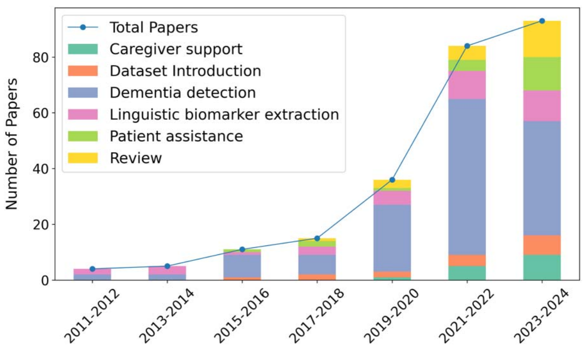

Figure 1: Number of papers per period by task family.

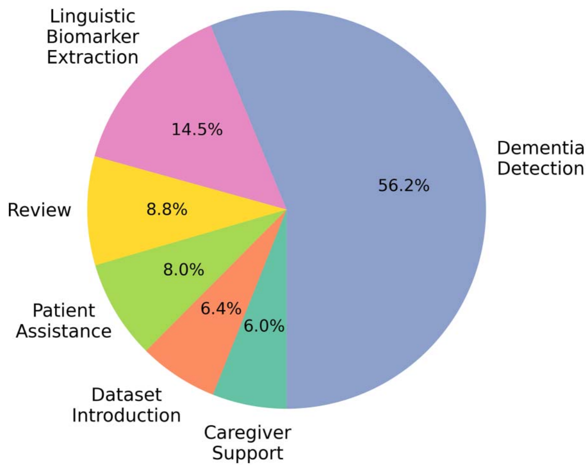

Figure 2: Distribution of task families across papers.

(where applicable). Among this cohort, we identified four distinct task families, described below.

3 Task Families

We identified four distinct research verticals differing by tasks and motivations: **dementia detection, linguistic biomarker extraction, caregiver support**, and **patient assistance** (see Section 3). Two additional, more general categories are **literature reviews** and **dataset introduction** papers, which we describe throughout our work. The number of studies on NLP for dementia has been steadily growing (Figure 1), reflecting a significant interest in the field. While there has been a rise in studies on patient assistance, caregiver support, and literature reviews, there is still an overwhelming focus on dementia detection (Figure 2). Naturally, this imbalance between the four tasks is reflected in our cohort, with over 130 dementia detection papers versus 15–60 papers for each other task family, affecting the depth of our analysis of specific dementia detection studies. Therefore, Appendix B offers further details on the dementia detection papers we reviewed.

Dementia Detection Can an algorithm accurately predict dementia from a provided text? Remarkably, 56% of the papers we reviewed address this exact question. This focus stems from two key factors. First, a *tangible impact*: NLP can improve the diagnostic process by making it faster, less invasive, and more affordable. Pathological changes can begin 15–20 years before cognitive symptoms are noticeable to others (Jack et al., 2010), and if NLP algorithms can detect early stages of dementia (for example, mild cognitive impairment [MCI]) they may enable earlier intervention, potentially slowing disease progression and delaying costly care.

The second motivation to pursue dementia detection is a *data-driven* one. NLP research is inherently empirical, and in this domain, data is often structured, spanning medical, audio, or textual modalities, and annotated with categorical labels such as 'healthy' or 'dementia.' This structure perfectly suits classification algorithms, making dementia detection a well-defined challenge for researchers.

NLP-based dementia detection has been an active research domain for over a decade, primarily leveraging well-known datasets such as the Pitt corpus (Becker et al., 1994), part of the Dementia-Bank cohort (Lanzi et al., 2023). The Pitt corpus provides transcribed recordings of cognitive assessments, such as picture descriptions. These transcribed text excerpts (e.g., 'there's a young boy, uh, going in a cookie jar, and there's a lit... a girl...') are annotated with the speaker's cognitive state (e.g., Healthy, MCI, Alzheimer's) allowing for text-based classification. The transcriptions are also annotated with demographic information and, for some participants, other cognitive assessment scores. Two other widely used data sources, the ADReSS and ADReSSo challenge datasets (Luz et al., 2021a,b, respectively), are derivatives of the Pitt corpus, offering refined transcripts and a more demographically balanced sample.

Unlike these datasets, which build on structured conversations, the widely used CCC dataset (Pope and Davis, 2011) contains transcriptions of spontaneous conversations, about memories, health, and daily life (e.g., 'hmm. This is my problem

(...) maybe you could help me find it'). These four popular datasets (Pitt, ADReSS, ADReSSo, and CCC) exemplify the most common data acquisition approach in dementia detection: transcribed conversations (whether from well-known datasets or self-collected clinical data) annotated with Dementia or Healthy Control labels.[2]

Across the 136 detection papers we reviewed, several algorithmic approaches emerged. Please note that we present representative citations here; for the complete list of references, see Appendix B. Before 2017, many studies relied on classic machine learning methods like SVM and Random Forests using straightforward features like N-grams (Jarrold et al., 2014; Fraser et al., 2016a; Zhou et al., 2016; Santos et al., 2017), achieving accuracy rates around 85%. From 2018 onwards, there was a surge in transformer-based classifiers, particularly applied to the ADReSS and ADReSSo challenges published around that time (Pappagari et al., 2020; Edwards et al., 2020; Haulcy and Glass, 2021; Balagopalan et al., 2021). These methods pushed accuracy over 90%, which are currently state-of-the-art results on the Pitt corpus and its subsets. Recently, large language models (LLMs) have been utilized for detection via prompt design, embedding extraction, and feature design (Agbavor and Liang, 2022; Wang et al., 2023a; Runde et al., 2024; BT and Chen, 2024; Bang et al., 2024).

Given the straightforward nature of the dementia detection (a text classification task), extensive body of research, and the impressive classification results, one might assume that dementia detection is a solved problem. From a modeling perspective, it has indeed evolved like other text classification tasks, from classic methods to word embeddings, neural networks, and LLMs. However, it remains fundamentally different, requiring tailored preprocessing, posing challenges in data collection from older adults, and demanding specialized linguistic metrics for evaluation, which are complexities we explore further in our review. As for the truly impressive results, they are (a) largely based on the Pitt corpus or its subsets: small, homogeneous datasets with specific characteristics (structured conversation, English only, etc.); (b) only 29% of studies report statistical significance or assess the robustness of their findings; and (c) despite strong performance, no NLP-based classification tool, to our knowledge, has been deployed in real-world applications. We explore these gaps further in Section 4.

Linguistic Bio-Markers Extraction Some studies leverage NLP for linguistic exploration, rather than straightforward detection. They often wish to confirm, on a large scale, the existing knowledge we have about the language of cognitively impaired individuals. For instance, studies such as those of Orimaye et al. (2014), Rosas et al. (2019) and Ilias and Askounis (2022a) used NLP methods to verify that linguistic cues associated with cognitive decline, such as repetitions (e.g., 'the... the...'), revisions (e.g., 'the woman, uh, mother'), and overuse of pronouns (e.g. 'she' instead of 'mother'), are indeed behaviors significantly more common in texts produced by the cognitively impaired (Clark et al., 2014; Fraser et al., 2016a; Voleti et al., 2019).

Other studies introduce linguistic markers and metrics to assess a speaker's cognitive state. For example, Roark et al. (2007) extracted *complexity scores* from parse trees, quantifying the extent to which sentences produced by cognitively impaired individuals are structurally and grammatically less complex. Sirts et al. (2017) calculated idea density to measure how efficiently dementia patients convey ideas, while Pompili et al. (2020a) quantified the number and order of topics mentioned by speakers in cognitive interviews. In other studies, such as Choi et al. (2019), the use of *meta-semantic terms* (words implying emotion, emphasis, or opinion) was analyzed in picture descriptions as an indicator of cognitive performance. The *frequency of disfluencies*, including silent pauses, reformulations, and context switches (Adhikari et al., 2021; Farzana et al., 2022; Williams et al., 2023), has also been identified as a strong linguistic marker, even enabling the longitudinal tracking of disease progression (Martinc et al., 2021; Robin et al., 2023).

Findings from this task family impact both the fields of NLP and medicine. From an NLP perspective, they underscore that the language of cognitively impaired individuals differs drastically from typical language, presenting unique challenges to standard NLP practices. For example, these studies show that certain preprocessing steps traditionally used in NLP, such as removing

[2]Some datasets offer more granular annotations, such as Dementia, MCI, Alzheimer's, etc.

repetitions and stop words, could inadvertently erase dementia-related linguistic signals that are critical for tasks like dementia detection. From a medical perspective, these studies have the potential to challenge conventional scoring practices in cognitive assessments. One example is Prud'hommeaux et al. (2011), who explored narrative recall tasks, where participants are asked to retell a story after hearing it. The researchers found that scoring this cognitive assessment based on the presence of *specific story elements* may be more effective for detecting MCI than scoring based on *the overall summary*, which is the traditional approach, thus challenging the current perspective on this task.

Notably, 75% of papers in this task family report statistical significance, which is the highest ratio among all task families. This likely stems from the motivation to ensure that existing knowledge, or any new linguistic insights, are robust.

Caregiver Support With the growing number of dementia patients, the demand for caregivers rises. In the US alone, an estimated 11 million Americans serve as caregivers for those with dementia, contributing billions of hours of care each year.[3] Research indicates that nearly half of these caregivers experience depression, and are at a higher risk of chronic health conditions (Huang, 2022). NLP methods can support these devoted caregivers by detecting emotional distress, providing answers to their concerns and offering companionship.

Around 6% of our cohort consists of studies on NLP for caregiver support, a field that emerged post-COVID-19. Early research focused on the emotional well-being of personal caregivers (family and friends) mainly through social media posts (Monfared et al., 2021; Azizi et al., 2024). For example, Sunmoo et al. (2022, 2023) used graph-based topic modeling and sentiment analysis to show that tweets posted throughout the pandemic shifted from practical care to emotional distress (e.g., depression, helplessness, elder abuse) and coping strategies (e.g., therapeutic reading). Posts beyond dementia-specific communities (Ni et al., 2022; Lal et al., 2023) and across social media profiles (Klein et al., 2022) reveal that caregivers express more than emotional distress, asking also for financial aid and legal advice.

Other studies focus on *professional caregivers*. One example is Zhu et al. (2022b), who analyzed clinical notes by nurses in aged care facilities. Disturbingly, their analysis shows dozens of distinct aggressive behaviors targeted towards the nursing team, such as pushing, shouting, and using profane language, involving over 50% of dementia patients studied. This line of research highlights the potential of NLP in detecting when to assist caregivers, whether through advice, emotional support, or enhanced workforce training.

Another growing body of research is exploring whether LLMs could provide caregivers with expert answers to dementia-related inquiries. Current studies show that models like ChatGPT offer relevant and factually correct advice and recommendations (Hristidis et al., 2023; Dosso et al., 2024), but often lack the depth of information available through Google Search or expert sources. LLMs also fall short in offering the emotional support needed when a caregiver seeks to manage a loved one's memory loss, confusion, or aggression (Aguirre et al., 2024). To bridge these gaps, Zaman et al. (2023) and Parmanto et al. (2024) explore the novel concept of fine-tuning LLMs to specifically address caregivers' delicate emotional needs. These studies pave the way for AI-based caregiver companions, a field likely to gain traction with the growing popularity of LLMs. While groundbreaking, these agent-based solutions also present critical ethical dilemmas, explored in Section 6.

Patient Assistance Individuals with dementia can live with the disease for decades,[4] and NLP solutions can significantly improve their daily lives. NLP can also help researchers decode the mental landscapes of patients, where agitation, depression, and cognitive decline intersect. Research indicates that depression affects 30%–50% of dementia patients (Lyketsos and Lee, 2003), with 15% experiencing suicidal thoughts (Naismith et al., 2022). This highlights the importance of studies such as Fraser et al. (2016b) and Ehghaghi et al. (2022), who aim to detect depression among dementia patients. These studies demonstrate that

[3]https://www.alzint.org/about/dementia-facts-figures/.

[4]https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses/later-stages-dementia.

it is extremely challenging to detect depression from the transcribed cognitive assessments alone, and that adding multi-modal data, particularly audio, drastically helps.

Apart from assessing the mental well-being of patients, NLP can also aid in text simplification, making online information more accessible to those who often find it unreadable (Espinosa-Zaragoza et al., 2023). Engineer et al. (2023) analyzed dementia-related texts online and found an average readability level equivalent to 12 years of education, making them potentially inaccessible to many at-risk communities, given the link between lower education and dementia susceptibility.[5] The study also reveals a consistently negative tone in the texts (e.g., emphasizing the grim nature of the disease), potentially affecting the mental state of patients who are already prone to depression. NLP tools are natural candidates to help simplify texts, enhance readability, and rephrase to a more supportive tone. They can also assist writers and content creators ensure their work does not promote or reinforce stigma against patients (Pilozzi and Huang, 2020).

Researchers are also trying to help patients sustain meaningful conversations and prevent communication breakdowns, shown to occur in over a third of interactions involving individuals with dementia. Studies such as Green et al. (2012) and Chinaei et al. (2017) have developed systems to detect confusion and disfluent speech patterns within transcribed interviews, suggesting real-time repair strategies to minimize misunderstandings. Such solutions may enhance conversations for dementia patients while also easing the burden on caregivers.

A growing body of research explores LLM-powered chatbots to enhance social engagement, provide cognitive stimulation, and reduce loneliness (Kostis et al., 2022; Xygkou et al., 2024; Qi, 2023; Gholizadeh et al., 2023). Studies show that participants appreciate the novelty and cognitive stimulation, reporting reduced loneliness and increased social support. However, they note some limitations such as handling emotionally sensitive conversations. Encouragingly, studies like Addlesee and Eshghi (2024) focus on improving chatbot patience and empathy, advancing the development of more suitable LLMs for patients.

[5]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6937498/.

Additionally, as with all AI tools, building trust with patients and caregivers is essential. Gilman et al. (2024) propose specialized GPT training for dementia patients as a way to help establish this trust. Studies such as Pacheco-Lorenzo et al. (2024) and Treder et al. (2024) interviewed patients, healthy individuals, and caregivers about their views on such technologies, revealing concerns about bias, data privacy, and emotional intelligence, alongside a heartwarming, cautious optimism. As one patient experimenting with a chatbot noted: *''I could talk to the robot longer than I could talk to a human... she didn't tell me if I've repeated myself... she [the bot] didn't think I was boring''*. (Xygkou et al., 2024)

Overview Figure 3 shows the contributions of different scientific communities to the discussed task families, including literature reviews and dataset introductions. While all communities engage in dementia detection and dataset publication, caregiver support remains unexplored in NLP and Speech research. Additionally, we found no literature review similar to ours in NLP venues, highlighting a significant gap. The next section explores these and other gaps, highlighting opportunities for deeper, more rigorous research, as well as ethical dilemmas and innovative directions.

4 Gaps and Opportunities

4.1 Data Variety

Classic We begin our discussion of gaps by addressing a core aspect of NLP research: the data. As noted in Section 3, the most popular datasets in the field are the Pitt corpus and its derivatives (ADReSS and ADReSSo), along with the CCC dataset (see additional details in Table 1). These datasets are foundational and have significantly shaped the current state of NLP for dementia. However, like all data sources, they have limitations, including size constraints and demographic biases. For instance, the Pitt corpus includes data from more women than men, as well as an imbalanced ratio of 35% healthy, 48% dementia, and 17% likely dementia participants. The ADReSS dataset was created as a balanced subset of Pitt, offering a smaller but evenly split sample by age, gender, and diagnosis (Ševčík and Rusko, 2022). Additionally, these datasets capture only the language of individuals

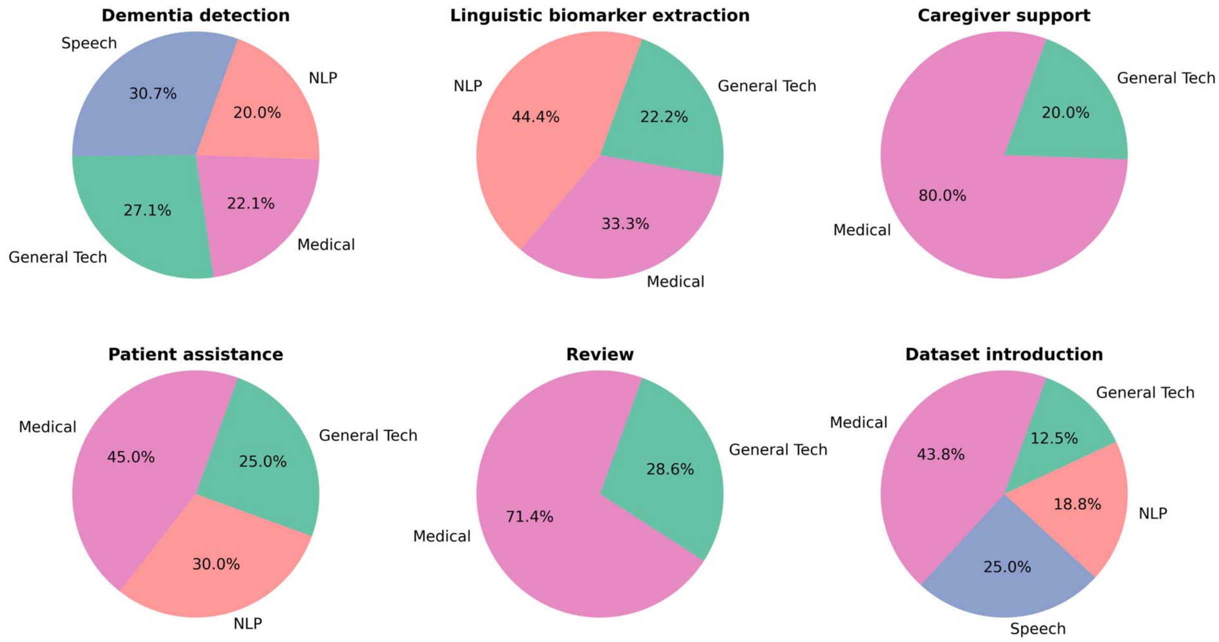

Figure 3: Distribution of venues across each of the task families.

