




Contents lists available at ScienceDirect

## The Journal of Prevention of Alzheimer's Disease

journal homepage: [www.elsevier.com/locate/tjpad](http://www.elsevier.com/locate/tjpad)

Special Article

## Multi-modal data analysis for early detection of alzheimer's disease and related dementias

Liming Wang<sup>a</sup>, Jim Glass<sup>a</sup>, Lampros Kourtis<sup>b</sup>, Rhoda Au<sup>c,\*</sup> <sup>a</sup> Massachusetts Institute of Technology, Cambridge, MA, USA<sup>b</sup> Gates Ventures, Seattle, WA, USA<sup>c</sup> Boston University Chobanian & Avedisian School of Medicine and School of Public Health, Boston, MA, USA

## ARTICLE INFO

## Keywords:

Dementia classification  
 Speech and language processing  
 State-space model  
 Multi-modal

## ABSTRACT

Until recently, accurate early detection of clinical symptoms associated with Alzheimer's disease (AD) and related dementias (ADRD) has been difficult. Digital technologies have created new opportunities to capture cognitive and other AD/ADRD related behaviors with greater sensitivity and specificity. Speech captured through digital recordings has shown recent promise at feasible levels of scalability because of the widespread penetration of smartphones. One such study is described in detail to illustrate the depth in which artificial intelligence (AI) analytic approaches can be used to amplify the value of audio recordings. Another modality that has also attracted research interest are ocular scans that have near term potential for validation as a digital biomarker and a point of entry for clinical care workflows. Single modality measures, however, are rapidly giving way to multi-modality sensors that are embedded in all smartphones and other internet-of-things connected devices. Artificial intelligence (AI) driven analytic approaches are able to divine clinical signals from these high dimensional digital data streams. These data driven findings are setting the stage for a future state in which AD/ADRD detection will be possible at the earliest possible stage of the neurodegenerative process and enable interventions that would significantly attenuate or alter the trajectory, preventing disease from reaching the clinical diagnosis threshold.

## 1. Introduction

With United States' (US) Federal Drug and Administration (FDA) approval of lecanemab and donenamab to slow progression of Alzheimer's disease (AD) clinical symptoms, early diagnosis of AD has become increasingly important for initiating timely treatment, slowing disease progression, and improving patients' quality of life and life expectancy [1,2]. But determining who to treat and when remains a significant challenge. Widespread AD in vivo pathological detection is now possible through the FDA's more recent approvals of AD blood-based biomarkers [3]. But presence of AD pathology does not always lead to clinical expression of disease [4]. Further, clinical symptoms, particularly in the earliest stages are highly variable, resulting in a significant challenge of identifying *clinically meaningful* symptoms. Compounding the clinical detection objective is that AD is often co-morbid with other pathologies, such as vascular and/or Lewy body pathologies, which also result in cognitive and related behavioral indicators, some of which overlap with those of AD. The ethical dilemma is potential treatment of those who are AD biomarker positive but will never progress to clinically

expressed disease given pharmacologically concomitant side effects. Thus, despite the exciting promise of AD blood biomarkers, the clinical utility of these blood tests will be constrained without confirmation that treatment is warranted.

Cognitive impairment is the most common clinical symptom of AD and related dementias (ADRD) and is a key primary outcome in AD clinical trials to determine intervention efficacy. Neuropsychological tests are widely used to assess cognitive state. They include multi-domain screening tests such as the widely used Mini-Mental State Examination (MMSE) [5] and domain specific assessments that typically rely on trained examiners administering standardized protocols. A comprehensive cognitive protocol is often labor-intensive to administer, comprised of tests that are culturally and educationally biased, and produce results that are variable making distinction from subtle or early-stage cognitive decline difficult to discern [6]. Manual scoring approaches also may miss nuanced indicators that signal cognitive impairment. Technological advances have emerged that can alleviate these issues. Digital capture of behaviors through smartphones, wearables and other internet of things (IoT) present an opportunity to

\* Corresponding author.

E-mail address: [rhodaau@bu.edu](mailto:rhodaau@bu.edu) (R. Au).<https://doi.org/10.1016/j.tjpad.2025.100399>

Received 4 September 2025; Received in revised form 15 October 2025; Accepted 20 October 2025

Available online 1 December 2025

2274-5807/© 2025 The Author(s). Published by Elsevier Masson SAS on behalf of SERDI Publisher. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

overcome the limitations of standardized neuropsychological testing and do so at a scale that has not been hereto for possible.