	Name	Year	Type	Population	Modalities	Size
Classic	Pitt	1994	Cognitive Assessments	Patients	Transcribed speech	196 Dem vs. 98 Ctrl
	ADReSS	2020	Cognitive Assessments	Patients	Transcribed speech	78 Dem vs. 78 Ctrl
	ADReSSo	2021	Cognitive Assessments	Patients	Transcribed speech	121 Dem vs. 116 Ctrl
	CCC	2011	Health and Wellbeing Conversations	Patients	Transcribed speech	125 Dem vs.125 Ctrl
Contemporary	CareD	2023	Social media posts	Caregivers	Caregivers written text	1005 Posts
	LoSST-AD	2024	Public interviews	Patients (Celebrities)	Transcribed Interviews	10 Dem vs. 10 Ctrl
	Li et-al	2023	Clinical Notes + Synthetic Sentences	Clinicians / Patients	(1) Human + Synthetic clinical notes (2) Synthetically generated sentences	30k sentences

Table 1: Comparison between classic, widely-used datasets focused on transcribed interviews and cognitive assessments, and examples of contemporary datasets showcasing caregiver data, synthetic data, and more. For the full list of 17 datasets encountered throughout our review, see Appendix B.

with the *access, ability, and willingness to participate in clinical trials*. As a result, communities with less access to such trials, limited education, or intellectual and physical disabilities (factors known to impact cognitive assessments) are not represented (Bruhn and Dammeyer, 2018). This underrepresentation also extends to rural populations, ethnic minorities, and individuals navigating dementia outside of formal healthcare systems, further limiting the datasets' generalizability.

Sixty percent of dementia patients live in low- and middle-income countries,[6] making it crucial to gather data multilingual data from such communities. Some studies try to bridge this gap using translation, transfer learning, or cross-lingual statistical methods for resource-limited languages (Drame et al., 2012; Fraser et al., 2019; Lindsay et al., 2021; Guo et al., 2020; Pérez-Toro et al., 2022, 2023; Kabir et al., 2023; Meng et al., 2023; Melistas et al., 2023). One such example is Nowenstein et al. (2024), who applied translation for dementia detection in Icelandic. They caution that despite their promise, such methods need extensive validation due to varying grammatical complexities and dementia-related nuances across languages. For instance, languages like Japanese omit pronouns in many contexts, which makes it harder to track referential errors, which are a known dementia marker in English (Hasegawa, 1985; Carlomagno et al., 2005). Translations may also remove culturally specific idioms, metaphors, or indirect expressions, potentially impacting diagnosis. Thus, while translation and cross-lingual

[6] `https://www.alzint.org/about/dementia-facts-figures/`.

methods are valuable, English-based data cannot fully substitute for original language data, especially when it comes to underrepresented linguistic and cultural communities.

Another concern with these classic datasets is their timeliness. For instance, the Pitt Corpus, published in 1994, had a median participant age of 67 years. There is a significant difference between 67-year-olds now versus 30 years ago, including changes in technological proficiency and language use. This raises questions about whether models trained on these datasets are effective for detection in today's elderly population, and whether these datasets may be reproduced with updated protocols.

Additionally, these classic datasets were manually transcribed. As such, they contain clean, standardized transcriptions with expert annotations of linguistic nuances such as disfluencies, pauses, and correction. However, manual transcription and annotation are neither scalable nor cost-effective. In real-world applications, texts will therefore likely be transcribed using automatic speech recognition (ASR), which produces noisier outputs and often fails to capture critical linguistic markers of cognitive decline (Al Hanai et al., 2018; Balagopalan et al., 2019; Gómez-Zaragozá et al., 2023; Heitz et al., 2024). As a result, models trained on these clean, expert-transcribed datasets may struggle to generalize to real-world settings.

A final limitation of these classic datasets stems from the dementia stakeholder they represent, namely, the *patient*. Other humans involved, such as caregivers, family members, and even the individuals conducting the cognitive assessments (Tahami Monfared et al., 2022), are not represented. Given the data-driven nature of NLP, this may naturally cause practitioners to focus on patients and inadvertently constrain the scope of research on all other dementia stakeholders.

Contemporary We now turn to data sources beyond clinical trials. *Social media posts* have proven valuable for capturing the broad spectrum of communication across various dementia stakeholders (discussed in Section 3). Some studies creatively used *non dementia-specific datasets, like IMDB*, leveraging their size and multilinguality or translation or embedding extraction for dementia-related tasks (Mirheidari et al., 2018; Li et al., 2019).

An additional direction is freely available data of *public figures diagnosed with dementia*. Petti et al. (2023a,b) and Asllani and Mullen (2023) examine personal writings and spontaneous speech from well-known individuals with dementia, while Petti and Korhonen (2024) curated a corpus of a similar nature (mentioned in Table C). This data source is unique in spanning celebrities of different cultures and origins, various types of data sources (interviews, written texts, etc.), and multiple modalities (video, audio, etc.). Additionally, it allows for tracking these public figures throughout their disease journey, exploring their individual progression. Berisha et al. (2015) and Wang et al. (2020a) emphasize this in their detailed case studies on the linguistic decline of President Ronald Reagan throughout his political career.

LLMs offer another promising approach to data collection, annotation, and augmentation, with proven success in enhancing clinical notes (Liu et al., 2023a; Koga et al., 2024; Latif and Kim, 2024). In a unique study, Li et al. (2023) showed that LLMs can label clinical records, generate clinical data, and spontaneous speech, *improving detection performance beyond that of purely human-created datasets*. The datasets they present are, to the best of our knowledge, the only dementia-related synthetic datasets currently available.

Future Historically, 'NLP for dementia' was synonymous with dementia detection, and the classic datasets containing annotated clinical records have proven well-suited for detection classifiers. However, to advance dementia research in all tasks and frontiers, NLP requires data that is abundant, diverse, and reflective of the disease's complexity. Table 1 compares the traditional clinical datasets with the more contemporary ones we discussed, representing a small part of the **17 datasets** we identified in our review (Table 2 in Appendix C).

The table summarizes each dataset's source and type (recorded, written, structured, spontaneous, synthetic, clinical, social media-based, and more) as well as its modalities (e.g., text, audio, images). We also describe dataset size and diagnosis distribution (when available), and provide references for each dataset so that readers can consult the original sources for full demographic and methodological data collection details. All 17 datasets are English-focused, in line with the scope of this review (as detailed in Sections 2 and 7).

While these 17 datasets have some limitations, such as modest size, incomplete demographic coverage, or older collection dates, they have nonetheless enabled a rich body of research and supported hundreds of studies, substantially advancing NLP for dementia. We hope that curating these resources into an accessible and comprehensive summary will encourage researchers to enter the field, make use of the growing body of data, and pursue creative approaches to data collection that reflect a broader range of dementia stakeholders, languages, genders, cultures, and educational backgrounds. This, in turn, can drive the development of NLP solutions that are inclusive and globally relevant.

4.2 Personalized Approaches

Every dementia patient experiences the disease progression uniquely, shaped by their individual characteristics and 'distinct personal language model'. As each generation produces more accessible data, a 60-year-old today has likely generated significantly more text than a 60-year-old two decades ago. Given this wealth of textual data sourced from older adults, NLP may facilitate personalized approaches for managing, tracking, or even diagnosing dementia by building individualized language models that monitor cognition on a personal level.

NLP-based digital twins for dementia patients represent a promising frontier in healthcare. These virtual models replicate a patient's cognitive and behavioral patterns through their language, enabling personalized care by monitoring cognitive decline and adjusting treatment plans based on simulated behaviors (Andargoli et al., 2024). A notable example is Hong et al. (2024), who developed digital twins by fine-tuning an LLM on longitudinal data from the I-CONECT clinical trial,[7] using these twins to evaluate a patient support chatbot.

An intriguing, yet unexplored, research direction involves creating personalized LLM companions modeled after loved ones, as cognitive stimuli from familiar individuals can positively impact dementia patients (Woods et al., 2012). An LLM trained on family-curated data could engage patients with their personal history, family stories, and interests, such as favorite bands or recipes, potentially offering a non-clinical treatment. Given the vast amount of personal data generated today, this approach seems both feasible and effective. With these research directions, NLP has the potential to shape the future of personalized dementia treatments, both medical and alternative.

[7]https://www.i-conect.org/.

4.3 Trust and Scientific Rigor

A key contribution of our work is drawing comparisons between the medical and technological communities, particularly in scientific modeling. Medical research typically begins with a hypothesis, followed by study design and data collection for analysis. In contrast, data scientists often start with pre-collected data, retrospectively identifying patterns without necessarily defining a specific hypothesis. This conceptual gap may be critical in medical contexts. Is retrospective data relevant to future cases (e.g., will a model trained on the 1994 Pitt corpus capture the same signals in today's dementia patients)? Do the results indicate correlation rather than causation? The medical community may be hesitant to adopt solutions they perceive as correlations derived from retrospectively collected, often small datasets (El-Sappagh et al., 2023).

Hesitance to adopt NLP-based solutions may grow when studies do not report statistical significance or assess robustness. We manually annotated each paper in our cohort, excluding literature reviews, to indicate whether statistical tests or p-values were reported. Among all reviewed papers, **only 33% report statistical significance**. While about half of NLP and medical papers include it, nearly 86% of Speech publications, which cover a significant portion of detection studies, particularly those using the popular ADReSS and ADReSSo datasets, do not (see Figure 4).

Another crucial aspect of scientific modeling is robustness. We found few studies in our cohort which explicitly addressed this topic or evaluated the robustness of their models. Two notable examples are Novikova (2021), who examined BERT-based dementia detection under data perturbations, and Favaro et al. (2024), who assessed the consistency of linguistic features across corpora.

Explainability, a key pillar of scientific rigor, remains rare in dementia detection, with few exceptions like Ilias and Askounis (2022a). A disconnect exists between the NLP community,

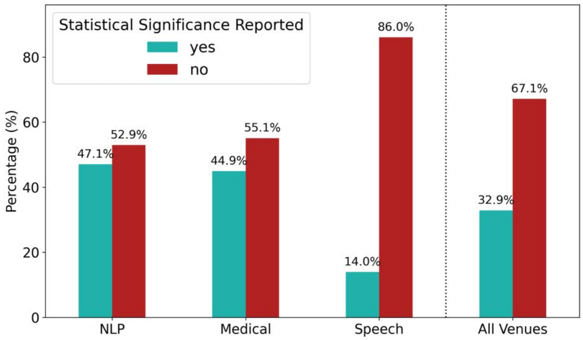

Figure 4: Statistical significance per publication type.

which favors deep learning for higher accuracy, and the medical community, which prioritizes transparency and tends to distrust ''black-box'' models (Adadi and Berrada, 2020). As a result, clinicians often prefer traditional, inherently interpretable models. Another gap exists as NLP practitioners' favor feature-attribution methods, explaining prediction by assigning importance scores to individual tokens or features (Karlekar et al., 2018; Ilias and Askounis, 2022a; Vimbi et al., 2024). While straightforward for researchers, these methods often produce overwhelmingly low-level explanations (e.g., assigning an importance score to every token in a lengthy transcript), potentially confusing both clinicians and patients, especially those with cognitive impairments. Addressing these challenges requires cross-disciplinary collaboration to develop interpretability methods that are both scientifically rigorous and clinically meaningful. We encourage researchers to consult surveys such as Stiglic et al. (2020), Calderon and Reichart (2024), and Viswan et al. (2024), which outline NLP interpretability paradigms applicable across various algorithms.

Ensuring scientific rigor is just as crucial, and equally challenging, for tasks beyond dementia detection. Studies on caregiver support and patient assistance often lack a gold standard or clear objective. For example, evaluating LLMs that answer medical questions or simplify texts for dementia patients poses challenges in assessment and statistical rigor. Developing benchmarks to evaluate these tools is therefore a vital research direction.

4.4 Application State-of-Mind

To combat dementia in real-world settings, NLP practitioners must not only adopt rigorous methods and build trust, they should also consider the applicability of their work, and ask themselves, 'How can my work be applied in practice?' Dementia detection exemplifies this gap. Despite decades of promising results and hundreds of surveyed papers, our research found no evidence that these approaches have been integrated into real-world detection processes.[8] What hinders the clinical adoption of such promising NLP techniques? First, we believe that trust issues (as discussed earlier) are a major factor, as integrating these tools into medical workflows is too risky without rigorous testing across diverse datasets and real-world scenarios. However, they are not the only barrier.

One applicative challenge in dementia detection lies in the reliance on spoken responses, which must be transcribed in real time or post-recording. Clinicians often cannot record assessments due to privacy concerns and technological constraints, and even if recordings were conducted regularly, transcription would remain a bottleneck. Manual transcription is costly and time-consuming, making ASR the most practical alternative. Some studies have explored ways to adapt ASR systems to dementia related tasks, such as training models on domain specific speech data (Abulimiti et al., 2020), or fine tuning state of the art models like Whisper on cognitively impaired speech (Li and Zhang, 2024; Radford et al., 2023). Others use ASR to generate transcripts automatically, extract features for clinical scoring, and compare the results to those derived from manual transcripts (Lehr et al., 2012, 2013). While not focused on ASR improvements per se, these studies further demonstrate that adapting ASR systems to the dementia domain can improve transcription quality for downstream domain-specific tasks. Still, ASR systems often introduce errors and miss critical cognitive signals like disfluencies and silent pauses, making it difficult to distinguish between patient incoherence and transcription errors (Li et al., 2024b). As long as NLP relies on ASR systems that introduce complexities and errors, they may add more burden than they alleviate.

Applicative thinking extends beyond dementia detection, relevant to all dementia-related tasks. Topaz et al. (2020), Zolnoori et al. (2023), and Ryvicker et al. (2024), who tailor their NLP applications to the home healthcare setting, help professional caretakers in identifying, clustering,

[8]Based on publicly available data.

and analyzing dementia symptoms and behaviors (e.g., anxiety, delusions, hoarding). These systems were designed for real-world deployment, aiming to flag dementia-related signals in undiagnosed individuals and facilitate information transfer across healthcare settings, supporting both professional and personal caregivers. Casu et al. (2024) present another form of applicative consideration, designing LLMs for dementia-related tasks and optimizing them for local fine-tuning to ensure data safety, an approach that might be more acceptable to the healthcare community compared to cloud-based training.

Demand for dementia-related applications is evident from a business perspectives as well. The dementia treatment market is projected to surpass US$36 billion by 2030, with independent technological solutions rapidly emerging.[9] For instance, the ChatGPT 'app store' now features over 30 dementia-related chatbots for cognitive stimulation or virtual companionship.[10] People are taking initiative, leveraging LLMs to create these tools, and scientists can follow suit, using this framework to develop LLM-based playgrounds for large-scale applicative experiments. Such applications may also empower older adults to administer self-assesment and monitoring, as shown by Skirrow et al. (2022).

To summarize, the need for practical NLP solutions is undeniable, and we encourage the community to focus on solutions that are not only innovative but also truly applicable in real-life settings.

4.5 NLP Insights for Medicine

With many scientific communities advancing dementia research, cross-disciplinary collaboration is key. While NLP and other technological fields draw from medical insights and data, we argue that the reverse, integrating NLP insights into medical science and practice, holds immense potential.

One example is the concept of *artificially degraded language models*, which explore 'healthy' vs. 'cognitively impaired' models replicating the linguistic behavior of dementia patients. Cohen and Pakhomov (2020) achieved this by fine-tuning an LSTM on the DementiaBank cohort, while Li et al. (2022a) degraded different layers of GPT-2. Both approaches yielded models that impressively demonstrated dementia-like linguistic patterns. The two studies utilized perplexity (a derivative of cross-entropy) as a measure of cognitive decline, a concept introduced in Roark et al. (2011). The degraded models exhibit significantly lower perplexity when exposed to dementia-related linguistic anomalies compared to 'healthy' models. We believe that this research area might shed light on the biological mechanisms of dementia and its pathologies; for example, degrading an LLM by mimicking protein deposits on its neurons might offer insights into the mechanisms of Alzheimer's.

Li et al. (2024a) use a degraded model to explore similarities between children's language and that of dementia patients. Their model shows lower perplexity than GPT-2 on picture descriptions by young children and dementia patients, and significantly higher perplexity on descriptions by children over 8, whose language aligns with that of with healthy adults. This work suggests that NLP can aid in developing dementia biomarkers or cognitive assessments inspired by children's language.

NLP can also help craft novel cognitive assessments that examine more than the patient's performance of the task at hand. Farzana et al. (2020), Nasreen et al. (2021a), Farzana and Parde (2022), and Dawalatabad et al. (2022) analyze interaction dynamics between patients and interviewers, demonstrating that rare dialogue acts, such as clarification exchanges and interviewer utterances, can signal cognitive variations. Reeves et al. (2020) propose an alternative assessment, focusing on narrative descriptions following video observations.

Another research direction can illuminate patients' lifestyle factors and experiences for the medical community. Studies such as Yi et al. (2020), Shen et al. (2022), Wu et al. (2023), and Zhou et al. (2018) focus on extracting patients' lifestyle factors from clinical records and social worker notes. They identify, for example, whether patients experience sleep deprivation, suffer from social isolation, or report substance abuse. Such NLP-based automated approaches can provide insights into a patient's unique experience of the disease and even allow for cognitive profiling (Upadhyay et al., 2022), potentially allowing clinicians to tailor interventions that improve individualized patient outcomes.

[9] `https://www.sphericalinsights.com/reports/dementia-drugs-market`.

[10] Searched on `https://chat.openai.com/`, August 2024.

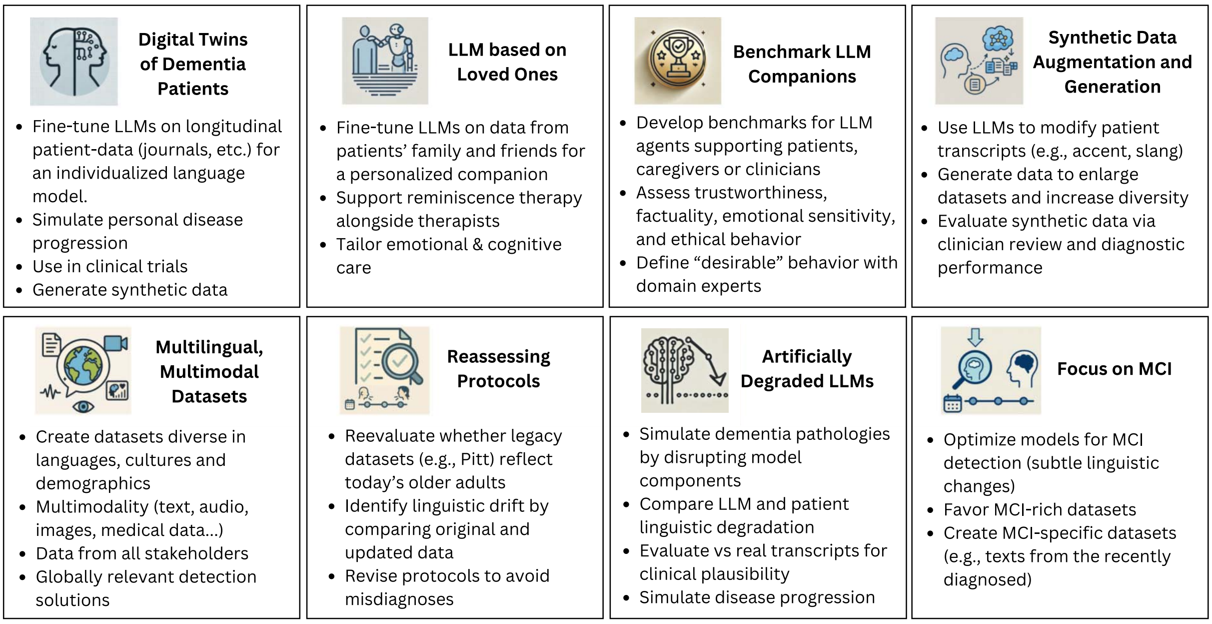

Figure 5: Key research directions we view as foundational, actionable, and yet-to-be fully explored.

Finally, NLP can prompt clinicians to reconsider existing hypotheses about dementia risk factors. Zhou et al. (2019, 2021) conducted extensive analyses of health records, uncovering contradictions with prevailing medical assumptions. While many medical studies emphasize cardiovascular risks, such as high-fat, high-calorie, and high-carbohydrate diets, these factors were rarely noted in dementia patient records. Instead, deficiencies in potassium, calcium, and other nutrients, as well as frequent mentions of dietary supplements (or their absence), appeared in over 25% of dementia patients' records. Actual lab results for nutrient levels, however, were rarely documented, a result which challenges existing hypotheses, and exposes potential gaps in medical procedures.

5 Directions for Future Work

Previous sections outlined the state of NLP for dementia and identified gaps and opportunities. Along the way, we highlighted open challenges that struck us as meaningful, practical, and yet-to-be fully explored (see Figure 5). We now distill these into concrete takeaways for researchers.

In Section 4.2, we discussed personalized LLMs as a research direction that holds immense potential. One example is **digital twins of dementia patients**, LLMs tuned on a patient's own linguistic data, such as clinical interviews, personal journals, or text messages, that can mirror how their language changes over time. These models may help simulate the progression of dementia under different treatment conditions, assist in tailoring interventions, or even serve as training tools for caregivers by generating simulated dialogues. Alternatively, **LLMs based on a patient's loved ones**, i.e., fine-tuned on texts from the patient's family and friends, could provide personalized, emotionally resonant interactions. Prompting a patient to recall life experiences or discuss their favorite movies and songs could potentially assist with reminiscence therapy.

However, to move from promising prototypes to reliable tools, **LLM companions require dedicated benchmarks** tailored to each use case. These should assess trustworthiness, factuality, emotional alignment, and ethical behavior: for example, avoiding hallucinations, adapting tone to emotional needs, and preserving the dignity and privacy of the user. Such benchmarks must be developed in collaboration with clinicians, caregivers, and domain experts, to define ''desired'' and ''undesired'' behaviors in these sensitive, high-stakes interactions.

We also encourage researchers to explore **synthetic data augmentation, alteration, and generation.** LLMs can generate synthetic texts, including patient speech and caregiver interactions, offering a scalable way to curate larger, linguistically inclusive datasets. Such datasets may support models that are more robust to distribution shifts and data perturbations. Another option is to synthetically enhance patient transcriptions to better represent populations excluded from clinical trials (see Section 4.1). This could involve altering transcripts to reflect diverse discourse styles, add slang, or incorporate contemporary language from a broader range of demographic backgrounds. For example, some participants in the Pitt corpus refer to the boy in the Cookie Theft picture as ''Johnny,'' a naming habit typical of a narrow generational and demographic group represented in the dataset. Such highly specific patterns may introduce linguistic bias, which synthetic augmentation could help alleviate. Crucially, the quality and effects of synthetic alterations and generation should be validated through both clinician inputs and empirical comparisons, i.e., evaluating detection accuracy with and without synthetic data.

Researchers should also prioritize **multilingual, multimodal datasets** to ensure that dementia research reflects true linguistic, demographic, and cultural diversity. NLP datasets underrepresent languages from low- and middle-income countries, where over half of dementia patients live (Prince et al., 2012; Tillmann et al., 2019). Countries with high dementia-related mortality, such as Finland, Slovakia, Vietnam, and Indonesia, also remain underrepresented in existing NLP datasets (World Population Review, 2025). Even countries that could benefit from English-based dementia NLP research still require multilingual solutions. For example, in Australia, 21% of dementia patients in aged care facilities are immigrants from non-English-speaking backgrounds, and language barriers with caregivers have been linked to increased agitation and aggression (Edith Cowan University, 2024). As noted in Section 4.1, addressing dementia on a global scale requires more than translating English datasets. Multilingual data should be actively collected to capture dialectal variation and cultural nuance. This data should also extend beyond patients to include caregivers, clinicians, and dementia-related content creators. Multimodal coverage is equally important, spanning speech, video, medical records, imaging, personal writing, social media, and more. These datasets could support a wide range of NLP tools, including detection models and LLM-based agents, that are linguistically robust, culturally sensitive, and globally relevant.

Another important direction is **reproducing classic protocols**, such as the Pitt interviews, to collect updated data from today's older adults. This would help assess the continued relevance of these protocols, detect generational shifts in language use, and evaluate whether models trained on legacy data generalize to current populations. Linguistic norms have changed: Older adults today are more likely to use technological terms, communicate online, and reflect greater sociocultural diversity. As a result, their discourse may differ significantly from that captured in the 1990s and early 2000s, when many benchmark datasets were created (Ladzekpo et al., 2023). Without revisiting these protocols, we risk biasing models against linguistically diverse older adults. Collecting new data using established tasks may allow researchers to compare linguistic patterns across generations and support more inclusive, up-to-date detection methods.

From a medical perspective, we see **artificially degraded LLMs** (discussed in Section 4.5) as a promising direction. By selectively disrupting internal components of an LLM, researchers might be able to simulate various dementia pathologies and explore structural and functional parallels between neural networks and the human brain. For example, they might alter specific layers to simulate frontotemporal dementia or deactivate neurons in a more broad manner to mimic Alzheimer's. These experiments may allow us to ask whether linguistic symptoms emerge in ways that align with clinical observations. For instance, does an LLM simulating frontotemporal symptoms display impaired inhibition or aggression in its language? Comparing outputs from degraded models to real patient transcripts could help evaluate clinical plausibility and identify which components of the LLM support fluency, coherence, or semantic integrity. These simulations also offer practical value: enabling continuous modeling of disease progression, supporting hypothesis testing, and generating synthetic patient data in a safe, controllable manner.

Finally, for NLP to meaningfully support early detection, models must focus on **the MCI stage**. While NLP studies report high accuracy in detecting advanced dementia stages, this has limited clinical utility: (a) such cases are often diagnosable by clinicians without NLP assistance, and

(b) few interventions remain effective at that stage. Instead, researchers must prioritize the detection of MCI, when both symptomatic and disease-modifying treatments may have the greatest impact (Rasmussen and Langerman, 2019; Sims et al., 2023). This requires developing models sensitive to the subtlest linguistic changes and optimizing for the unique characteristics of early-stage decline. It also demands suitable data: Researchers should prioritize datasets with sufficient MCI representation or curate MCI-specific resources using writings or interviews from recently diagnosed individuals.

These research paths demand careful ethical considerations, as we discuss in the next section.

6 Ethical Discussion

Ethical dilemmas regarding AI and healthcare have been widely examined (Rigby, 2019; Gerke et al., 2020; Keskinbora, 2019; Harrer, 2023). However, in this section, we highlight those specific to NLP and cognitive decline. One example is collecting data from the cognitively impaired, which is an ethically delicate task, as these individuals often cannot provide informed consent, raising questions about autonomy and data use. Fairness and diversity is another challenge, as NLP systems may misinterpret silent pauses, repetitions, or corrections, mistaking them for cognitive decline when they actually stem from non-native speech or even simply from stress.

The classic precision-recall trade-off also takes on new ethical weight in dementia research, particularly in detection. For treatable conditions like certain cancers, prioritizing recall to err on the side of caution is often justified. But for untreatable, progressive illnesses like dementia, false positives can be deeply harmful. A misdiagnosis, especially at a younger age, can lead to severe emotional distress, depression, and even suicidal thoughts (Lyketsos and Lee, 2003; Naismith et al., 2022). Misdiagnosed individuals may even be prescribed newly emerging drugs and treatments, exposing them to serious side effects like brain edema and hemorrhages- despite not needing the medication.

Another challenge may arise from using LLMs to generate patient speech. LLM hallucinations (non-factual outputs; Smith et al., 2023) and confabulations (false memories without intent to deceive; Brown et al., 2017) are problematic for many NLP tasks but are also key characteristics of dementia patients. How does one distinguish between LLM fabrications and authentic simulations of cognitive impairment? This raises both technical and ethical concerns: While some generated patterns (e.g., excessive empty speech or extreme hallucinations) may enhance accuracy, they can also introduce bias and reduce interpretability. Misleading data risks distorting our understanding of dementia, its cognition, and lived experience. Establishing benchmarks and ethical guidelines for synthetic data in this domain is essential.

Finally, researchers must ensure that models communicate predictions responsibly and safely. Dementia patients, often experiencing impaired judgment, may overshare sensitive information, act on unsafe advice, or follow suggestions with unintended consequences. For instance, if a chatbot recommends taking a walk to feel better, a patient might do so and become lost, which is an entirely plausible scenario. Addressing these ethical challenges demands cross-disciplinary collaboration, rigorous risk assessment, robust evaluation, and a strong sense of responsibility and accountability.

7 Limitations

Like all surveys, our work represents a snapshot in time. With NLP advancing extremely rapidly, new relevant studies are highly likely emerge after this review's publication. Additional limitations stem from deliberate choices to maintain a manageable scope and focused analysis. For instance, we relied on relatively broad dementia-related terms rather than expanding to all specific pathologies (e.g., Lewy Body, Vascular) or other potentially related medical conditions (e.g., Parkinson's disease). Additionally, we focused on studies using data that is fully or partially in English, thus excluding research conducted solely in other languages. Finally, we confined our survey to peer-reviewed articles, acknowledging that in NLP, innovation often originates in open-source publications. Despite these limitations, we believe our review provides comprehensive coverage of the NLP for Dementia domain and provides a solid starting point for researchers looking to advance this field.

8 Conclusions

Through a cohort of 242 papers, we uncovered the depth and richness of NLP research for dementia. Multiple scientific communities are working to advance early detection, support patients, and caregivers, and much more, all through language. Yet, many gaps and opportunities remain, and the data is out there. We urge researchers to collaborate, share knowledge through rigorous research, and uphold the highest ethical standards to make a true impact and help combat this disease.

References

Nadine Abdelhalim, Ingy Abdelhalim, and Riza Theresa Batista-Navarro. 2023. Training models on oversampled data and a novel multi-class annotation scheme for dementia detection. In *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 118–124. `https://doi.org/10.18653/v1/2023.clinicalnlp-1.15`

Ayimnisagul Ablimit, Catarina Botelho, Alberto Abad, Tanja Schultz, and Isabel Trancoso. 2022a. Exploring dementia detection from speech: Cross corpus analysis. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6472–6476. IEEE. `https://doi.org/10.1109/ICASSP43922.2022.9747167`

Ayimnisagul Ablimit, Karen Scholz, and Tanja Schultz. 2022b. Deep learning approaches for detecting Alzheimer's dementia from conversational speech of ilse study. In *Interspeech*, pages 3348–3352. `https://doi.org/10.21437/Interspeech.2022-10942`

Ayimunishagu Abulimiti, Jochen Weiner, and Tanja Schultz. 2020. Automatic speech recognition for ilse-interviews: Longitudinal conversational speech recordings covering aging and cognitive decline. In *Interspeech*, pages 3795–3799. `https://doi.org/10.21437/Interspeech.2020-2829`

Amina Adadi and Mohammed Berrada. 2020. Explainable ai for healthcare: From black box to interpretable models. In *Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco*, pages 327–337. Springer. `https://doi.org/10.1007/978-981-15-0947-6_31`

Angus Addlesee and Arash Eshghi. 2024. You have interrupted me again!: Making voice assistants more dementia-friendly with incremental clarification. *Frontiers in Dementia*, 3:1343052. `https://doi.org/10.3389/frdem.2024.1343052`, PubMed: 39081607

Surabhi Adhikari, Surendrabikram Thapa, Priyanka Singh, Huan Huo, Gnana Bharathy, and Mukesh Prasad. 2021. A comparative study of machine learning and NLP techniques for uses of stop words by patients in diagnosis of Alzheimer's disease. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE. `https://doi.org/10.1109/IJCNN52387.2021.9534449`

Felix Agbavor and Hualou Liang. 2022. Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health*, 1(12):e0000168. `https://doi.org/10.1371/journal.pdig.0000168`, PubMed: 36812634

Alyssa Aguirre, Robin Hilsabeck, Tawny Smith, Bo Xie, Daqing He, Zhendong Wang, and Ning Zou. 2024. Assessing the quality of chatgpt responses to dementia caregivers' questions: Qualitative analysis. *JMIR Aging*, 7:e53019. `https://doi.org/10.2196/53019`, PubMed: 38722219

Tuka Al Hanai, Rhoda Au, and James Glass. 2018. Role-specific language models for processing recorded neuropsychological exams. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 746–752. `https://doi.org/10.18653/v1/N18-2117`

Ahmed M. Al-Harrasi, Ehtesham Iqbal, Konstantinos Tsamakis, Judista Lasek, Romayne Gadelrab, Pinar Soysal, Enno Kohlhoff, Dimitrios Tsiptsios, Emmanouil Rizos, Gayan Perera, et al. 2021. Motor signs in Alzheimer's disease and vascular dementia: Detection through natural language processing, co-morbid features and relationship to adverse outcomes. *Experimental Gerontology*, 146:111223. `https://doi.org/10.1016/j.exger.2020.111223`, PubMed: 33450346

Ali Amin-Nejad, Julia Ive, and Sumithra Velupillai. 2020. Exploring transformer text generation for medical dataset augmentation. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4699–4708.

Samad Amini, Boran Hao, Jingmei Yang, Cody Karjadi, Vijaya B. Kolachalama, Rhoda Au, and Ioannis C. Paschalidis. 2024. Prediction of Alzheimer's disease progression within 6 years using speech: A novel approach leveraging language models. *Alzheimer's & Dementia*. `https://doi.org/10.1002/alz.13886`, PubMed: 38924662

Samad Amini, Boran Hao, Lifu Zhang, Mengting Song, Aman Gupta, Cody Karjadi, Vijaya B. Kolachalama, Rhoda Au, and Ioannis Ch Paschalidis. 2023. Automated detection of mild cognitive impairment and dementia from voice recordings: A natural language processing approach. *Alzheimer's & Dementia*, 19(3):946–955. `https://doi.org/10.1002/alz.12721`, PubMed: 35796399

Amirhossein Eslami Andargoli, Nalika Ulapane, Tuan Anh Nguyen, Nadeem Shuakat, John Zelcer, and Nilmini Wickramasinghe. 2024. Intelligent decision support systems for dementia care: A scoping review. *Artificial Intelligence in Medicine*, page 102815. `https://doi.org/10.1016/j.artmed.2024.102815`, PubMed: 38553156

Saurav K. Aryal, Ujjawal Shah, Howard Prioleau, and Legand Burge. 2023. Ensembling and modeling approaches for enhancing Alzheimer's disease scoring and severity assessment. In *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1364–1370. IEEE. `https://doi.org/10.1109/CSCI62032.2023.00224`

Beni Asllani and Deborah M. Mullen. 2023. Using personal writings to detect dementia: A text mining approach. *Health Informatics Journal*, 29(4):14604582231204409. `https://doi.org/10.1177/14604582231204409`, PubMed: 37800542

Mehrnoosh Azizi, Ali Akbar Jamali, and Raymond J. Spiteri. 2024. Identifying x (formerly twitter) posts relevant to dementia and covid-19: Machine learning approach. *JMIR Formative Research*, 8:e49562. `https://doi.org/10.2196/49562`, PubMed: 38833288

Aparna Balagopalan, Benjamin Eyre, Jessica Robin, Frank Rudzicz, and Jekaterina Novikova. 2021. Comparing pre-trained and feature-based models for prediction of Alzheimer's disease based on speech. *Frontiers in Aging Neuroscience*, 13:635945. `https://doi.org/10.3389/fnagi.2021.635945`, PubMed: 33986655

Aparna Balagopalan, Benjamin Eyre, Frank Rudzicz, and Jekaterina Novikova. 2020. To BERT or not to BERT: Comparing speech and language-based approaches for Alzheimer's disease detection. *arXiv preprint arXiv:2008.01551*. `https://doi.org/10.21437/Interspeech.2020-2557`

Aparna Balagopalan, Ksenia Shkaruta, and Jekaterina Novikova. 2019. Impact of asr on Alzheimer's disease detection: All errors are equal, but deletions are more equal than others. *arXiv preprint arXiv:1904.01684*. `https://doi.org/10.18653/v1/2020.wnut-1.21`

Jeong-Uk Bang, Seung-Hoon Han, and Byung-Ok Kang. 2024. Alzheimer's disease recognition from spontaneous speech using large language models. *ETRI Journal*, 46(1):96–105. `https://doi.org/10.4218/etrij.2023-0356`

James T. Becker, François Boiler, Oscar L. Lopez, Judith Saxton, and Karen L. McGonigle. 1994. The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594. `https://doi.org/10.1001/archneur.1994.00540180063015`, PubMed: 8198470

Visar Berisha, Shuai Wang, Amy LaCross, and Julie Liss. 2015. Tracking discourse complexity preceding Alzheimer's disease diagnosis: A case study comparing the press conferences of presidents ronald reagan and george herbert walker bush. *Journal of Alzheimer's Disease*, 45(3):959–963. `https://doi.org/10.3233/JAD-142763`, PubMed: 25633673

Angela A. Botros, Narayan Schütz, Hugo Saner, Philipp Buluschek, and Tobias Nef. 2020. A simple two-dimensional location embedding for passive infrared motion-sensing based home monitoring applications. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 5826–5830. IEEE. https://doi.org/10.1109/EMBC44109.2020.9175351, PubMed: 33019299

Mondher Bouazizi, Chuheng Zheng, and Tomoaki Ohtsuki. 2022. Dementia detection using language models and transfer learning. In *Proceedings of the 2022 5th International Conference on Software Engineering and Information Management*, pages 152–157. https://doi.org/10.1145/3520084.3520108

Mondher Bouazizi, Chuheng Zheng, Siyuan Yang, and Tomoaki Ohtsuki. 2023. Dementia detection from speech: What if language models are not the answer? *Information*, 15(1):2. https://doi.org/10.3390/info15010002

Jerrod Brown, D. Huntley, S. Morgan, K. D. Dodson, and J. Cich. 2017. Confabulation: A guide for mental health professionals. *International Journal of Neurology and Neurotherapy*, 4:070. https://doi.org/10.23937/2378-3001/1410070

Peter Bruhn and Jesper Dammeyer. 2018. Assessment of dementia in individuals with dual sensory loss: Application of a tactile test battery. *Dementia and Geriatric Cognitive Disorders Extra*, 8(1):12–22. https://doi.org/10.1159/000486092, PubMed: 29515619

Balamurali B. T. and Jer-Ming Chen. 2024. Performance assessment of chatGPT versus bard in detecting Alzheimer's dementia. *Diagnostics*, 14(8):817. https://doi.org/10.3390/diagnostics14080817, PubMed: 38667463

Joseph Bullard, Cecilia Ovesdotter Alm, Xumin Liu, Qi Yu, and Ruben A. Proano. 2016. Towards early dementia detection: fusing linguistic and non-linguistic clinical data. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 12–22. https://doi.org/10.18653/v1/W16-0302

Hongmin Cai, Xiaoke Huang, Zhengliang Liu, Wenxiong Liao, Haixing Dai, Zihao Wu, Dajiang Zhu, Hui Ren, Quanzheng Li, Tianming Liu, et al. 2023. Multimodal approaches for Alzheimer's detection using patients' speech and transcript. In *International Conference on Brain Informatics*, pages 395–406. Springer. https://doi.org/10.1007/978-3-031-43075-6_34

Nitay Calderon and Roi Reichart. 2024. On behalf of the stakeholders: Trends in NLP model interpretability in the era of LLMs. *arXiv preprint arXiv:2407.19200*.