In this perspective, we highlight two areas in which digital technologies are being used in AD/ADRD because of their non-invasive and feasibly scalable nature. The first section provides a detailed description of research being done with speech as an illustration of how both technological advances are being applied to both data collection and analysis. The second section summarizes ocular scans research to introduce the potential realm of digital biomarkers. The final section looks beyond these highlighted efforts to present a vision of the future, well beyond the single modality approach and where much scientific discovery remains to be done.

## 2. Speech as a measure of early AD related symptoms

Many IoT devices have speech recording capacity and because speaking is a cognitively complex task, in the context of cognitive impairment detection have already led to automatic dementia classification (ADC) systems that infer cognitive state directly from digitally recorded speech of neuropsychological tests [6–8]. The potential of analyzing speech as an alternative approach to assessing cognitive state is particularly intriguing because it taps into multiple cognitive domains to produce intelligible content and most people speak, regardless of their education, culture, sex, or language.

Individuals with AD and other forms of dementia exhibit measurable changes in their speech production, seen in both the acoustic domain and the language domain [9]. These changes often precede noticeable cognitive related symptoms [10]. Digital recordings of structured speech, such as from neuropsychological testing, also allow for concurrent validation of speech markers as a surrogate measure of cognition compared to neuropsychological tests [11].

ADC systems that analyze speech recordings aim to detect subtle linguistic and paralinguistic cues (e.g., hesitations, disfluencies, semantic anomalies) indicative of a neurodegenerative disorder. In addition, these systems can also mitigate well-known test question biases [8, 12,13] because the speech analysis can analyze all content and is not limited to scoring parameters that are impacted by test item relevance. Early work on ADC tasks for AD detection (ADD) used classical machine learning algorithms with hand-crafted speech and linguistic features [7, 14]. More recent systems leverage deep learning architectures such as convolutional [11] and recurrent [15] neural networks and neural embeddings from pretrained speech representation models such as wav2vec 2.0 [16,17] and Whisper [18,19] as well as text language models [18,20]. These speech processing methods have led to published studies that evidence the power of speech analysis across the dementia progression spectrum. Fraser et al. (2016), using a computational approach with 370 linguistic and acoustic features, achieved up to 82 % accuracy in classifying AD versus controls from picture descriptions [21]. Eyigöz et al demonstrated that linguistic variables derived from a pre-recorded picture description task could predict future AD onset (almost 15 years in advance) from a cognitively normal baseline with a significant area under the curve (AUC) of 0.74 and an accuracy of 0.70 [22]. Pan et al , exploring different automatic speech recognition (ASR) paradigms and bidirectional encoder representations from transformers (BERT)-based classification from the DementiaBank 2021 publicly available audio only speech dataset called Alzheimer's Dementia Recognition through Spontaneous Speech (ADReSS). They reported test results for their best acoustic-only model at 74.65 % accuracy and their best linguistic-only model at 84.51 % accuracy [23]. García-Gutiérrez et al. (2024), leveraging paralinguistic (acoustic) features combined with socio-demographic data, showed the ability to identify individuals with AD (F1-score = 0.92) and MCI (F1-score = 0.84). They found differentiating MCI from SCD (Subjective Cognitive Decline) yielded an AUC of 0.80, and MCI from AD had an AUC of 0.73 [24]. Agbavor and Liang (2022) found that GPT-3 based text embeddings notably outperformed conventional acoustic feature-based approaches for AD classification [25].

Their model achieved 80.3 % accuracy, 72.3 % precision, 97.1 % recall, and an 82.9 % F1-score on the ADReSSo unseen test set. For Mini Mental State Exam score prediction, a short dementia screening assessment, they found that a Ridge regression model (RMSE) using acoustic features had an RMSE of 6.250, while their GPT-3 Babbage model achieved a lower RMSE of 5.163. Heitz et al. (2024) leveraged GPT-4 to extract five semantic features from spontaneous speech transcripts with up to 93.1 % accuracy on manual transcripts and 90.5 % on ASR transcripts [26]. Amini et al. (2024) used natural language processing (NLP) and machine learning (ML) techniques on recorded neuropsychological test interviews to predict progression from MCI to AD within six years and achieved an accuracy of 78.5 % and a sensitivity of 81.1 %, with a moderate specificity of 75 % [27].

Despite progress in algorithmic design, existing work still focuses on sentence-level speech segments and small datasets such as ADReSS [6, 14] and the Framingham Heart Study [7]. Their study surprisingly found that AD is possible with examiner speech only, indicating examiner bias in administration of standardized neuropsychological tests [8,12,13] that are presumed to follow prescriptive administration protocols . Existing speech-specific ADC systems are fundamentally limited in their ability to process neuropsychological test recordings that are typically long in duration, with many published protocols taking 1+ hours to administer. This constraint often forces segment-level inference using forced alignment or manual heuristics [7,17,18], leading to context fragmentation and a drop in fine-grained classification performance [28].