Edward L. Campbell, Laura Docío Fernández, Javier Jiménez Raboso, and Carmen García-Mateo. 2021. Alzheimer's dementia detection from audio and language modalities in spontaneous speech. In *IberSPEECH*. https://doi.org/10.21437/IberSPEECH.2021-57

Sergio Carlomagno, Anna Santoro, Antonella Menditti, Maria Pandolfi, and Andrea Marini. 2005. Referential communication in Alzheimer's type dementia. *Cortex*, 41(4):520–534. https://doi.org/10.1016/S0010-9452(08)70192-8, PubMed: 16042028

Filippo Casu, Enrico Grosso, Andrea Lagorio, and Giuseppe A. Trunfio. 2024. Optimizing and evaluating pre-trained large language models for Alzheimer's disease detection. In *2024 32nd Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pages 277–284. IEEE. https://doi.org/10.1109/PDP62718.2024.00046

Zhengping Che, Yu Cheng, Shuangfei Zhai, Zhaonan Sun, and Yan Liu. 2017. Boosting deep learning risk prediction with generative adversarial networks for electronic health records. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 787–792. IEEE. https://doi.org/10.1109/ICDM.2017.93

Jun Chen, Ji Zhu, and Jieping Ye. 2019. An attention-based hybrid network for automatic detection of Alzheimer's disease from narrative speech. *Interspeech*. https://doi.org/10.21437/Interspeech.2019-2872

Liu Chen, Hiroko H. Dodge, and Meysam Asgari. 2020. Topic-based measures of conversation for detecting mild cognitive impairment. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, volume 2020, page 63. NIH Public Access. https://doi.org/10.18653/v1/2020.nlpmc-1.9, PubMed: 33642674

Minchuan Chen, Chenfeng Miao, Jun Ma, Shaojun Wang, and Jing Xiao. 2023a. Exploring multi-task learning and data augmentation in dementia detection with self-supervised pretrained models. *Interspeech*. https://doi.org/10.21437/Interspeech.2023-1623

Zhaoyi Chen, Hansi Zhang, Xi Yang, Songzi Wu, Xing He, Jie Xu, Jingchuan Guo, Mattia Prosperi, Fei Wang, Hua Xu, Yong Chen, Hui Hu, Steven T. DeKosky, Matthew Farrer, Yi Guo, Yonghui Wu, and Jiang Bian. 2023b. Assess the documentation of cognitive tests and biomarkers in electronic health records via natural language processing for Alzheimer's disease and related dementias. *International Journal of Medical Informatics*, 170:104973. https://doi.org/10.1016/j.ijmedinf.2022.104973, PubMed: 36577203

Hamidreza Chinaei, Leila Chan Currie, Andrew Danks, Hubert Lin, Tejas Mehta, and Frank Rudzicz. 2017. Identifying and avoiding confusion in dialogue with people with Alzheimer's disease. *Computational Linguistics*, 43(2):377–406. https://doi.org/10.1162/COLI_a_00290

Bharath Chintagunta, Namit Katariya, Xavier Amatriain, and Anitha Kannan. 2021. Medically aware GPT-3 as a data generator for medical dialogue summarization. In *Machine Learning for Healthcare Conference*, pages 354–372. PMLR. https://doi.org/10.18653/v1/2021.nlpmc-1.9

Jinho D. Choi, Mengmei Li, Felicia Goldstein, and Ihab Hajjar. 2019. Meta-semantic representation for early detection of Alzheimer's disease. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 82–91. https://doi.org/10.18653/v1/W19-3309

D. G. Clark, V. G. Wadley, P. Kapur, T. P. DeRamus, B. Singletary, A. P. Nicholas, P. D. Blanton, K. Lokken, H. Deshpande, D. Marson, and G. Deutsch. 2014. Lexical factors and cerebral regions influencing verbal fluency performance in MCI. *Neuropsychologia*, 54:98–111. https://doi.org/10.1016/j.neuropsychologia.2013.12.010, PubMed: 24384308

Natasha Clarke, Thomas R. Barrick, and Peter Garrard. 2021. A comparison of connected speech tasks for detecting early Alzheimer's disease and mild cognitive impairment using natural language processing and machine learning. *Frontiers in Computer Science*, 3:634360. https://doi.org/10.3389/fcomp.2021.634360

Natasha Clarke, Peter Foltz, and Peter Garrard. 2020. How to do things with (thousands of) words: Computational approaches to discourse analysis in Alzheimer's disease. *Cortex*, 129:446–463. https://doi.org/10.1016/j.cortex.2020.05.001, PubMed: 32622173

Trevor Cohen and Serguei Pakhomov. 2020. A tale of two perplexities: Sensitivity of neural language models to lexical retrieval deficits in dementia of the Alzheimer's type. *arXiv preprint arXiv:2005.03593*. https://doi.org/10.18653/v1/2020.acl-main.176

Ziyun Cui, Wen Wu, Wei-Qiang Zhang, Ji Wu, and Chao Zhang. 2023. Transferring speech-generic and depression-specific knowledge for Alzheimer's disease detection. In *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 1–8. IEEE. https://doi.org/10.1109/ASRU57964.2023.10389785

Nicholas Cummins, Yilin Pan, Zhao Ren, Julian Fritsch, Venkata Srikanth Nallanthighal, Heidi Christensen, Daniel Blackburn, Björn W. Schuller, Mathew Magimai Doss, Helmer Strik, and Aki Härmä. 2020. A comparison of acoustic and linguistics methodologies for Alzheimer's dementia recognition. *Interspeech*. https://doi.org/10.21437/Interspeech.2020-2635

Louis Daudet, Nikhil Yadav, Matthew Perez, Christian Poellabauer, Sandra Schneider, and Alan Huebner. 2016. Portable mTBI assessment using temporal and frequency analysis of speech. *IEEE Journal of Biomedical and Health Informatics*, 21(2):496–506. https://doi.org/10.1109/JBHI.2016.2633509, PubMed: 27913365

Nauman Dawalatabad, Yuan Gong, Sameer Khurana, Rhoda Au, and James Glass. 2022. Detecting dementia from long neuropsychological interviews. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5270–5283. https://doi.org/10.18653/v1/2022.findings-emnlp.386

Sofia De la Fuente Garcia, Craig W. Ritchie, and Saturnino Luz. 2020. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: A systematic review. *Journal of Alzheimer's Disease*, 78(4):1547–1574. https://doi.org/10.3233/JAD-200888, PubMed: 33185605

Hongshun Deng, Hongqing Liu, Yi Zhou, and Gan Lu. 2022. Alzheimer's disease detection using acoustic and linguistic features. In *2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, pages 2280–2284. IEEE. https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00337

Aniket Dey and Sanam Mittal. 2022. An integrated approach to non-invasive diagnosis of dementia using natural language processing and machine learning. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, pages 75–79. IEEE. https://doi.org/10.1109/ICDSCA56264.2022.9987931

Flavio Di Palo and Natalie Parde. 2019. Enriching neural models with targeted features for dementia detection. *ACL 2019*, page 302. https://doi.org/10.18653/v1/P19-2042

Zhongren Dong, Zixing Zhang, Weixiang Xu, Jing Han, Jianjun Ou, and Björn W. Schuller. 2024. Hafformer: A hierarchical attention-free framework for Alzheimer's disease detection from spontaneous speech. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11246–11250. IEEE. https://doi.org/10.1109/ICASSP48485.2024.10446795

Grazia D'Onofrio, Daniele Sancarlo, Francesco Ricciardi, Francesco Panza, Davide Seripa, Filippo Cavallo, Francesco Giuliani, and Antonio Greco. 2017. Information and communication technologies for the activities of daily living in older patients with dementia: A systematic review. *Journal of Alzheimer's Disease*, 57(3):927–935. https://doi.org/10.3233/JAD-161145, PubMed: 28304297

Jill A. Dosso, Jaya N. Kailley, and Julie M. Robillard. 2024. What does chatgpt know about dementia? A comparative analysis of information quality. *Journal of Alzheimer's Disease*, (Preprint):1–7. https://doi.org/10.3233/JAD-230573, PubMed: 38143345

Khadim Drame, Gayo Diallo, and Fleur Mougin. 2012. Towards a bilingual Alzheimer's disease terminology acquisition using a parallel corpus. In *Quality of Life Through Quality of Information*, pages 179–183. IOS Press.

Junwen Duan, Fangyuan Wei, Jin Liu, Hongdong Li, Tianming Liu, and Jianxin Wang. 2023. Cda: A contrastive data augmentation method for Alzheimer's disease detection. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1819–1826. https://doi.org/10.18653/v1/2023.findings-acl.114

Edith Cowan University. 2024. Language barriers could contribute to higher aggression in people with dementia. https://www.sciencedaily.com/releases/2024/02/240209134456.htm. Accessed: 2025-05-26.

Erik Edwards, Charles Dognin, Bajibabu Bollepalli, Maneesh Kumar Singh, and Verisk Analytics. 2020. Multiscale system for Alzheimer's dementia recognition through spontaneous speech. In *Interspeech*, pages 2197–2201. https://doi.org/10.21437/Interspeech.2020-2781

Malikeh Ehghaghi, Frank Rudzicz, and Jekaterina Novikova. 2022. Data-driven approach to differentiating between depression and dementia from noisy speech and language data. *arXiv preprint arXiv:2210.03303*.

Shaker El-Sappagh, Jose M. Alonso-Moral, Tamer Abuhmed, Farman Ali, and Alberto Bugarín-Diz. 2023. Trustworthy artificial intelligence in Alzheimer's disease: State of the art, opportunities, and challenges. *Artificial Intelligence Review*, 56(10):11149–11296. https://doi.org/10.1007/s10462-023-10415-5

Shaker El-Sappagh, Hager Saleh, Farman Ali, Eslam Amer, and Tamer Abuhmed. 2022. Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time. *Neural Computing and Applications*, 34(17):14487–14509. https://doi.org/10.1007/s00521-022-07263-9

Margi Engineer, Sushant Kot, and Emma Dixon. 2023. Investigating the readability and linguistic, psychological, and emotional characteristics of digital dementia information written in the english language: Multitrait-multimethod text analysis. *JMIR Formative Research*, 7:e48143. https://doi.org/10.2196/48143, PubMed: 37878351

Isabel Espinosa-Zaragoza, José Abreu-Salas, Elena Lloret, Paloma Moreda Pozo, and Manuel Palomar. 2023. A review of research-based automatic text simplification tools. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 321–330. https://doi.org/10.26615/978-954-452-092-2_036

Ben Eyre, Aparna Balagopalan, and Jekaterina Novikova. 2020. Fantastic features and where to find them: Detecting cognitive impairment with a subsequence classification guided approach. In *Proceedings of the Sixth Workshop on Noisy User-Generated Text (W-NUT 2020)*, pages 193–199. https://doi.org/10.18653/v1/2020.wnut-1.25

Ali Pourramezan Fard, Mohammad H. Mahoor, Muath Alsuhaibani, and Hiroko H. Dodge. 2024. Linguistic-based mild cognitive impairment detection using informative loss. *Computers in Biology and Medicine*, 176:108606. https://doi.org/10.1016/j.compbiomed.2024.108606, PubMed: 38763068

Shahla Farzana, Ashwin Deshpande, and Natalie Parde. 2022. How you say it matters: Measuring the impact of verbal disfluency tags on automated dementia detection. In *Proceedings of the 21st Workshop on Biomedical Language Processing*, pages 37–48. https://doi.org/10.18653/v1/2022.bionlp-1.4

Shahla Farzana and Natalie Parde. 2020. Exploring mmse score prediction using verbal and non-verbal cues. In *Interspeech*, pages 2207–2211. https://doi.org/10.21437/Interspeech.2020-3085

Shahla Farzana and Natalie Parde. 2022. Are interaction patterns helpful for task-agnostic dementia detection? An empirical exploration. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 172–182. https://doi.org/10.18653/v1/2022.sigdial-1.18

Shahla Farzana and Natalie Parde. 2023. Towards domain-agnostic and domain-adaptive dementia detection from spoken language. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11965–11978. https://doi.org/10.18653/v1/2023.acl-long.668

Shahla Farzana, Edoardo Stoppa, Alex Leow, Tamar Gollan, Raeanne Moore, David Salmon, Douglas Galasko, Erin Sundermann, and Natalie Parde. 2024. Slacad: A spoken language corpus for early Alzheimer's disease detection. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 14877–14897.

Shahla Farzana, Mina Valizadeh, and Natalie Parde. 2020. Modeling dialogue in conversational cognitive health screening interviews. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1167–1177.

Anna Favaro, Najim Dehak, Thomas Thebaud, Jesús Villalba, Esther Oh, and Laureano Moro-Velázquez. 2024. Discovering invariant patterns of cognitive decline via an automated analysis of the cookie thief picture description task. In *Proceedings of The Speaker and Language Recognition Workshop (Odyssey 2024)*, pages 201–208. https://doi.org/10.21437/odyssey.2024-29

Kathleen C. Fraser, Nicklas Linz, Bai Li, Kristina Lundholm Fors, Frank Rudzicz, Alexandra König, Jan Alexandersson, Philippe Robert, and Dimitrios Kokkinakis. 2019. Multilingual prediction of Alzheimer's disease through domain adaptation and concept-based language modelling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and*

Short Papers), pages 3659–3670. https://doi.org/10.18653/v1/N19-1367

Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2016a. Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422. https://doi.org/10.3233/JAD-150520, PubMed: 26484921

Kathleen C. Fraser, Frank Rudzicz, and Graeme Hirst. 2016b. Detecting late-life depression in Alzheimer's disease through analysis of speech and language. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 1–11. https://doi.org/10.18653/v1/W16-0301

Kathleen C. Fraser, Frank Rudzicz, and Elizabeth Rochon. 2013. Using text and acoustic features to diagnose progressive aphasia and its subtypes. In *Interspeech*, pages 2177–2181. https://doi.org/10.21437/Interspeech.2013-514

Julian Fritsch, Sebastian Wankerl, and Elmar Nöth. 2019. Automatic diagnosis of Alzheimer's disease using neural network language models. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5841–5845. IEEE. https://doi.org/10.1109/ICASSP.2019.8682690

Gloria Gagliardi. 2024. Natural language processing techniques for studying language in pathological ageing: A scoping review. *International Journal of Language & Communication Disorders*, 59(1):110–122. https://doi.org/10.1111/1460-6984.12870, PubMed: 36960885

Muskan Garg and Sunghwan Sohn. 2023. Cared: Caregiver's experience with cognitive decline in reddit posts. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, pages 581–587. IEEE. https://doi.org/10.1109/ICHI57859.2023.00104, PubMed: 38384500

Sara Gerke, Timo Minssen, and Glenn Cohen. 2020. Ethical and legal challenges of artificial intelligence-driven healthcare, *Artificial Intelligence in Healthcare*, pages 295–336. Elsevier. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Shadi Gholizadeh, Amy K. Freeman, and Mark P. Botticelli. 2023. Conversational artificial intelligence for people living with dementia and their care partners: A scoping review. *Alzheimer's & Dementia*, 19:e071965. https://doi.org/10.1002/alz.071965

Elizabeth S. Gilman, Sushant Kot, Margi Engineer, and Emma Dixon. 2024. Training adults with mild to moderate dementia in chatGPT: Exploring best practices. In *Companion Proceedings of the 29th International Conference on Intelligent User Interfaces*, pages 101–106. https://doi.org/10.1145/3640544.3645230

Dimitris Gkoumas, Matthew Purver, and Maria Liakata. 2023. Reformulating nlp tasks to capture longitudinal manifestation of language disorders in people with dementia. *arXiv preprint arXiv:2310.09897*. https://doi.org/10.18653/v1/2023.emnlp-main.986

Dimitris Gkoumas, Bo Wang, Adam Tsakalidis, Maria Wolters, Matthew Purver, Arkaitz Zubiaga, and Maria Liakata. 2024. A longitudinal multi-modal dataset for dementia monitoring and diagnosis. *Language Resources and Evaluation*, pages 1–20. https://doi.org/10.1007/s10579-023-09718-4, PubMed: 39323983

Luka Gligic, Andrey Kormilitzin, Paul Goldberg, and Alejo Nevado-Holgado. 2020. Named entity recognition in electronic health records using transfer learning bootstrapped neural networks. *Neural Networks*, 121:132–139. https://doi.org/10.1016/j.neunet.2019.08.032, PubMed: 31541881

Lucía Gómez-Zaragozá, Simone Wills, Cristian Tejedor-Garcia, Javier Marín-Morales, Mariano Alcañiz, and Helmer Strik. 2023. Alzheimer disease classification through asr-based transcriptions: Exploring the impact of punctuation and pauses. *arXiv preprint arXiv:2306.03443*. https://doi.org/10.21437/Interspeech.2023-1734

Míriam Gónzalez Atienza, José Andrés González López, and Antonio M. Peinado. 2021. An automatic system for dementia detection using acoustic and linguistic features. ISCA. https://doi.org/10.21437/IberSPEECH.2021-56

Nancy Green, Curry I. Guinn, and Ronnie W. Smith. 2012. Assisting social conversation between persons with Alzheimer's disease and their conversational partners. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies*, pages 37–46.