To date, many voice-related analyses have focused on acoustic features. The advantage is acoustic features can be easily extracted regardless of native language spoken. Available open-source automatic speech recognizer (ASR) such as Whisper have led to transcriptive based pipelines that can utilize natural language processing for analysis, but ASR accuracy is more variable outside of the most commonly spoken languages (e.g., English, Spanish, etc.). ASR suffer from loss of acoustic information as well as error propagation, especially in noisy, spontaneous and multi-speaker conversational settings, whereas acoustics only suffer from loss of language information, such as word choices, sentence structure and content richness. Further, speech generated in a natural context is generally longer in length and involves exchange between two or more speakers. To address the challenges of analyzing long-length digital recordings that include interactive speech, we have proposed leveraging state-space models (SSMs) [29,30], a family of architectures designed for efficient long-sequence modeling.

## 3. Joint acoustic and linguistic analysis of interactive long-length speech recordings

SSMs scale linearly in both space and time, making them ideal for modeling longer-length speech recordings without segmentation. For example, the dementia information in recording of neuropsychological test administration can be subtle and sporadic, with many conversational turns offering little diagnostic value [12]. SSMs' natural capacity for temporal compression allows them to distill salient patterns with minimal information loss, making them particularly well-suited for the ADC task. Below, we present *Dememba* [31], a memory- and compute-efficient architecture trained on over 1000 h of neuropsychological tests with balanced representation across dementia stages.

Fig. 1 shows the overall design of our multimodal dementia classification system, which combines both speech and language analysis. The system consists of four main components: (1) a speech segmenter, (2) an automatic speech recognizer (ASR), (3) an audio-based dementia classifier, and (4) a text-based dementia classifier.

## 4. Audio analysis

The **speech segmenter** takes an hour-long neuropsychological test recording and breaks it into shorter, more manageable segments. It

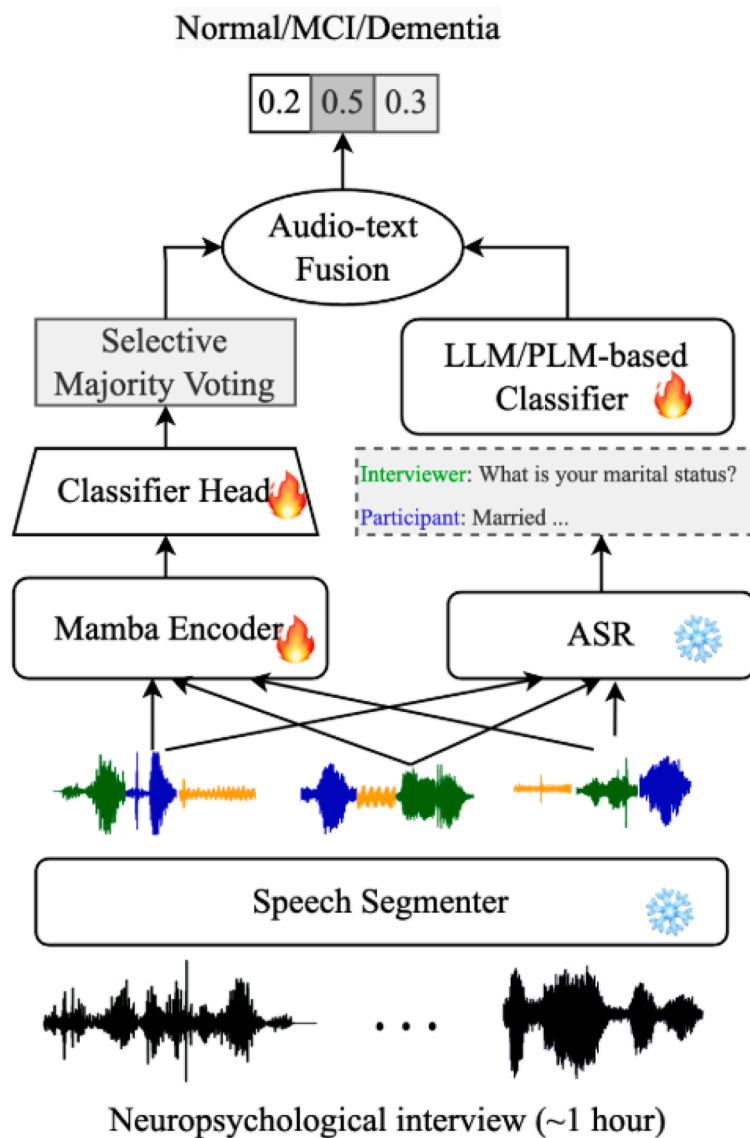


Fig. 1. Overall architecture of the proposed ADC model. The frozen speech segmenter divides the hour-long recording into shorter segments, a trainable SSM-based audio classifier and a trainable text-based text classifier. The predictions from the two classifiers are then combined via late fusion.

separates examiner speech, participant speech and pauses. This not only makes the system more efficient but also allows us to study how factors such as silence duration or examiner/participant speaking turns affect dementia classification.