Yue Guo, Changye Li, Carol Roan, Serguei Pakhomov, and Trevor Cohen. 2021. Crossing the ''cookie theft'' corpus chasm: Applying what bert learns from outside data to the adress challenge dementia detection task. *Frontiers in Computer Science*, 3:642517. https://doi.org/10.3389/fcomp.2021.642517, PubMed: 40535703

Zhiqiang Guo, Zhaoci Liu, Zhenhua Ling, Shijin Wang, Lingjing Jin, and Yunxia Li. 2020. Text classification by contrastive learning and cross-lingual data augmentation for Alzheimer's disease detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6161–6171. https://doi.org/10.18653/v1/2020.coling-main.542

Priyanka Gupta, Pankaj Malhotra, Jyoti Narwariya, Lovekesh Vig, and Gautam Shroff. 2020. Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research*, 4(2):112–137. https://doi.org/10.1007/s41666-019-00062-3, PubMed: 35415440

Christopher A. Hane, Vijay S. Nori, William H. Crown, Darshak M. Sanghavi, and Paul Bleicher. 2020. Predicting onset of dementia using clinical notes and machine learning: Case-control study. *JMIR Medical Informatics*, 8(6):e17819. https://doi.org/10.2196/17819, PubMed: 32490841

Stefan Harrer. 2023. Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*, 90. https://doi.org/10.1016/j.ebiom.2023.104512, PubMed: 36924620

Nobuko Hasegawa. 1985. On the so-called ''zero pronouns'' in japanese. https://doi.org/10.1515/tlir.1985.4.4.289

R'mani Haulcy and James Glass. 2021. Classifying Alzheimer's disease using audio and text-based representations of speech. *Frontiers in Psychology*, 11:624137. https://doi.org/10.3389/fpsyg.2020.624137, PubMed: 33519651

Jonathan Heitz, Gerold Schneider, and Nicolas Langer. 2024. The influence of automatic speech recognition on linguistic features and automatic Alzheimer's disease detection from spontaneous speech. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15955–15969.

Pamela Herd, Deborah Carr, and Carol Roan. 2014. Cohort profile: Wisconsin longitudinal study (wls). *International Journal of Epidemiology*, 43(1):34–41. https://doi.org/10.1093/ije/dys194, PubMed: 24585852

Mrutyunjaya S. Hiremath and Rajashekhar C. Biradar. 2023. Early stage detection of Alzheimer's using hybrid artificial intelligence model: A review. In *2023 3rd International Conference on Intelligent Technologies (CONIT)*, pages 1–6. IEEE. https://doi.org/10.1109/CONIT59222.2023.10205681

Anna Hlédiková, Dominika Woszczyk, Alican Akman, Soteris Demetriou, and Björn Schuller. 2022. Data augmentation for dementia detection in spoken language. *arXiv preprint arXiv:2206.12879*.

Junyuan Hong, Wenqing Zheng, Han Meng, Siqi Liang, Anqing Chen, Hiroko H. Dodge, Jiayu Zhou, and Zhangyang Wang. 2024. A-conect: Designing AI-based conversational chatbot for early dementia intervention. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*.

Sheng-Yi Hong, Li-Hung Yao, Wen-Ting Cheah, Wei-Der Chang, Li-Chen Fu, and Yu-Ling Chang. 2019. A novel screening system for Alzheimer's disease based on speech transcripts using neural network. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 2440–2445. IEEE. https://doi.org/10.1109/SMC.2019.8914628

Toshiro Horigome, Kimihiro Hino, Hiroyoshi Toyoshiba, Norihisa Shindo, Kei Funaki, Yoko Eguchi, Momoko Kitazawa, Takanori Fujita, Masaru Mimura, and Taishiro Kishimoto. 2022. Identifying neurocognitive disorder using vector representation of free conversation. *Scientific Reports*, 12(1):12461. https://doi.org/10.1038/s41598-022-16204-4, PubMed: 35922457

Vagelis Hristidis, Nicole Ruggiano, Ellen L. Brown, Sai Rithesh Reddy Ganta, and Selena Stewart. 2023. ChatGPT vs Google for queries related to dementia and other cognitive decline: Comparison of results. *Journal of Medical Internet Research*, 25:e48966. https://doi.org/10.2196/48966, PubMed: 37490317

Si-Sheng Huang. 2022. Depression among caregivers of patients with dementia: Associative factors and management approaches. *World Journal of Psychiatry*, 12(1):59. https://doi.org/10.5498/wjp.v12.i1.59, PubMed: 35111579

Loukas Ilias and Dimitris Askounis. 2022a. Explainable identification of dementia from transcripts using transformer networks. *IEEE Journal of Biomedical and Health Informatics*, 26(8):4153–4164. https://doi.org/10.1109/JBHI.2022.3172479, PubMed: 35511841

Loukas Ilias and Dimitris Askounis. 2022b. Multimodal deep learning models for detecting dementia from speech and transcripts. *Frontiers in Aging Neuroscience*, 14:830943. https://doi.org/10.3389/fnagi.2022.830943, PubMed: 35370608

Clifford R. Jack, David S. Knopman, William J. Jagust, Leslie M. Shaw, Paul S. Aisen, Michael W. Weiner, Ronald C. Petersen, and John Q. Trojanowski. 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128. https://doi.org/10.1016/S1474-4422(09)70299-6, PubMed: 20083042

William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar. 2014. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 27–37. https://doi.org/10.3115/v1/W14-3204

Ashir Javeed, Ana Luiza Dallora, Johan Sanmartin Berglund, Arif Ali, Liaqata Ali, and Peter Anderberg. 2023. Machine learning for dementia prediction: A systematic review and future research directions. *Journal of Medical Systems*, 47(1):17. https://doi.org/10.1007/s10916-023-01906-7, PubMed: 36720727

Md Shariar Kabir, Jakia Khanom, Md Atik Bhuiyan, Zerin Nasrin Tumpa, Sk Fazlee Rabby, and Saurabh Bilgaiyan. 2023. The early detection of dementia disease using machine learning approach. In *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6. IEEE. https://doi.org/10.1109/ICCCI56745.2023.10128411

Edith Kaplan, Harold Goodglass, and Sandra Weintraub. 2001. Boston naming test. *The Clinical Neuropsychologist*.

Sweta Karlekar, Tong Niu, and Mohit Bansal. 2018. Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models. *arXiv preprint arXiv:1804.06440*. https://doi.org/10.18653/v1/N18-2110

Kadircan H. Keskinbora. 2019. Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, 64:277–282. https://doi.org/10.1016/j.jocn.2019.03.001, PubMed: 30878282

Yusera Farooq Khan, Baijnath Kaushik, Mohammad Khalid Imam Rahmani, and Md Ezaz Ahmed. 2022a. Hsi-lfs-bert: Novel hybrid swarm intelligence based linguistics feature selection and computational intelligent model for Alzheimer's prediction using audio transcript. *IEEE Access*, 10:126990–127004. https://doi.org/10.1109/ACCESS.2022.3223681

Yusera Farooq Khan, Baijnath Kaushik, Mohammad Khalid Imam Rahmani, and Md Ezaz Ahmed. 2022b. Stacked deep dense neural network model to predict Alzheimer's dementia using audio transcript data. *IEEE Access*, 10:32750–32765. https://doi.org/10.1109/ACCESS.2022.3161749

Hana Kim, Argye E. Hillis, and Charalambos Themistocleous. 2024. Machine learning classification of patients with amnestic mild cognitive impairment and non-amnestic mild cognitive impairment from written picture description tasks. *Brain Sciences*, 14(7):652. https://doi.org/10.3390/brainsci14070652, PubMed: 39061392

Ari Z. Klein, Arjun Magge, Karen O'Connor, and Graciela Gonzalez-Hernandez. 2022. Automatically identifying twitter users for interventions to support dementia family caregivers: Annotated data set and benchmark classification models. *JMIR Aging*, 5(3):e39547. https://doi.org/10.2196/39547, PubMed: 36112408

Philipp Klumpp, Julian Fritsch, and Elmar Nöth. 2018. Ann-based Alzheimer's disease classification from bag of words. In *Speech Communication; 13th ITG-Symposium*, pages 1–4. VDE.

Shunsuke Koga, Nicholas B. Martin, and Dennis W. Dickson. 2024. Evaluating the performance of large language models: ChatGPT and google bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders. *Brain Pathology*, 34(3):e13207. https://doi.org/10.1111/bpa.13207, PubMed: 37553205

Junghyun Koo, Jie Hwan Lee, Jaewoo Pyo, Yujin Jo, and Kyogu Lee. 2020. Exploiting multi-modal features from pre-trained networks for Alzheimer's dementia recognition. *arXiv preprint arXiv:2009.04070*.

Ioannis-Aris Kostis, Konstantinos Karamitsios, Konstantinos Kotrotsios, Magda Tsolaki, and Anthoula Tsolaki. 2022. AI-enabled conversational agents in service of mild cognitive impairment patients. In *2022 International Conference on Electrical and Information Technology (IEIT)*, pages 69–74. IEEE. https://doi.org/10.1109/IEIT56384.2022.9967865

Gershon Kofi Ladzekpo, Collins Kwabla Amekor, and Margaret Akrobotu. 2023. Language and communication in the digital age: The study of how new technologies and digital media are affecting language use, communication patterns, and sociolinguistic dynamics. *Journal of Literature and Linguistics Studies*, 1(1):24–31.

Mily Lal, Manisha Bhende, Akanksha Goel, Poi Tamrakar, and Saurabh Saoji. 2023. A hybrid deep learning approach for sentiment analysis of dementia care. In *2023 IEEE Engineering Informatics*, pages 1–7. IEEE. https://doi.org/10.1109/IEEECONF58110.2023.10520374

Alyssa M. Lanzi, Anna K. Saylor, Davida Fromm, Houjun Liu, Brian MacWhinney, and Matthew L. Cohen. 2023. Dementiabank: Theoretical rationale, protocol, and illustrative analyses. *American Journal of Speech-Language Pathology*, 32(2):426–438. https://doi.org/10.1044/2022_AJSLP-22-00281, PubMed: 36791255

Atif Latif and Jihie Kim. 2024. Evaluation and analysis of large language models for clinical text augmentation and generation. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3384496

John Laurentiev, Dae Hyun Kim, Mufaddal Mahesri, Kuan-Yuan Wang, Lily G. Bessette, Cassandra York, Heidi Zakoul, Su Been Lee, Li Zhou, and Kueiyu Joshua Lin. 2024. Identifying functional status impairment in people living with dementia through natural language processing of clinical documents: Cross-sectional study. *Journal of Medical Internet Research*, 26:e47739. https://doi.org/10.2196/47739, PubMed: 38349732

Maider Lehr, Emily Tucker Prud'hommeaux, Izhak Shafran, and Brian Roark. 2012. Fully automated neuropsychological assessment for detecting mild cognitive impairment. In *Interspeech*, volume 2012. https://doi.org/10.21437/Interspeech.2012-306

Maider Lehr, Izhak Shafran, Emily Prud'Hommeaux, and Brian Roark. 2013. Discriminative joint modeling of lexical variation and acoustic confusion for automated narrative retelling assessment. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 211–220.

Bai Li, Yi-Te Hsu, and Frank Rudzicz. 2019. Detecting dementia in mandarin chinese using transfer learning from a parallel corpus. *arXiv preprint arXiv:1903.00933*.

Changye Li, David Knopman, Weizhe Xu, Trevor Cohen, and Serguei Pakhomov. 2022a. GPT-d: Inducing dementia-related linguistic anomalies by deliberate degradation of artificial neural language models. *arXiv preprint arXiv:2203.13397*.

Changye Li, Jacob Solinsky, Trevor Cohen, and Serguei Pakhomov. 2024a. A curious case of retrogenesis in language: Automated analysis of language patterns observed in dementia patients and young children. *Neuroscience Informatics*, 4(1):100155. https://doi.org/10.1016/j.neuri.2023.100155, PubMed: 38433986

Changye Li, Weizhe Xu, Trevor Cohen, and Serguei Pakhomov. 2024b. Useful blunders: Can automated speech recognition errors improve downstream dementia classification? *Journal of Biomedical Informatics*, 150:104598. https://doi.org/10.1016/j.jbi.2024.104598, PubMed: 38253228

Jinpeng Li and Wei-Qiang Zhang. 2024. Whisper-based transfer learning for alzheimer disease classification: Leveraging speech segments with full transcripts as prompts. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11211–11215. IEEE. https://doi.org/10.1109/ICASSP48485.2024.10448004

Renxuan Albert Li, Ihab Hajjar, Felicia Goldstein, and Jinho D. Choi. 2020. Analysis of hierarchical multi-content text classification model on b-sharp dataset for early detection of Alzheimer's disease. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 358–365. https://doi.org/10.18653/v1/2020.aacl-main.38

Rumeng Li, Xun Wang, and Hong Yu. 2023. Two directions for clinical data generation with large language models: Data-to-label and label-to-data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2023, page 7129. NIH Public Access. https://doi.org/10.18653/v1/2023.findings-emnlp.474, PubMed: 38213944

Yuanchao Li, Catherine Lai, Divesh Lala, Koji Inoue, and Tatsuya Kawahara. 2022b. Alzheimer's dementia detection through spontaneous dialogue with proactive robotic listeners. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 875–879. IEEE.

Kaiying Lin and Peter Y. Washington. 2024. Multimodal deep learning for dementia classification using text and audio. *Scientific Reports*, 14(1):13887. https://doi.org/10.1038/s41598-024-64438-1, PubMed: 38880810

Hali Lindsay, Johannes Tröger, and Alexandra König. 2021. Language impairment in Alzheimer's disease–robust and explainable evidence for ad-related deterioration of spontaneous speech through multilingual machine learning. *Frontiers in Aging Neuroscience*, 13:642033. https://doi.org/10.3389/fnagi.2021.642033, PubMed: 34093165

Ming Liu, Richard Beare, Taya Collyer, Nadine Andrew, and Velandai Srikanth. 2023a. Leveraging natural language processing and clinical notes for dementia detection. In *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 150–155. https://doi.org/10.18653/v1/2023.clinicalnlp-1.20

Ning Liu, Kexue Luo, Zhenming Yuan, and Yan Chen. 2022a. A transfer learning method for detecting Alzheimer's disease based on speech and natural language processing. *Frontiers in Public Health*, 10:772592. https://doi.org/10.3389/fpubh.2022.772592, PubMed: 35493375

Ning Liu and Lingxing Wang. 2023. An approach for assisting diagnosis of Alzheimer's disease based on natural language processing. *Frontiers in Aging Neuroscience*, 15:1281726. https://doi.org/10.3389/fnagi.2023.1281726, PubMed: 38035270

Ning Liu, Zhenming Yuan, and Qingfeng Tang. 2022b. Improving Alzheimer's disease detection for speech based on feature purification network. *Frontiers in Public Health*, 9:835960. https://doi.org/10.3389/fpubh.2021.835960, PubMed: 35310782

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ziming Liu, Muskan Garg, Sunyang Fu, Surjodeep Sarkar, Maria Vassilaki, Ronald C. Petersen, Jennifer St Sauver, and Sunghwan Sohn. 2023b. Harnessing transfer learning for dementia prediction: Leveraging sex-different mild cognitive impairment prognosis. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2097–2100. IEEE. `https://doi.org/10.1109/BIBM58861.2023.10385516`

Ziming Liu, Lauren Proctor, Parker N. Collier, and Xiaopeng Zhao. 2021. Automatic diagnosis and prediction of cognitive decline associated with Alzheimer's dementia through spontaneous speech. In *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 39–43. IEEE. `https://doi.org/10.1109/ICSIPA52582.2021.9576784`

Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney. 2021a. Detecting cognitive decline using speech only: The adresso challenge. *arXiv preprint arXiv:2104.09356*.

Saturnino Luz, Fasih Haider, Sofia de la Fuente Garcia, Davida Fromm, and Brian MacWhinney. 2021b. Alzheimer's dementia recognition through spontaneous speech.

Constantine G. Lyketsos and Hochang B. Lee. 2003. Diagnosis and treatment of depression in Alzheimer's disease: A practical update for the clinician. *Dementia and Geriatric Cognitive disorders*, 17(1–2):55–64. `https://doi.org/10.1159/000074277`, PubMed: 14564126

Laura C. Maclagan, Mohamed Abdalla, Daniel A. Harris, Therese A. Stukel, Branson Chen, Elisa Candido, Richard H. Swartz, Andrea Iaboni, R. Liisa Jaakkimainen, and Susan E. Bronskill. 2023. Can patients with dementia be identified in primary care electronic medical records using natural language processing? *Journal of Healthcare Informatics Research*, 7(1):42–58. `https://doi.org/10.1007/s41666-023-00125-6`, PubMed: 36910911

Pranav Mahajan and Veeky Baths. 2021. Acoustic and language based deep learning approaches for Alzheimer's dementia detection from spontaneous speech. *Frontiers in Aging Neuroscience*, 13:623607. `https://doi.org/10.3389/fnagi.2021.623607`, PubMed: 33613269

Chengsheng Mao, Jie Xu, Luke Rasmussen, Yikuan Li, Prakash Adekkanattu, Jennifer Pacheco, Borna Bonakdarpour, Robert Vassar, Li Shen, Guoqian Jiang, et al. 2023. Ad-BERT: Using pre-trained language model to predict the progression from mild cognitive impairment to Alzheimer's disease. *Journal of Biomedical Informatics*, 144:104442. `https://doi.org/10.1016/j.jbi.2023.104442`, PubMed: 37429512

Matej Martinc, Fasih Haider, Senja Pollak, and Saturnino Luz. 2021. Temporal integration of text transcripts and acoustic features for Alzheimer's diagnosis based on spontaneous speech. *Frontiers in Aging Neuroscience*, 13:642647. `https://doi.org/10.3389/fnagi.2021.642647`, PubMed: 34194313

Matej Martinc and Senja Pollak. 2020. Tackling the adress challenge: A multimodal approach to the automated recognition of Alzheimer's dementia. In *Interspeech*, pages 2157–2161. `https://doi.org/10.21437/Interspeech.2020-2202`

Vaden Masrani, Gabriel Murray, Thalia Field, and Giuseppe Carenini. 2017. Detecting dementia through retrospective analysis of routine blog posts by bloggers with dementia. In *BioNLP 2017*, pages 232–237. `https://doi.org/10.18653/v1/W17-2329`

Lovro Matošević and Alan Jović. 2022. Accurate detection of dementia from speech transcripts using roberta model. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, pages 1478–1484. IEEE. `https://doi.org/10.23919/MIPRO55190.2022.9803462`

Guy M. McKhann, David S. Knopman, Howard Chertkow, Bradley T. Hyman, Clifford R.