Next, the **audio dementia classifier** evaluates how the participant sounds. It uses an advanced deep learning model [30,32–35] to capture long-range dependency in the speech. Each segment is assigned a probability of belonging to a dementia category. Since dementia-related cues may appear only in certain moments, the system highlights the most informative segments and weighs them more heavily in the final decision.

### 5. Text analysis

To complement the audio, the text dementia classifier looks at what the participant says. The ASR system, Whisper [19], automatically transcribes the entire recording into text. A large language model (LLM) [36,37] or pretrained language model (PLM) [38] then analyzes the transcripts, assessing whether the participant's responses are coherent, accurate, and consistent with normal cognition. Beyond simply assigning a diagnosis, the LLM can also generate a clinician-friendly narrative

summary, improving interpretability for medical use.

### 6. Decision fusion

Lastly, the system integrates the predictions from both audio and text branches to determine overall dementia severity. This combined approach improves accuracy compared to using speech or text alone. In particular, for 2-class classification, the AUC of the best text-only and audio-only approaches are 91 % and 87 %, while the combined approach achieves an AUC of 95 %. The reason for this synergy may be that audio models tend to capture prosodic cues such as hesitation and intonation, whereas text models tend to capture lexical/linguistic patterns like filler words and semantic incoherence.

### 7. Clinical evaluation

We tested the system on the Framingham Heart Study (FHS) dataset [12], which includes about 11,000 h-long neuropsychological test recordings, 2058 of which have been adjudicated and labeled as cognitively intact ( $n = 936$ ) and cognitively impaired ( $n = 1122$ ). For 3-class fine-grained classification task, we further divide the cognitively

impaired class into mild cognitive impairment (MCI) and dementia classes.

In the two-class setting (normal vs. dementia), our system (Demenba-medium) achieved an area under curve (AUC) of 92 %, surpassing a 6 % improvement over prior methods. For the three-class task (intact, MCI, dementia), performance remained strong with an AUC of 83 %, a 14 % AUC gain over the previous state of the art. Importantly, the advantage of our method grew when distinguishing more subtle differences between MCI and dementia, suggesting that the system is particularly sensitive to early signs of cognitive decline. Our method consistently outperformed prior methods, with the largest gains observed in the more challenging 3-class setting. Importantly, the system scales effectively, handling over 1000 h of training data and reliably analyzing speech segments up to 6 min long — capabilities essential for real-world clinical recordings. Details about the methods described above are provided in Wang et al. [31].

## 8. Concluding remarks regarding speech

Our Demenba analyses highlight the complementary roles of acoustic patterns (how someone speaks), speaker dynamics (who is speaking), and linguistic content (what is said) in dementia classification. These findings provide new insights into how different aspects of speech and language reflect cognitive status. Looking ahead to enhancing the potential of speech, we aim to enhance both the performance and interpretability of our models by integrating them with multimodal LLMs, which can combine speech, language and other clinical as well as digital data. Another challenge is to improve the accuracy of our approach by handling data variability introduced by factors such as accent, gender, age and additional health conditions.

In addition to the publicly available DementiaBank ADReSS, the emergence of other speech datasets such as that from the Alzheimer's Drug Discovery Foundation SpeechDx, a multi-center, longitudinal study that is collecting longitudinal voice recordings from approximately 2000 participants with strong clinical characterization profiles provide resources from which to accelerate the translation of audio recording research to clinical utility. Future directions enabled by these datasets include studying finer-grained classification of dementia subtypes and generalization performance on other dementia datasets, to ensure broader clinical applicability.

At the same time, the increasing diversity of datasets underscores important challenges the field must address. Harmonization across sites and recording protocols is critical to ensure that models trained on one cohort remain valid across others. Equally important are anonymization and de-identification strategies, which safeguard participant privacy while retaining the clinical richness of speech data. Developing methods that balance privacy with utility, while also addressing variability in recording conditions, accents, and disease progression, will be essential for making speech-based biomarkers both reliable and ethically deployable in real-world healthcare.

Further, while our method is a step toward more fine-grained categorization of dementia, it remains to be tested whether our method can be extended to detect more subtle cues of dementia in the early stages of dementia, which are much more challenging than differentiating MCI and AD. More clinically relevant metrics such as the RMSE of predicting the MMSE score can also be assessed, provided that the ground truth scores are available.

The broad penetration of smartphones provides a ubiquitous tool for capturing and processing natural speech during everyday phone conversations, offering a rich and passive means of collecting longitudinal speech and language data. This continual, real-world sampling enables for the detection of subtle changes in vocal, lexical, and syntactic patterns and identify early signs of cognitive decline if used over time.

## 9. Beyond speech: eye as a window to the brain and as a potential digital biomarker

Digital ocular image instruments can track eye movements, the analysis of which is playing an increasingly significant role in AD research due to its potential as a non-invasive tool for early detection and monitoring of the disease. However, like all other cognitive domain testing, the study of eye movements relies on the respective stimuli context such as smooth pursuit, scene viewing, visual search and other. Amongst the various stimuli, reading is a well-defined task that occurs numerous times per day on mobile devices without any prompt. The average person who is literate reads 3000–10,000 words per day on their mobile device. Reading metrics are a highly applicable biomarker in AD research due to the reading process' standardized nature as a well-defined, complex cognitive process whose underlying mechanisms are profoundly impacted by early AD-related cognitive and neurological changes, resulting in quantifiable alterations in eye movement patterns such as increased gaze duration, more fixations, and a loss of the contextual predictability effect.

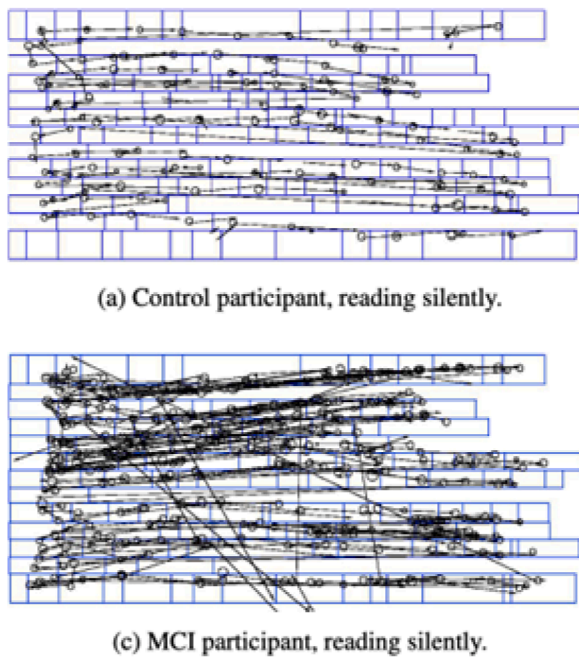
For those who are literate and are not visually impaired, reading is a complex cognitive activity that requires the fine integration of attention, ocular movements, word recognition, language comprehension, working memory, and semantic memory. Many of these cognitive processes, such as attention, inhibitory control, working memory, and decision-making, have been well-documented to be impaired in the early stages of AD/ADRD. Subtle alterations in movement coordination and planning, which are often unnoticed in early AD/ADRD when performing other fine motor tasks like writing, can be precisely detected through eye movement analysis during reading. This is because neurological connectivity changes occur early in the course of AD, disrupting controlled information processing that is critical for reading.

## 10. Previous studies demonstrating validity of ocular movements as a potential AD biomarker

The eye offers an intriguing opportunity to explore the concept of digital indices as digital biomarkers. The advantage of ocular research is the finite and well-defined measurements and the transparency in analysis. To move from a digital measure to a digital biomarker requires validation that can be easily reproduced or replicated. The few studies summarized below evidence the type of results that have much more direct clinical translation, which is necessary for FDA approval as a biomarker. There is also an existing clinical pathway for conducting eye scans, which would allow more easy integration of an ocular biomarker into the clinical workflow.

A series of studies have been able to show that ocular measures are able to capture in a quantifiable manner, cognitively related natural behaviors, such as reading. In an early study, Fernandez et al. 2013 found a sizeable difference of 23 % in outgoing saccade size between AD and Controls [39]. In another study, Fernandez et al. found that participants with mild AD did not show reduced gaze duration when reading highly predictable text as compared to cognitively intact controls, suggesting early impairments in memory-related mechanisms that support contextual word processing [40]. In another follow up study, Fernandez et al. (2015) further validated that participants with mild AD exhibited significantly more total, first-pass, and especially second pass fixations compared to cognitively healthy controls, indicating increased rereading behavior during both regular and high-predictability sentence reading [41]. They also showed fewer single fixations and shorter outgoing saccades, suggesting impaired contextual word processing (see Fig. 2). In general, cognitively healthy readers adjust their gaze based on the predictability of preceding and upcoming words, while those with AD do not, reflecting early deficits in semantic anticipation and memory-guided eye movement control.