Jack Jr., Claudia H. Kawas, William E. Klunk, Walter J. Koroshetz, Jennifer J. Manly, Richard Mayeux, Richard C. Mohs, John C. Morris, Martin N. Rossor, Philip Scheltens, Maria C. Carrillo, Bill Thies, Sandra Weintraub, and Creighton H. Phelps. 2011. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):263–269. https://doi.org/10.1016/j.jalz.2011.03.005, PubMed: 21514250

Thomas Melistas, Lefteris Kapelonis, Nikos Antoniou, Petros Mitseas, Dimitris Sgouropoulos, Theodoros Giannakopoulos, Athanasios Katsamanis, Shrikanth Narayanan, and NCSR Demokritos. 2023. Cross-lingual features for Alzheimer's dementia detection from speech. *Interspeech*. https://doi.org/10.21437/Interspeech.2023-1934

Helen Meng, Brian Mak, Man-Wai Mak, Helene Fung, Xianmin Gong, Timothy Kwok, Xunying Liu, Vincent Mok, Patrick Wong, Jean Woo, et al. 2023. Integrated and enhanced pipeline system to support spoken language analytics for screening neurocognitive disorders. In *Proceedings of InterSpeech*, pages 1713–1717. https://doi.org/10.21437/Interspeech.2023-2249

Mario Merone, Sebastian Luca D'Addario, Pierandrea Mirino, Francesca Bertino, Cecilia Guariglia, Rossella Ventura, Adriano Capirchio, Gianluca Baldassarre, Massimo Silvetti, and Daniele Caligiore. 2022. A multi-expert ensemble system for predicting alzheimer transition using clinical features. *Brain Informatics*, 9(1):20. https://doi.org/10.1186/s40708-022-00168-2, PubMed: 36056985

Bahman Mirheidari, Daniel Blackburn, and Heidi Christensen. 2022. Automatic cognitive assessment: Combining sparse datasets with disparate cognitive scores. In *Interspeech*, pages 2463–2467. https://doi.org/10.21437/Interspeech.2022-10205

Bahman Mirheidari, Daniel Blackburn, Markus Reuber, Traci Walker, and Heidi Christensen. 2016. Diagnosing people with dementia using automatic conversation analysis. In *Proceedings of Interspeech*, pages 1220–1224. ISCA. https://doi.org/10.21437/Interspeech.2016-857

Bahman Mirheidari, Daniel Blackburn, Traci Walker, Annalena Venneri, Markus Reuber, and Heidi Christensen. 2018. Detecting signs of dementia using word vector representations. In *Interspeech*, pages 1893–1897. https://doi.org/10.21437/Interspeech.2018-1764

Bahman Mirheidari, Yilin Pan, Daniel Blackburn, Ronan O'Malley, and Heidi Christensen. 2021. Identifying cognitive impairment using sentence representation vectors. In *Interspeech*, pages 2941–2945. https://doi.org/10.21437/Interspeech.2021-915

David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *Annals of Internal Medicine*, 151(4):264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135, PubMed: 19622511

Amir Abbas Tahami Monfared, Yaakov Stern, Stephen Doogan, Margaret Bray, and Quanwu Zhang. 2021. Understanding the impact of covid-19 pandemic on patients with Alzheimer's disease and caregivers using online narratives on social media. *Alzheimer's & Dementia*, 17:e055998. https://doi.org/10.1002/alz.055998

Kimberly D. Mueller, Bruce Hermann, Jonilda Mecollari, and Lyn S. Turkstra. 2018. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*, 40(9):917–939. https://doi.org/10.1080/13803395.2018.1446513, PubMed: 29669461

Hamish Naismith, Robert Howard, Robert Stewart, Alexandra Pitman, and Christoph Mueller. 2022. Suicidal ideation in dementia: Associations with neuropsychiatric symptoms and subtype diagnosis. *International Psychogeriatrics*, 34(4):399–406. https://doi.org/10.1017/S1041610222000126, PubMed: 35331357

Anjana S. Nambiar, Kanigolla Likhita, K. V. S. Sri Pujya, Deepa Gupta, Susmitha Vekkot, and S. Lalitha. 2022. Comparative study of deep classifiers for early dementia detection using speech transcripts. In *2022 IEEE 19th India Council International Conference (INDICON)*, pages 1–6. IEEE. https://doi.org/10.1109/INDICON56171.2022.10039705

Shamila Nasreen, Julian Hough, and Matthew Purver. 2021a. Rare-class dialogue act tagging for Alzheimer's disease diagnosis. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 290–300. https://doi.org/10.18653/v1/2021.sigdial-1.32

Shamila Nasreen, Morteza Rohanian, Julian Hough, and Matthew Purver. 2021b. Alzheimer's dementia recognition from spontaneous speech using disfluency and interactional features. *Frontiers in Computer Science*, 3:640669. https://doi.org/10.3389/fcomp.2021.640669

Congning Ni, Bradley Malin, Lijun Song, Angela Jefferson, Patricia Commiskey, and Zhijun Yin. 2022. ''rough day. . . need a hug'': Learning challenges and experiences of the Alzheimer's disease and related dementia caregivers on reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 711–722. https://doi.org/10.1609/icwsm.v16i1.19328

Jekaterina Novikova. 2021. Robustness and sensitivity of BERT models predicting Alzheimer's disease from text. *arXiv preprint arXiv:2109.11888*. https://doi.org/10.18653/v1/2021.wnut-1.37

Iris E. Nowenstein, Marija Stanojevic, Gunnar Örnólfsson, María Kristín Jónsdóttir, Bill Simpson, Jennifer Sorinas Nerin, Bryndís Bergþórsdóttir, Kristín Hannesdóttir, Jekaterina Novikova, and Jelena Curcic. 2024. Speech and language biomarkers of neurodegenerative conditions: Developing cross-linguistically valid tools for automatic analysis. In *Proceedings of the Fifth Workshop on Resources and Processing of Linguistic, Para-linguistic and Extra-linguistic Data from People with Various Forms of Cognitive/psychiatric/developmental impairments@ LREC-COLING 2024*, pages 26–33.

Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Judyanne Sharmini Gilbert Fernandez. 2016. Deep-deep neural network language models for predicting mild cognitive impairment. In *Advances in Bioinformatics and Artificial Intelligence 2016*, pages 14–20. Rheinisch-Westfaelische Technische Hochschule Aachen.

Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Karen Jennifer Golden. 2014. Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 78–87.

Moisés R. Pacheco-Lorenzo, Luis E. Anido-Rifón, Manuel J. Fernández-Iglesias, and Sonia M. Valladares-Rodríguez. 2024. Will senior adults accept being cognitively assessed by a conversational agent? A user-interaction pilot study. *Applied Intelligence*, pages 1–16. https://doi.org/10.1007/s10489-024-05558-z

Swati Padhee, Anurag Illendula, Megan Sadler, Valerie L. Shalin, Tanvi Banerjee, Krishnaprasad Thirunarayan, and William L. Romine. 2020. Predicting early indicators of cognitive decline from verbal utterances. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 477–480. IEEE. https://doi.org/10.1109/BIBM49941.2020.9313106

Yilin Pan, Bahman Mirheidari, Markus Reuber, Annalena Venneri, Daniel Blackburn, and Heidi Christensen. 2019. Automatic hierarchical attention neural network for detecting ad. In *Proceedings of Interspeech 2019*, pages 4105–4109. International Speech Communication Association (ISCA). https://doi.org/10.21437/Interspeech.2019-1799

Yilin Pan, Bahman Mirheidari, Zehai Tu, Ronan O'Malley, Traci Walker, Annalena Venneri, Markus Reuber, Daniel Blackburn, and Heidi Christensen. 2020. Acoustic feature extraction with interpretable deep neural network for neurodegenerative related disorder classification. In *Proceedings of Interspeech 2020*, pages 4806–4810. International Speech Communication Association (ISCA). https://doi.org/10.21437/Interspeech.2020-2684

Samin Panahi, Jamie Mayo, Eamonn Kennedy, Lee Christensen, Sreekanth Kamineni, Hari Krishna Raju Sagiraju, Tyler Cooper, David F. Tate, Randall Rupper, and Mary Jo Pugh. 2024. Identifying clinical phenotypes of frontotemporal dementia in post-9/11 era veterans using natural language processing. *Frontiers in Neurology*, 15:1270688. https://doi.org/10.3389/fneur.2024.1270688, PubMed: 38426171

Raghavendra Pappagari, Jaejin Cho, Sonal Joshi, Laureano Moro-Velázquez, Piotr Zelasko, Jesús Villalba, and Najim Dehak. 2021. Automatic detection and assessment of Alzheimer disease using speech and language technologies in low-resource scenarios. In *Interspeech*, volume 2021, pages 3825–3829. https://doi.org/10.21437/Interspeech.2021-1850

Raghavendra Pappagari, Jaejin Cho, Laureano Moro-Velazquez, and Najim Dehak. 2020. Using state of the art speaker recognition and natural language processing technologies to detect Alzheimer's disease and assess its severity. In *Interspeech*, pages 2177–2181. https://doi.org/10.21437/Interspeech.2020-2587

Bambang Parmanto, Bayu Aryoyudanta, Timothius Wilbert Soekinto, I. Made Agus Setiawan, Yuhan Wang, Haomin Hu, Andi Saptono, and Yong Kyung Choi. 2024. A reliable and accessible caregiving language model (calm) to support tools for caregivers: Development and evaluation study. *JMIR Formative Research*, 8:e54633. https://doi.org/10.2196/54633, PubMed: 39083337

Mahboobeh Parsapoor. 2023. AI-based assessments of speech and language impairments in dementia. *Alzheimer's & Dementia*, 19(10):4675–4687. https://doi.org/10.1002/alz.13395, PubMed: 37578167

Jay Patel and Huanmei Wu. 2024. Utilizing electronic dental records to predict neuro-degenerative diseases in a dental setting: A pilot study, *MEDINFO 2023–The Future Is Accessible*. IOS Press, pages 1322–1326. https://doi.org/10.3233/SHTI231179

Robert B. Penfold, David S. Carrell, David J. Cronkite, Chester Pabiniak, Tammy Dodd, Ashley M. H. Glass, Eric Johnson, Ella Thompson, H. Michael Arrighi, and Paul E. Stang. 2022. Development of a machine learning model to predict mild cognitive impairment using natural language processing in the absence of screening. *BMC Medical Informatics and Decision Making*, 22(1):129. https://doi.org/10.1186/s12911-022-01864-z, PubMed: 35549702

Beatriz Peres and Pedro F. Campos. 2024. A systematic review of reminder and guidance systems for Alzheimer's disease and related dementias patients: Context, barriers and facilitators. *Disability and Rehabilitation: Assistive Technology*, 19(6):2133–2146. https://doi.org/10.1080/17483107.2023.2277821, PubMed: 37987633

P. A. Pérez-Toro, T. Arias-Vergara, F. Braun, F. Hönig, C. A. Tobón-Quintero, D. Aguillón, F. Lopera, L. Hincapié-Henao, M. Schuster, K. Riedhammer, Andreas Maier, Elmar Noeth, and Juan Rafael Orozco-Arroyave. 2023. Automatic assessment of Alzheimer's across three languages using speech and language features. *Interspeech*. https://doi.org/10.21437/Interspeech.2023-2079

Paula Andrea Pérez-Toro, Sebastian P. Bayerl, Tomás Arias-Vergara, Juan Camilo Vásquez-Correa, Philipp Klumpp, Maria Schuster, Elmar Nöth, Juan Rafael Orozco-Arroyave, and Korbinian Riedhammer. 2021. Influence of the interviewer on the automatic assessment of Alzheimer's disease in the context of the adresso challenge. In *Interspeech*, pages 3785–3789. https://doi.org/10.21437/Interspeech.2021-1589

Paula Andrea Pérez-Toro, Philipp Klumpp, Abner Hernandez, Tomas Arias, Patricia Lillo, Andrea Slachevsky, Adolfo Martín García, Maria Schuster, Andreas K. Maier, Elmar Noeth, et al. 2022. Alzheimer's detection from english to spanish using acoustic and linguistic embeddings. In *Interspeech*, pages 2483–2487. https://doi.org/10.21437/Interspeech.2022-10883

Ronald Carl Petersen, Paul S. Aisen, Laurel A. Beckett, Michael C. Donohue, Anthony Collins Gamst, Danielle J. Harvey, C. R. Jack Jr., William J. Jagust, Leslie M. Shaw, Arthur W. Toga, J. Q. Trojanowski, and M. W. Weiner. 2010. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25, PubMed: 20042704

Ulla Petti, Simon Baker, and Anna Korhonen. 2020. A systematic literature review of automatic Alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association*, 27(11):1784–1797. https://doi.org/10.1093/jamia/ocaa174, PubMed: 32929494

Ulla Petti, Simon Baker, Anna Korhonen, and Jessica Robin. 2023a. The generalizability of longitudinal changes in speech before alzheimer's disease diagnosis. *Journal of Alzheimer's Disease*, 92(2):547–564. https://doi.org/10.3233/JAD-220847, PubMed: 36776053

Ulla Petti, Simon Baker, Anna Korhonen, and Jessica Robin. 2023b. How much speech data is needed for tracking language change in Alzheimer's disease? A comparison of random length, 5-min, and 1-min spontaneous speech samples. *Digital Biomarkers*, 7(1):157–166. https://doi.org/10.1159/000533423, PubMed: 38029002

Ulla Petti and Anna Korhonen. 2024. Losst-ad: A longitudinal corpus for tracking Alzheimer's disease related changes in spontaneous speech. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10813–10821.

Alexander Pilozzi and Xudong Huang. 2020. Overcoming Alzheimer's disease stigma by leveraging artificial intelligence and blockchain technologies. *Brain Sciences*, 10(3):183. https://doi.org/10.3390/brainsci10030183, PubMed: 32210011

Anna Pompili, Alberto Abad, David Martins de Matos, and Isabel Pavao Martins. 2018. Topic coherence analysis for the classification of Alzheimer's disease. In *IberSPEECH*, pages 281–285. https://doi.org/10.21437/IberSPEECH.2018-59

Anna Pompili, Alberto Abad, David Martins de Matos, and Isabel Pavão Martins. 2020a. Pragmatic aspects of discourse production for the automatic identification of Alzheimer's disease. *IEEE Journal of Selected Topics in Signal Processing*, 14(2):261–271. https://doi.org/10.1109/JSTSP.2020.2967879

Anna Pompili, Thomas Rolland, and Alberto Abad. 2020b. The inesc-id multi-modal system for the adress 2020 challenge. *arXiv preprint arXiv:2005.14646*. https://doi.org/10.21437/Interspeech.2020-2833

Charlene Pope and Boyd H. Davis. 2011. Finding a balance: The carolinas conversation collection. https://doi.org/10.1515/cllt.2011.007

Chloé Pou-Prom and Frank Rudzicz. 2018. Learning multiview embeddings for assessing dementia. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2812–2817. https://doi.org/10.18653/v1/D18-1304

Martin Prince, Daisy Acosta, Cleusa P. Ferri, Mariella Guerra, Yueqin Huang, Juan J. Llibre Rodriguez, Aquiles Salas, Ana Luisa Sosa, Joseph D. Williams, Michael E. Dewey, Isaac Acosta, Amuthavalli T. Jotheeswaran, and Zhaorui Liu. 2012. Dementia incidence and mortality in middle-income countries, and associations with indicators of cognitive reserve: A 10/66 dementia research group population-based cohort study. *The Lancet*, 380(9836):50–58. https://doi.org/10.1016/S0140-6736(12)60399-7, PubMed: 22626851

Martin Prince, Anders Wimo, Maëlenn Guerchet, Gemma-Claire Ali, Yu-Tzu Wu, and Matthew Prina. 2015. *World Alzheimer Report 2015. The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends*. Ph.D. thesis, Alzheimer's Disease International.

Emily Prud'hommeaux, Margaret Mitchell, and Brian Roark. 2011. Using patterns of narrative recall for improved detection of mild cognitive impairment. *Age*, 81:79–7.

Emily Prud'hommeaux and Brian Roark. 2015. Graph-based word alignment for clinical language evaluation. *Computational Linguistics*, 41(4):549–578. https://doi.org/10.1162/COLI_a_00232, PubMed: 34334943

Emily T. Prud'hommeaux and Brian Roark. 2011. Extraction of narrative recall patterns for neuropsychological assessment. In *Twelfth Annual Conference of the International Speech Communication Association*. https://doi.org/10.21437/Interspeech.2011-756

Xiang Qi. 2023. ChatGPT: A promising tool to combat social isolation and loneliness in older adults with mild cognitive impairment. *Neurology Live*, NA–NA.

Xiaoke Qi, Qing Zhou, Jian Dong, and Wei Bao. 2023. Noninvasive automatic detection of Alzheimer's disease from spontaneous speech: A review. *Frontiers in Aging Neuroscience*, 15:1224723. https://doi.org/10.3389/fnagi.2023.1224723, PubMed: 37693647

Yu Qiao, Xuefeng Yin, Daniel Wiechmann, and Elma Kerz. 2021. Alzheimer's disease detection from spontaneous speech through combining linguistic complexity and (dis) fluency features with pretrained language models. *arXiv preprint arXiv:2106.08689*. https://doi.org/10.21437/Interspeech.2021-1415

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR.