Biondi et al. 2018 reported performance of a softmax classifier using a series of features like first pass fixations, unique fixations and multiple



**Fig. 2.** Reproduced from Fraser et al. 2017, showing less scattered single fixations and outgoing saccades during silent reading from (a) control compared to (c) MCI participants.

fixations Refixations was 88.7 % to classify Controls and 91.0 % for AD patients [42]. Taken together, these studies show that ocular measures can differentiate with specificity between those with and without AD. These results are illustrative of the types of ocular studies underway that lend credence to the eventual validation of digital ocular AD biomarkers.

## 11. Digital interactions

Given that over 6.8 billion people are estimated to use a smartphones [43] it is worth emphasizing the unique opportunity they provide for continuous remote monitoring of AD/ADRD related behavioral changes. Embedded within each smartphone are the multiple sensors described above that can collect the raw 3-dimensional digital data streams that AI driven algorithms are interpreting into behavioral measures. Patterns of sensor rhythms reflect executive function, sleep-wake stability, and behavioral regularity, all of which deteriorate with mild cognitive impairment [44]. Kaye et al., (2011) used activity sensors and digital markers that included phone use to track daily regularity as a proxy for cognitive decline [45]. Previous studies have demonstrated that frequent app switching and short dwell times can reflect impulsivity or distractibility while reduced diversity may reflect narrowing interests or cognitive fatigue [46]. Changes in session structure can reflect attention span, fatigue, or executive function breakdown. Slower typing and higher error rates may correlate with motor or processing issues [47]. These studies in the aggregate provide further evidence that cognitive decline is associated with disrupted routines or decreased behavioral regularity, which can all be measured through sensors embedded in every smartphone [48].

## 12. Multi-Modal: the next frontier

Despite the promise of speech and ocular biomarkers, limiting early detection of AD/ADRD clinical symptoms to a single modality would be short-sighted. Cognition is reflected in virtually all bodily movement. Sensors embedded in smartphones, wearables and in home devices also collect behavioral movements. The different sensor modalities in combination, provide a comprehensive multi-modal assessment platform

from which to detect early changes in cognitive and other related behaviors. Further, multi-modal assessments can help circumvent limitations of any single modality measurement. For example, despite the promise of reading-based ocular biomarkers, there are multiple factors beyond literacy levels that can impact accuracy of measurement including educational attainment levels, baseline or concomitant decline in visual acuity from cataracts, glaucoma, macular degeneration or other age-related eye disorders, whether reading materials are in the person's native language, language, etc. Other comorbid health conditions such as hearing loss, musculoskeletal related problems, breathing difficulties, are other examples of factors that can impact data collected from any single digital format. Successful interpretation of high volume and highly variable fluctuating patterns in different person-specific combinations of digital data will likely lead to much more accurate differentiation between normal behavioral fluctuations from those that are reflective of early neuropathological progression. While previous work centered on single sensor modality accounts for much of the current digital adoption into clinical research, trials and care, an inflection point is nearing where there will be a shift to multi-dimensional data streams uniquely customized to the individual, but fueled largely by AI analytic methods that are still able to effectively extract common meaningful information from them. If these AI solutions result in high sensitivity and specificity for AD/ADRD at much lower cost and far greater reach, they will accelerate the digital revolution that is already underway.

Other issues regarding digital biomarkers that needs significant consideration are the validation process and the transition to clinical care. Acquisition of data through digital devices does not automatically mean the resulting measurement is a "digital biomarker". Neither digital voice or eye movements can yet be considered digital biomarkers within the US until they go through the same rigorous validation process that both AD imaging and plasma biomarkers have gone through following FDA guidelines, which includes specific context of use [49]. Further FDA approval alone will not lead to clinical use in the US. Many factors impact the path post-validation such as whether the test is widely accessible, whether test results are easily interpretable or whether clear treatment guidelines are available, particularly when specialized expertise is unavailable [50]. Advances in research that capitalize on approaches that are ubiquitously obtainable can ensure this trend does not have to continue, despite the rise in the number of dementia cases around the world,. Combination therapies that are increasingly being touted [51], comprised of both pharmacologic and non-pharmacologic interventions, could have potential to disrupt the trajectory of AD/ADRD progression to the point that disease symptoms at the clinical diagnosis threshold is never reached. But this vision is contingent on detecting those changes many years if not decades earlier in the insidious onset process. Through smartphones and other IoT devices, single modality digital assessment tools are rapidly giving way to multi-modal ones. AI innovations today may soon be replaced by an even more powerful quantum computing environment. While much is unknown, what is certain is that technological advances are solving the challenge of early detection of AD/ADRD to the entire at risk global population, bringing with it great optimism in eradicating them.

## Consent statement

Consent was not necessary for the purposes of this perspective piece.

## CRediT authorship contribution statement

**Liming Wang:** Writing – review & editing, Writing – original draft. **Jim Glass:** Writing – review & editing, Writing – original draft. **Lampros Kourtis:** Writing – review & editing, Writing – original draft, Conceptualization. **Rhoda Au:** Writing – review & editing, Writing – original draft, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rhoda Au reports a relationship with Novo Nordisk Inc that includes: consulting or advisory, speaking and lecture fees, and travel reimbursement. Rhoda Au reports a relationship with Signant Health that includes: consulting or advisory. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Funding

This work was informed by research supported by the American Heart Association (20SFRN35360180), the National Institute on Aging (2P30AG013846, AG062109, AG068753, AG072654), the Alzheimer's Drug Discovery Foundation (201902–2017835) and Gates Ventures.

## References

- Szekely CA, Thorne JE, Zandi PP, Ek M, Messias E, Breitner JC, Goodman SN. Nonsteroidal antiinflammatory drugs for the prevention of Alzheimer's disease: a systematic review. *Neuroepidemiology* 2004;23(4):159–69.
- Chuang YF, An Y, Bilgel M, Wong DF, Troncoso JC, O'Brien RJ, Breitner JC, Ferrucci L, Resnick SM, Thambisetty M. Midlife adiposity predicts earlier onset of Alzheimer's dementia, neuropathology and presymptomatic cerebral amyloid accumulation. *Mol Psychiatry* 2016;21(7):910–5. Jul. <https://www.fda.gov/news-events/press-announcements/fda-clears-first-blood-test-used-diagnosing-alzheimers-disease#:~:text=The%20Lumipulse%20%20pTau217%2F%C3%9F,and%20symptoms%20of%20the%20disease>.
- de Vries LE, Huitinga I, Kessels HW, Swaab DF, Verhaagen J. The concept of resilience to Alzheimer's Disease: current definitions and cellular and molecular mechanisms. *Mol Neurodegener* 2024 Apr 8;19(1):33. <https://doi.org/10.1186/s13024-024-00719-7>. PMID: 38589893; PMCID: PMC11003087.
- Kurlowicz L, Wallace M. The mini-mental state examination (mmse). *J Gerontol Nurs* 1999;25(5):8–9.
- Luz S, Haider F, de la Fuente S, Fromm D, MacWhinney B. Detecting cognitive decline using speech only: the ADReSSo Challenge. *Interspeech* 2021;2021:3780–4. <https://doi.org/10.21437/Interspeech.2021-1220>. Oct.
- Alhanai T, Au R, Glass J. Spoken language biomarkers for detecting cognitive impairment. In: 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE; 2017. p. 409–16.
- Dawalatabad N, Gong Y, Khurana S, Au R, Glass J. Detecting dementia from long neuropsychological interviews. In: Findings of the Association for Computational Linguistics: EMNLP 2022; 2022. p. 5270–83. Association for Computational Linguistics.
- van den Berg RL, de Boer C, Zwan MD, et al. Digital remote assessment of speech acoustics in cognitively unimpaired adults: feasibility, reliability and associations with amyloid pathology. *Alz Res Ther* 2024;16:176. <https://doi.org/10.1186/s13195-024-01543-3>.
- Young CB, Smith V, Karjadi C, Grogan SM, Ang TFA, Insel PS, Henderson VW, Sumner M, Poston KL, Au R, Mormino EC. Speech patterns during memory recall relates to early tau burden across adulthood. *Alzheimers Dement* 2024;20(4):2552–63. <https://doi.org/10.1002/alz.13731>. Apr; Epub 2024 Feb 13. PMID: 38348772; PMCID: PMC11032578.
- Ding H, Mandapati A, Karjadi C, Ang TFA, Lu S, Miao X, Glass J, Au R, Lin H. Association between acoustic features and neuropsychological test performance in the Framingham Heart Study: observational study. *J Med Internet Res* 2022 Dec 22;24(12):e42886. <https://doi.org/10.2196/42886>. PMID: 36548029; PMCID: PMC9816957.
- Al Hanai T, Au R, Glass J. 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; 2018. p. 746–52. Jun; Volume 2 (Short Papers). Association for Computational Linguistics.
- Pérez-Toro P, Bayer L, Arias-Vergara T, Vázquez-Correa J, Klumpp P, Schuster M, Nöth E, Orozco-Arroyave J, Riedhammer K. Influence of the interviewer on the automatic assessment of Alzheimer's disease in the context of the ADReSSo Challenge. *Proc Interspeech* 2021;2021:3785–9.