Jill Rasmussen and Haya Langerman. 2019. Alzheimer's disease–why we need early diagnosis. *Degenerative Neurological and Neuromuscular Disease*, 123–130. https://doi.org/10.2147/DNND.S228939, PubMed: 31920420

Kritesh Rauniyar, Shuvam Thakur, Aayush Nevatia, and Prashant Giridhar Shambharkar. 2023. Early detection of Alzheimer's disease: The importance of speech analysis. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pages 1069–1073. IEEE. https://doi.org/10.1109/ICAAIC56838.2023.10140703

Stephanie Reeves, Victoria Williams, Francisco M. Costela, Rocco Palumbo, Olivia Umoren, Mikaila M. Christopher, Deborah Blacker, and Russell L. Woods. 2020. Narrative video scene description task discriminates between levels of cognitive impairment in Alzheimer's disease. *Neuropsychology*, 34(4):437. https://doi.org/10.1037/neu0000621, PubMed: 31999169

Michael J. Rigby. 2019. Ethical dimensions of using artificial intelligence in health care. *AMA Journal of Ethics*, 21(2):121–124. https://doi.org/10.1001/amajethics.2019.121

Brian Roark, Margaret Mitchell, and Kristy Hollingshead. 2007. Syntactic complexity measures for detecting mild cognitive impairment. In *Biological, Translational, and Clinical Language Processing*, pages 1–8. https://doi.org/10.3115/1572392.1572394

Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead, and Jeffrey Kaye. 2011. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2081–2090. https://doi.org/10.1109/TASL.2011.2112351, PubMed: 22199464

Rosebud O. Roberts, Yonas E. Geda, David S. Knopman, Ruth H. Cha, V. Shane Pankratz, Bradley F. Boeve, Robert J. Ivnik, Eric G. Tangalos, Ronald C. Petersen, and Walter A. Rocca. 2008. The mayo clinic study of aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30(1):58–69. https://doi.org/10.1159/000115751, PubMed: 18259084

Jessica Robin, Mengdan Xu, Aparna Balagopalan, Jekaterina Novikova, Laura Kahn, Abdi Oday, Mohsen Hejrati, Somaye Hashemifar, Mohammadreza Negahdar, William Simpson, et al. 2023. Automated detection of progressive speech changes in early Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(2):e12445. https://doi.org/10.1002/dad2.12445, PubMed: 37361261

Morteza Rohanian, Julian Hough, and Matthew Purver. 2021a. Alzheimer's dementia recognition using acoustic, lexical, disfluency and speech pause features robust to noisy inputs.

arXiv preprint arXiv:2106.15684. https://doi.org/10.21437/Interspeech.2021-1633

Morteza Rohanian, Julian Hough, and Matthew Purver. 2021b. Multi-modal fusion with gating using audio, lexical and disfluency features for Alzheimer's dementia recognition from spontaneous speech. *arXiv preprint arXiv:2106.09668*. https://doi.org/10.21437/Interspeech.2020-2721

Damián Solís Rosas, Saúl Tovar Arriaga, and Marco Antonio Aceves Fernández. 2019. Search for dementia patterns in transcribed conversations using natural language processing. In *2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, pages 1–6. IEEE. https://doi.org/10.1109/ICEEE.2019.8884572

Benjamin S. Runde, Ajit Alapati, and Nicolas G. Bazan. 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. https://doi.org/10.3390/brainsci14030211, PubMed: 38539600

Miriam Ryvicker, Yolanda Barrón, Jiyoun Song, Maryam Zolnoori, Shivani Shah, Julia G. Burgdorf, James M. Noble, and Maxim Topaz. 2024. Using natural language processing to identify home health care patients at risk for diagnosis of Alzheimer's disease and related dementias. *Journal of Applied Gerontology*, page 07334648241242321. https://doi.org/10.1177/07334648241242321, PubMed: 38556756

Tausifa Jan Saleem, Syed Rameem Zahra, Fan Wu, Ahmed Alwakeel, Mohammed Alwakeel, Fathe Jeribi, and Mohammad Hijji. 2022. Deep learning-based diagnosis of Alzheimer's disease. *Journal of Personalized Medicine*, 12(5):815. https://doi.org/10.3390/jpm12050815, PubMed: 35629237

Ploypaphat Saltz, Shih Yin Lin, Sunny Chieh Cheng, and Dong Si. 2021. Dementia detection using transformer-based deep learning and natural language processing models. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, pages 509–510. IEEE. https://doi.org/10.1109/ICHI52183.2021.00094

Leandro Santos, Edilson Anselmo Corrêa Júnior, Osvaldo Oliveira Jr., Diego Amancio, Letícia Mansur, and Sandra Aluísio. 2017. Enriching complex networks with word embeddings for detecting mild cognitive impairment from speech transcripts. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vol. 1*. https://doi.org/10.18653/v1/P17-1118

Ita Daryanti Saragih, Chun-Wang Wei, Sakti Oktaria Batubara, Ice Septriani Saragih, and Bih-O Lee. 2023. Effects of technology-assisted interventions for people with dementia: A systematic review and meta-analysis. *Journal of Nursing Scholarship*, 55(1):291–303. https://doi.org/10.1111/jnu.12808, PubMed: 36056586

Utkarsh Sarawgi, Wazeer Zulfikar, Nouran Soliman, and Pattie Maes. 2020. Multimodal inductive transfer learning for detection of Alzheimer's dementia and its severity. *arXiv preprint arXiv:2009.00700*. https://doi.org/10.21437/Interspeech.2020-3137

Thomas Searle, Zina Ibrahim, and Richard Dobson. 2020. Comparing natural language processing techniques for Alzheimer's dementia prediction in spontaneous speech. *arXiv preprint arXiv:2006.07358*. https://doi.org/10.21437/Interspeech.2020-2729

Adam Ševčík and Milan Rusko. 2022. A systematic review of Alzheimer's disease detection based on speech and natural language processing. In *2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA)*, pages 1–5. IEEE. https://doi.org/10.1109/RADIOELEKTRONIKA54537.2022.9764938

Nagamani H. Shahapure, N. Poornima, Spoorthi Kulkarni, V. Shilpa, Rajiv Ranjan Singh, and Patel Kavan. 2022. NLP based word predictor for dementia patients: A systematic review. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1538–1543. IEEE. https://doi.org/10.1109/ICICCS53718.2022.9788447

Arezo Shakeri, Shaima Ahmad Freja, Yeganeh Hallaj, and Mina Farmanbar. 2024. Uncovering linguistic patterns: A machine learning exploration for early dementia detection in speech transcripts. In *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)*, pages 1–8. IEEE. https://doi.org/10.1109/ICAPAI61893.2024.10541296

Ravi Kumar Sharma, Shyam Sunder Jannu Soloman, and Nagaraju Baydeti. 2024. Detection of Alzheimer's disease using machine learning classification. In *2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, pages 21–26. IEEE. https://doi.org/10.1109/ECTIDAMTNCON60518.2024.10479989

Zitao Shen, Dalton Schutte, Yoonkwon Yi, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 2022. Classifying the lifestyle status for Alzheimer's disease from clinical notes using deep learning with weak supervision. *BMC Medical Informatics and Decision Making*, 22(Suppl 1):88. https://doi.org/10.1186/s12911-022-01819-4, PubMed: 35799294

Mengke Shi, Gary Cheung, and Seyed Reza Shahamiri. 2023. Speech and language processing with deep learning for dementia diagnosis: A systematic review. *Psychiatry Research*, page 115538. https://doi.org/10.1016/j.psychres.2023.115538, PubMed: 37864994

John R. Sims, Jennifer A. Zimmer, Cynthia D. Evans, Ming Lu, Paul Ardayfio, JonDavid Sparks, Alette M. Wessels, Sergey Shcherbinin, Hong Wang, Emel Serap Monkul Nery, Emily C. Collins, Paul Solomon, Stephen Salloway, Liana G. Apostolova, Oskar Hansson, Craig Ritchie, Dawn A. Brooks, Mark Mintun, Daniel M. Skovronsky, and TRAILBLAZER-ALZ 2 Investigators. 2023. Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*, 330(6): 512–527. https://doi.org/10.1001/jama.2023.13239, PubMed: 37459141

Kairit Sirts, Olivier Piguet, and Mark Johnson. 2017. Idea density for predicting Alzheimer's disease from transcribed speech. *arXiv preprint arXiv:1706.04473*. https://doi.org/10.18653/v1/K17-1033

Caroline Skirrow, Marton Meszaros, Udeepa Meepegama, Raphael Lenain, Kathryn V. Papp, Jack Weston, and Emil Fristed. 2022. Validation of a remote and fully automated story recall task to assess for early cognitive impairment in older adults: Longitudinal case-control observational study. *JMIR Aging*, 5(3):e37090. https://doi.org/10.2196/37090, PubMed: 36178715

Andrew L. Smith, Felix Greaves, and Trishan Panch. 2023. Hallucination or confabulation? Neuroanatomy as metaphor in large language models. *PLOS Digital Health*, 2(11):e0000388. https://doi.org/10.1371/journal.pdig.0000388, PubMed: 37910473

Aradhana Soni, Benjamin Amrhein, Matthew Baucum, Eun Jin Paek, and Anahita Khojandi. 2021. Using verb fluency, natural language processing, and machine learning to detect Alzheimer's disease. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2282–2285. IEEE. https://doi.org/10.1109/EMBC46164.2021.9630371, PubMed: 34891742

Gregor Stiglic, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, and Leona Cilar. 2020. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5):e1379. https://doi.org/10.1002/widm.1379

Edoardo Stoppa, Guido Walter Di Donato, Isabella Poles, Eleonora D'Arnese, Natalie Parde, and Marco Domenico Santambrogio. 2023. A graph machine learning approach to automatic dementia detection. In *2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–4. IEEE. https://doi.org/10.1109/BHI58575.2023.10313453

Yoon Sunmoo, Peter Broadwell, Carmela Alcantara, Nicole Davis, Lee Haeyoung, Amanda Bristol, Dante Tipiani, Joo Young Nho, and Mary Mittelman. 2022. Analyzing topics and sentiments from twitter to gain insights to refine interventions for family caregivers of persons with Alzheimer's disease and related dementias (adrd) during covid-19 pandemic. *Studies in Health Technology and Informatics*, 289:170. https://doi.org/10.3233/SHTI210886

Yoon Sunmoo, Peter Broadwell, Dante Tipiani, Amanda Bristol, Moon Soyoung, Yoon Brian, Liu Jianfang, Niya Huang, and Nicole Davis. 2023. Comparing emotional valence scores of twitter messages from human coding and machine learning algorithms among hispanic and african american family caregivers of persons with dementia. *Studies in Health Technology and Informatics*, 305:440. https://doi.org/10.3233/SHTI230526

Muhammad Shehram Shah Syed, Zafi Sherhan Syed, Margaret Lech, and Elena Pirogova. 2020. Automated screening for Alzheimer's dementia through spontaneous speech. In *Interspeech*, volume 2020, pages 2222–2226.

Zafi Sherhan Syed, Muhammad Shehram Shah Syed, Margaret Lech, and Elena Pirogova. 2021. Tackling the adresso challenge 2021: The muet-rmit system for Alzheimer's dementia recognition from spontaneous speech. In *Interspeech*, pages 3815–3819.

Behrad Taghibeyglou and Frank Rudzicz. 2023. Who needs context? Classical techniques for Alzheimer's disease detection. In *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 102–107. https://doi.org/10.18653/v1/2023.clinicalnlp-1.13

Amir Abbas Tahami Monfared, Yaakov Stern, Stephen Doogan, Michael Irizarry, and Quanwu Zhang. 2022. Stakeholder insights in Alzheimer's disease: Natural language processing of social media conversations. *Journal of Alzheimer's Disease*, 89(2):695–708. https://doi.org/10.3233/JAD-220422, PubMed: 35938254

Calvin Thomas, Vlado Keselj, Nick Cercone, Kenneth Rockwood, and Elissa Asp. 2005. Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. In *IEEE International conference mechatronics and automation, 2005*, volume 3, pages 1569–1574. IEEE. https://doi.org/10.1109/ICMA.2005.1626789

Judith Tillmann, Johannes Just, Rieke Schnakenberg, Klaus Weckbecker, Birgitta Weltermann, and Eva Münster. 2019. Challenges in diagnosing dementia in patients with a migrant background-a cross-sectional study among german general practitioners. *BMC Family Practice*, 20:1–10. https://doi.org/10.1186/s12875-019-0920-0, PubMed: 30803438

Maxim Topaz, Victoria Adams, Paula Wilson, Kyungmi Woo, and Miriam Ryvicker. 2020. Free-text documentation of dementia symptoms in home healthcare: A natural language processing study. *Gerontology and Geriatric Medicine*, 6:2333721420959861. https://doi.org/10.1177/2333721420959861, PubMed: 33029550

Matthias S. Treder, Sojin Lee, and Kamen A. Tsvetanov. 2024. Introduction to large language models (LLMs) for dementia care and research. *Frontiers in Dementia*, 3:1385303. https://doi.org/10.3389/frdem.2024.1385303, PubMed: 39081594

Kosha Upadhyay, Jason Zhang, and Eric Lloyd. 2022. Cognitive profiling and personalized therapy recommendation for dementia through a language aware multi-model artificially intelligent system. In *2022 IEEE MIT Undergraduate Research Technology Conference (URTC)*, pages 1–5. IEEE. https://doi.org/10.1109/URTC56832.2022.10002250

Akshay Valsaraj, Ithihas Madala, Nikhil Garg, and Veeky Baths. 2021. Alzheimer's dementia detection using acoustic & linguistic features and pre-trained BERT. In *2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pages 171–175. IEEE. https://doi.org/10.1109/ISCMI53840.2021.9654804

Viswan Vimbi, Noushath Shaffi, and Mufti Mahmud. 2024. Interpreting artificial intelligence models: A systematic review on the application of lime and SHAP in Alzheimer's disease detection. *Brain Informatics*, 11(1):10. https://doi.org/10.1186/s40708-024-00222-1, PubMed: 38578524

Vimbi Viswan, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, and Faizal

Hajamohideen. 2024. Explainable artificial intelligence in Alzheimer's disease classification: A systematic review. *Cognitive Computation*, 16(1):1–44. https://doi.org/10.1007/s12559-023-10192-x

Rohit Voleti, Julie M. Liss, and Visar Berisha. 2019. A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE journal of selected topics in signal processing*, 14(2):282–298. https://doi.org/10.1109/JSTSP.2019.2952087, PubMed: 33907590

M. K. Vrindha, V. Geethu, P. R. Anurenjan, S. Deepak, and K. G. Sreeni. 2023. A review of Alzheimer's disease detection from spontaneous speech and text. In *2023 International Conference on Control, Communication and Computing (ICCC)*, pages 1–5. IEEE. https://doi.org/10.1109/ICCC57789.2023.10165507

Changyu Wang, Siru Liu, Aiqing Li, and Jialin Liu. 2023a. Text dialogue analysis for primary screening of mild cognitive impairment: Development and validation study. *Journal of Medical Internet Research*, 25, e51501. https://doi.org/10.2196/51501, PubMed: 38157230

Ning Wang, Yupeng Cao, Shuai Hao, Zongru Shao, and K. P. Subbalakshmi. 2021. Modular multi-modal attention network for Alzheimer's disease detection using patient audio and language data. In *Interspeech*, pages 3835–3839. https://doi.org/10.21437/Interspeech.2021-2024

Ning Wang, Fan Luo, Vishal Peddagangireddy, Koduvayur P. Subbalakshmi, and Rajarathnam Chandramouli. 2020a. Personalized early stage Alzheimer's disease detection: A case study of president reagan's speeches. *arXiv preprint arXiv:2005.12385*. https://doi.org/10.1101/2020.05.01.20087627

Yi Wang, Jiajun Deng, Tianzi Wang, Bo Zheng, Shoukang Hu, Xunying Liu, and Helen Meng. 2023b. Exploiting prompt learning with pre-trained language models for Alzheimer's disease detection. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE. https://doi.org/10.1109/ICASSP49357.2023.10095993

Yi Wang, Tianzi Wang, Zi Ye, Lingwei Meng, Shoukang Hu, Xixin Wu, Xunying Liu, and Helen Meng. 2022. Exploring linguistic feature and model combination for speech recognition based automatic ad detection. *Interspeech*. https://doi.org/10.21437/Interspeech.2022-723

Zixu Wang, Julia Ive, Sinéad Moylett, Christoph Mueller, Rudolf N. Cardinal, Sumithra Velupillai, John O'Brien, and Robert Stewart. 2020b. Distinguishing between dementia with lewy bodies (dlb) and Alzheimer's disease (ad) using mental health records: A classification approach. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.clinicalnlp-1.19

Sebastian Wankerl, Elmar Nöth, and Stefan Evert. 2017. An n-gram based approach to the automatic diagnosis of Alzheimer's disease from spoken language. https://doi.org/10.21437/Interspeech.2017-1572

Davy Weissenbacher, Travis A Johnson, Laura Wojtulewicz, Amylou Dueck, Dona Locke, Richard Caselli, and Graciela Gonzalez. 2016. Automatic prediction of linguistic decline in writings of subjects with degenerative dementia. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1198–1207. https://doi.org/10.18653/v1/N16-1143

Bingyang Wen, Ning Wang, Koduvayur Subbalakshmi, Rajarathnam Chandramouli, et al. 2023. Revealing the roles of part-of-speech taggers in Alzheimer disease detection: Scientific discovery using one-intervention causal explanation. *JMIR Formative Research*, 7(1):e36590. https://doi.org/10.2196/36590, PubMed: 37129944

Simon Williams, Clea Tanner, and Claire Lancaster. 2023. Inhibitory control and the production of disfluencies in speakers with Alzheimer's disease. https://doi.org/10.21437/DiSS.2023-4

Robert T. Woods, E. Bruce, R. T. Edwards, R. Elvish, Z. Hoare, B. Hounsome, J. Keady, E. D. Moniz-Cook, V. Orgeta, M. Orrell,

J. Rees, and I. T. Russell. 2012. REMCARE: Reminiscence groups for people with dementia and their family caregivers-effectiveness and costeffectiveness pragmatic multicentre randomised trial. *Health Technology Assessment*, 16(48). https://doi.org/10.3310/hta16480, PubMed: 23211271

World Population Review. 2025. Dementia rates by country 2025.

Wenbo Wu, Kaes J. Holkeboer, Temidun O. Kolawole, Lorrie Carbone, and Elham Mahmoudi. 2023. Natural language processing to identify social determinants of health in Alzheimer's disease and related dementia from electronic health records. *Health Services Research*, 58(6):1292–1302. https://doi.org/10.1111/1475-6773.14210, PubMed: 37534741

Anna Xygkou, Chee Siang Ang, Panote Siriaraya, Jonasz Piotr Kopecki, Alexandra Covaci, Eiman Kanjo, and Wan-Jou She. 2024. Mindtalker: Navigating the complexities of AI-enhanced social engagement for people with early-stage dementia. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–15. https://doi.org/10.1145/3613904.3642538

Maria Yancheva, Kathleen C. Fraser, and Frank Rudzicz. 2015. Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias. In *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, pages 134–139. https://doi.org/10.18653/v1/W15-5123

Maria Yancheva and Frank Rudzicz. 2016. Vector-space topic models for detecting Alzheimer's disease. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2337–2346. https://doi.org/10.18653/v1/P16-1221

Qin Yang, Xin Li, Xinyun Ding, Feiyang Xu, and Zhenhua Ling. 2022. Deep learning-based speech analysis for Alzheimer's disease detection: A literature review. *Alzheimer's Research & Therapy*, 14(1):186. https://doi.org/10.1186/s13195-022-01131-3, PubMed: 36517837

Yoonkwon Yi, Zitao Shen, Anusha Bompelli, Fang Yu, Yanshan Wang, and Rui Zhang. 2020. Natural language processing methods to extract lifestyle exposures for Alzheimer's disease from clinical notes. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–2. IEEE. https://doi.org/10.1109/ICHI48887.2020.9374320

Jiahong Yuan, Yuchen Bian, Xingyu Cai, Jiaji Huang, Zheng Ye, and Kenneth Church. 2020. Disfluencies and fine-tuning pre-trained language models for detection of Alzheimer's disease. In *Interspeech*, volume 2020, pages 2162–2166. https://doi.org/10.21437/Interspeech.2020-2516

Kimia Tuz Zaman, Wordh Ul Hasan, Juan Li, and Cui Tao. 2023. Empowering caregivers of Alzheimer's disease and related dementias (adrd) with a GPT-powered voice assistant: Leveraging peer insights from social media. In *2023 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–7. IEEE. https://doi.org/10.1109/ISCC58397.2023.10218142

Chuheng Zheng, Mondher Bouazizi, and Tomoaki Ohtsuki. 2022. An evaluation on information composition in dementia detection based on speech. *IEEE Access*, 10:92294–92306. https://doi.org/10.1109/ACCESS.2022.3203068

Luke Zhou, Kathleen C. Fraser, and Frank Rudzicz. 2016. Speech recognition in Alzheimer's disease and in its assessment. In *Interspeech*, volume 2016, pages 1948–1952. https://doi.org/10.21437/Interspeech.2016-1228

Sicheng Zhou, Dalton Schutte, Aiwen Xing, Jiyang Chen, Julian Wolfson, Zhe He, Fang Yu, and Rui Zhang. 2021. Identification of dietary supplement use from electronic health records using transformer-based language models. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, pages 513–514. IEEE. https://doi.org/10.1109/ICHI52183.2021.00096

Xin Zhou, Hongfang Liu, and Yanshan Wang. 2018. A comparison of lifestyle interventions for Alzheimer's disease extracted from clinical notes and literature. In *2018 IEEE International Conference on Healthcare Informatics Workshop (ICHI-W)*, pages 74–75. IEEE.

https://doi.org/10.1109/ICHI-W.2018.00026

Xin Zhou, Yanshan Wang, Sunghwan Sohn, Terry M. Therneau, Hongfang Liu, and David S. Knopman. 2019. Automatic extraction and assessment of lifestyle exposures for Alzheimer's disease using natural language processing. *International Journal of Medical Informatics*, 130:103943. https://doi.org/10.1016/j.ijmedinf.2019.08.003, PubMed: 31476655

Youxiang Zhu, Xiaohui Liang, John A. Batsis, and Robert M. Roth. 2021. Exploring deep transfer learning techniques for Alzheimer's dementia detection. *Frontiers in Computer Science*, 3:624683. https://doi.org/10.3389/fcomp.2021.624683, PubMed: 34046588

Youxiang Zhu, Xiaohui Liang, John A. Batsis, and Robert M. Roth. 2022a. Domain-aware intermediate pretraining for dementia detection with limited data. In *Interspeech*, volume 2022, page 2183. NIH Public Access. https://doi.org/10.21437/Interspeech.2022-10862

Yunshu Zhu, Ting Song, Zhenyu Zhang, Chao Deng, Mohammad Alkhalaf, Wanqing Li, Mengyang Yin, Hui Chen Chang, and Ping Yu. 2022b. Agitation prevalence in people with dementia in australian residential aged care facilities: Findings from machine learning of electronic health records. *Journal of Gerontological Nursing*, 48(4):57–64. https://doi.org/10.3928/00989134-20220309-01, PubMed: 35343838

Maryam Zolnoori, Yolanda Barrón, Jiyoun Song, James Noble, Julia Burgdorf, Miriam Ryvicker, and Maxim Topaz. 2023. HomeADScreen: Developing Alzheimer's disease and related dementia risk identification model in home healthcare. *International Journal of Medical Informatics*, 177:105146. https://doi.org/10.1016/j.ijmedinf.2023.105146, PubMed: 37454558

A Cohort Construction

A.1 Search Methodology

We searched in six data sources: ACL Anthology, PubMed, DBLP, IEEE Xplore, Springer, and Wiley. The database search was conducted between April and July 2024, with an additional search in January 2025 as part of our revised submission. Our research strategy focused on papers addressing dementia, Alzheimer's, and Mild Cognitive Impairment (MCI) using NLP, including but not limited to large language models (LLMs). We searched for papers including the keywords (''NLP'' OR ''Natural Language Processing'' OR ''LLM'' OR ''Language Model'' OR ''ChatGPT'') AND (''Dementia'' OR ''Alzheimer'' OR ''MCI'' OR ''Cognitive Impairment'') in their title, abstract, and/or keywords. While we do not include keywords for specific dementia pathologies (e.g., Lewy body), we believe that using the general term dementia in the title, abstract and/or keywords will effectively capture papers on various types of the condition.

A.2 Paper Inclusion and Exclusion Criteria

We screened the extracted articles against the following criteria:

1. Only fully accessible, English-language papers were included.
2. Only peer-reviewed papers were included.
3. Only full academic papers (excluding posters and theses) were included.
4. Studies focusing on text as the primary modality or employing multi-modal approaches that incorporate textual analysis were included.
5. Papers focusing solely on audio, visual, or other non-textual data were excluded.
6. Papers that, despite mentioning our defined keywords, do not directly deal with NLP applications to Dementia, were excluded.

Papers were deemed irrelevant to our literature review if they contained dementia- and NLP-related terms in their title, abstract, or keywords but focused on entirely unrelated topics. For example, Botros et al. (2020) mentions 'dementia' in the abstract and 'language model' in the keywords, but primarily focuses on sensor locations. Similarly, Daudet et al. (2016) includes 'Alzheimer's' and 'Language Model' as keywords but primarily discusses the design of an application for diagnosing brain injuries using youth speech data.

As mentioned in Section 2, the first three steps of our screening process were automated, while the remaining steps, including gauging general relevance, were conducted manually by two PhD students from our NLP lab, one of whom has a medical background in Alzheimer-related projects. We encountered discrepancies in fewer than 10% of reviewed works, and resolved them through discussion until a final consensus was reached.

A PRISMA flowchart describing our search process can be seen in Figure 6.

A.3 Dataset Inclusion and Exclusion Criteria

Throughout our review, we discuss classic and contemporary datasets, summarized in Table C. The datasets mentioned in the paper and listed in this table were screened against the following criteria:

1. Only datasets referenced in the reviewed papers were included.
2. The dataset must contain data that is fully or partially in English.
3. Sufficient information about the dataset must be available online or obtainable through private requests to researchers or institutions.
4. Datasets do not need to be publicly accessible to be included in our review.
5. As these datasets are part of the peer-reviewed literature we selected, we assume they have also undergone peer review.

A.4 Paper Annotation

For each paper, we extracted the title, authors, year of publication, and venue. We then manually annotated the following:

1. Task Family: Dementia detection, linguistic bio-marker extraction, caregiver support, patient assistance, literature review, or dataset introduction. While some papers may address multiple categories, we selected the most prominent motivation and novelty as described by the authors.
2. Venue type: NLP (e.g., ACL), Medical (e.g., NIH), Speech (e.g., Interspeech), or Other Technological (e.g., Frontiers in CS).
3. Datasets used (if applicable).
4. Technologies applied (if applicable), such as SVM or BERT.
5. Statistical significance reported: yes or no, depending on whether any statistical tests were conducted on data or algorithmic results.

B Extended Mention of Studies

In Section 3, we cited a representative sample of dementia detection studies we reviewed. Below, we cite the remainder of the papers according to internal categorization used for our internal analysis.

Detection papers using **classic ML classifiers** on a wide range of features, from straightforward N-grams and part-of-speech tagging to information content units extracted from picture descriptions and even dementia patients' dental records and motor signs: Prud'hommeaux and Roark (2011); Fraser et al. (2013); Jarrold et al. (2014); Prud'hommeaux and Roark (2015); Fraser et al. (2016a); Yancheva and Rudzicz (2016); Zhou et al. (2016); Weissenbacher et al. (2016); Mirheidari et al. (2016); Santos et al. (2017); Wankerl et al. (2017); Klumpp et al. (2018); Pou-Prom and Rudzicz (2018); Eyre et al. (2020); Chen et al. (2020); Searle et al. (2020); Martinc and Pollak (2020); Hane et al. (2020); Al-Harrasi et al. (2021); Nasreen et al. (2021b); Clarke et al. (2021); Gónzalez Atienza et al. (2021); Soni et al. (2021); Penfold et al. (2022); Ablimit et al. (2022b); Dey and Mittal (2022); Amini et al. (2023); Liu and Wang (2023); Taghibeyglou and Rudzicz 2023; Wen et al. (2023); Patel and Wu (2024); Sharma et al. (2024).

Studies using classic neural models such as LSTMs and transformer-based models like BERT: Karlekar et al. (2018), Pompili et al. (2018), Pou-Prom and Rudzicz (2018), Hong et al. (2019), Di Palo and Parde (2019), Chen et al. (2019), Fritsch et al. (2019), Pan et al. (2019), Edwards et al. (2020), Syed et al. (2020), Koo et al. (2020), Sarawgi et al. (2020), Pompili et al. (2020b), Cummins et al. (2020), Balagopalan et al. (2020), Yuan et al. (2020), Pappagari et al. (2020), Haulcy and Glass (2021), Valsaraj et al. (2021), Balagopalan et al. (2021), Wang et al. (2021), Qiao et al. (2021), Campbell et al. (2021), Nambiar et al. (2022), Khan et al. (2022b), Rohanian et al. (2021a), Syed et al. (2021), Pappagari et al. (2021), Pérez-Toro et al. (2021), Mirheidari et al. (2021),

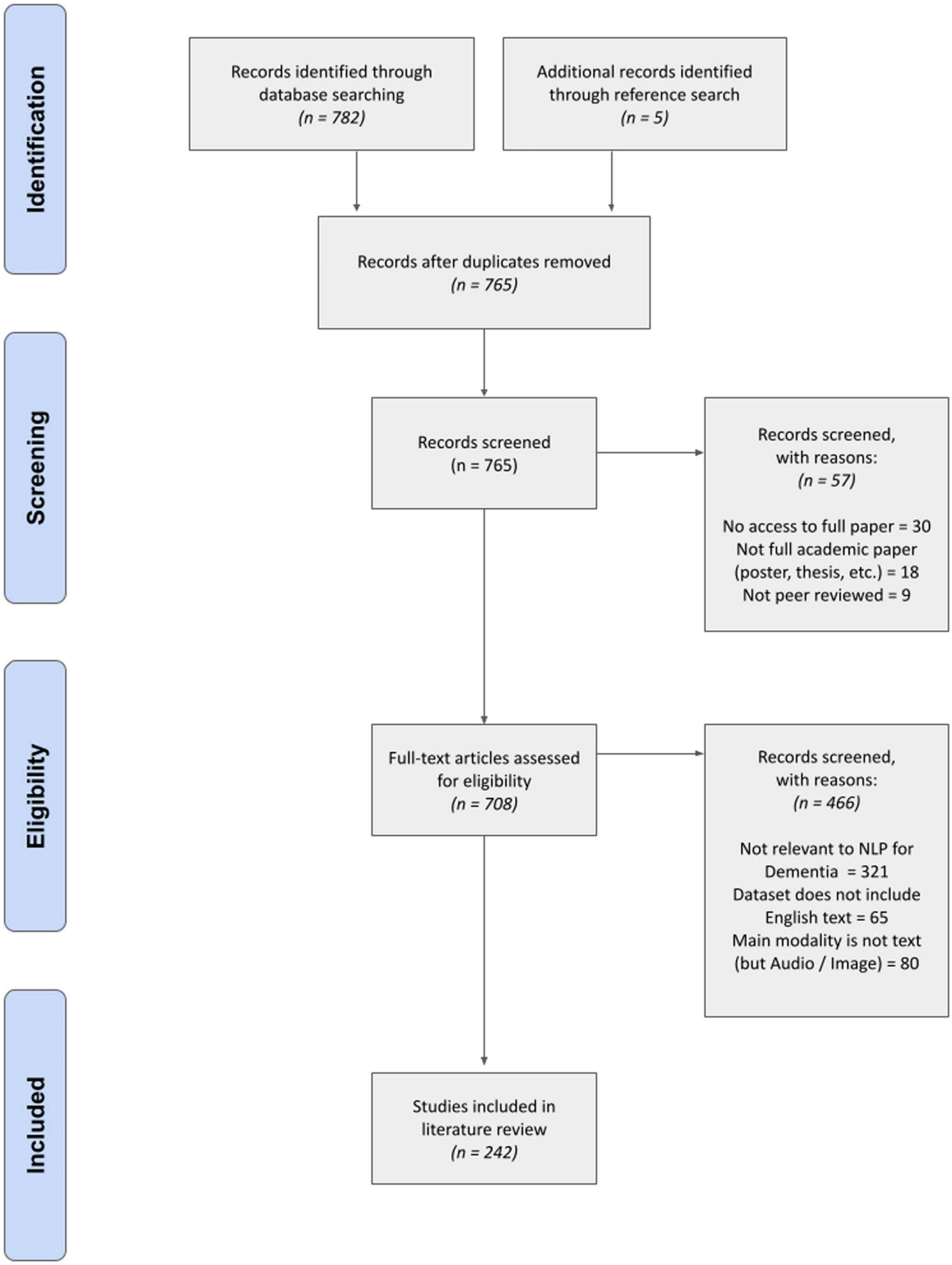

Figure 6: PRISMA flowchart displaying study screening and selection process.

Liu et al. (2021), Zhu et al. (2021), Saltz et al. (2021), Khan et al. (2022a), Deng et al. (2022), Bouazizi et al. (2022), Horigome et al. (2022), Zheng et al. (2022), Liu et al. (2022b), Liu et al. (2022a), Wang et al. (2022), Matošević and Jović (2022), Amini et al. (2023), Rauniyar et al. (2023), Abdelhalim et al. (2023), Bouazizi et al. (2023), Gkoumas et al. (2023), Cai et al. (2023), Wang et al. 2023b), Shakeri et al. (2024), Dong et al. (2024), Kim et al. (2024), Laurentiev et al. (2024).

Studies classifying between more granular stages of the dementia (e.g., 'MCI', 'Alzheimer's'), assess disease progression or differentiate between dementia types (e.g., Alzheimer's versus Frontotemporal), through classification: Thomas et al. (2005); Orimaye et al. (2016); Wang et al. (2020b); Padhee et al. (2020); Merone et al. (2022); Mao et al. (2023); Liu et al. (2023b); Amini et al. (2024); Laurentiev et al. (2024); Panahi et al. (2024). Studies focusing on text-based MMSE score regression for dementia detection: Yancheva et al. (2015), Pou-Prom and Rudzicz (2018), Farzana and Parde (2020), Rohanian et al. (2021b), (Stoppa et al., 2023),Aryal et al. (2023).

Studies presenting multi-modal approaches (with audio / video) as well as augmentation methods using synthetic, noisy, cross-domain, or semantically and structurally different datasets: Che et al. (2017); Gligic et al. (2020); Gupta et al. (2020); Amin-Nejad et al. (2020); Mahajan and Baths; Guo et al. (2021); Zhu et al. (2021, 2022a); Ablimit et al. (2022a); Mirheidari et al (2022); Hlédiková et al. (2022); Ilias and Askounis (2022b); El-Sappagh et al. (2022); Li et al. (2022b); Maclagan et al. (2023); Chen et al. (2023b); Farzana and Parde (2023); Cui et al. (2023); Chen et al. (2023a); Cai et al. (2023); Lin and Washington (2024); Fard et al. (2024).

Studies leveraging LLMs for detection or data augmentation: Bullard et al. (2016), Liu et al. (2019), Chintagunta et al. (2021), Liu et al. (2023a), Cai et al. (2023), Duan et al. (2023), Koga et al. (2024), Latif and Kim (2024), Casu et al. (2024).

C Full Dataset Table

Table 2 showcases the dementia-related datasets encountered throughout our review. Datasets such as the Reagan Library and IMDB, though creatively used in some papers, are excluded as they are not dementia-specific. Private health records or unreleased clinical notes are excluded as well.

General	Data Source					Modalities						Participants	
Dataset	**Longi-tudinal**	**Clinical setting**	**Social media**	**Non-clinic. interviews**	**Synt. data**	**Trans. speech**	**Written text**	**Audio**	**Video**	**Medical imaging**	**Physical markers**	**Patients**	**Other**
DementiaBank (Becker et al., 1994)	x	x	–	–	–	x	–	x	–	–	–	196 Dem 98 Ctrl	–
ADReSS (Luz et al., 2021b)	–	x	–	–	–	x	–	x	–	–	–	78 Dem 78 Ctrl	–
ADReSSo (Luz et al., 2021a)	–	x	–	–	–	–	–	x	–	–	–	121 Dem 116 Ctrl	–
CCC (Pope and Davis, 2011)	x	–	–	x	–	x	–	x	x	–	–	124 Dem 125 Ctrl	–
B-SHARP (Li et al., 2020)	–	x	–	–	–	x	–	x	–	–	–	144 Dem 185 Ctrl	–
IVA (Pan et al., 2020)	–	x	–	–	–	–	–	x	–	–	–	45 Dem 25 Ctrl	–
Farmington Heart Study	x	x	–	–	–	x	x	x	–	x	x	>15,000 over decades	–
CareD (Garg and Sohn, 2023)	–	–	x	–	–	–	x	–	–	–	–	–	1005 caregiver posts
ADNI (Petersen et al., 2010)	x	x	–	–	–	–	x	–	–	x	x	>2500 over several cohorts	–
Wisconsin WLS (Herd et al., 2014)	x	x	–	x	–	x	x	–	–	–	x	>10,000 over 60 years	–
MCSA (Roberts et al., 2008)	x	x	–	–	–	–	x	–	–	x	x	>3000	–
Layton Aging and Alz. Disease Research Center	–	x	–	x	–	x	–	x	x	x	x	Depending Sub-cohort	
I-CONECT	x	–	–	x	–	x	–	x	x	–	–	320*	–
LoSST-AD (Petti and Korhonen, 2024)	x	–	–	x	–	x	–	–	–	–	–	10 Dem 10 Ctrl	–
SLaCAD (Farzana et al., 2024)	x	x	–	–	–	x	–	x	–	–	x	9 Dem 82 Ctrl	–
Li et al. (2023)	–	x	–	–	x	–	x	–	–	–	–	–	(1) Clinicians (2) Synt. Annotated (3) Synt. Generated 16,000 sentences each.
Gkoumas et al. (2024)	x	–	–	x	–	x	x	x	–	–	–	12 Dem 10 Ctrl	–

Table 2: Overview of datasets reviewed. Most datasets originate from transcribed or written speech in clinical settings, with quite a few offering longitudinal data. Two under-represented categories here are synthetic datasets and social media-based datasets, with only one of each category. For additional sources of social media data we refer the readers to Sunmoo et al. (2023) and Masrani et al. (2017). While they do not publish full datasets, they provide code to replicate their scraping process. For non-English datasets, see Yang et al. (2022).
** Estimated number of participants in the I-CONECT clinical trial are based on publicly available data.