- Luz S, Haider F, de la Fuente S, Fromm D, MacWhinney B. Alzheimer's Dementia Recognition through spontaneous speech: the ADReSSo Challenge. In: Proceedings of INTERSPEECH 2020; 2020. p. 2172–6. Oct 25–29.
- Rohanian M, Hough J, Purver M. Alzheimer's dementia recognition using acoustic, lexical, disfluency and speech pause features robust to noisy inputs. *Proc Interspeech* 2021;2021:3820–4.
- Baevski A, Zhou H, Mohamed A, Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations. *Adv Neural Inf Process Syst* 2020; 33:12449–60.
- Balogopalan A, Novikova J. Comparing acoustic-based approaches for Alzheimer's disease detection. *Proc Interspeech* 2021;2021:3800–4.
- Li J., Song K., Li J., Zheng B., Li D., Wu X., Liu X., Meng H. Leveraging pretrained representations with task-related keywords for Alzheimer's disease detection. *arXiv [Preprint]*. 2023 Apr 12;arXiv:2304.06035.
- Radford A, Kim JW, Xu T, Brockman G, McLevey C, Sutskever I. Robust speech recognition via large-scale weak supervision. In: International Conference on Machine Learning (ICML). PMLR; 2023 Jul 23–29. Honolulu, HI. *Proc Mach Learn Res*. 2023;2023:28492–28518.
- Haulcy R, Glass J. Classifying Alzheimer's disease using audio and text-based representations of speech. *Front Psychol* 2021 Mar 26;11:62413721.
- Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis* 2016;49(2):407–22. <https://doi.org/10.3233/JAD-150520>. PMID: 26484921. Epub 2016 Oct 20.
- Eyigoz E, Mathur S, Santamaria M, Cecchi G, Naylor M. Linguistic markers predict onset of Alzheimer's disease. *EclinicalMedicine* 2020 Oct 22;28:100583. <https://doi.org/10.1016/j.eclinm.2020.100583>. PMID: 33294808; PMCID: PMC7700896.
- Pan Y, Mirheidari B, Harris JM, Thompson JC, Jones M, Snowden JS, Blackburn D, Christensen H. Using the outputs of different automatic speech recognition paradigms for acoustic- and BERT-based Alzheimer's dementia detection through spontaneous speech. In: *Proc. Interspeech* 2021; 2021. p. 3810–4. <https://doi.org/10.21437/Interspeech.2021-1519>.
- García-Gutiérrez F, Alegret M, Marquí M, Muñoz N, Ortega G, Cano A, De Rojas I, García-González P, Olivé C, Puerta R, García-Sánchez A, Capdevila-Bayo M, Montreal L, Pytel V, Rosende-Roca M, Zaldua C, Gabirondo P, Tárraga L, Ruiz A, Boada M, Valero S. Unveiling the sound of the cognitive status: machine Learning-based speech analysis in the Alzheimer's disease spectrum. *Alzheimers Res Ther* 2024 Feb 2;16(1):26. <https://doi.org/10.1186/s13195-024-01394-y>. PMID: 38308366; PMCID: PMC10835900.
- Agbavor F, Liang H. Predicting dementia from spontaneous speech using large language models. *PLOS Digit Health* 2022 Dec 22;1(12):e0000168. <https://doi.org/10.1371/journal.pdig.0000168>. PMID: 36812634; PMCID: PMC9931366.
- Heita J, Scheider G, Lander M. The influence of automatic speech recognition on linguistic features and automatic Alzheimer's Disease detection from spontaneous speech. In: The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024); 2024. p. 15955–69. Association for Computational Linguistics.
- Amini S, Hao B, Yang J, Karjadi C, Kolachalama VB, Au R, Paschalidis IC. Prediction of Alzheimer's disease progression within 6 years using speech: a novel approach leveraging language models. *Alzheimers Dement* 2024;20(8):5262–70. <https://doi.org/10.1002/alz.13886>. Aug; Epub 2024 Jun 25. PMID: 38924662; PMCID: PMC11350035.
- Bhati S, Gong Y, Karlinsky L, Kuehne H, Feris R, Glass J. DASS: distilled audio State space models are stronger and more duration-scalable learners. In: Proceedings of the 2024 IEEE Spoken Language Technology Workshop (SLT); 2024. p. 1015–22. Dec.
- Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. In: *Int Conf Learn Represent (ICLR)*; 2022.
- Gu A., Dao T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv [Preprint]*. 2023 Dec 1;arXiv:2312.00752.
- Wang L, Bhati S, Karjadi C, Au R, Glass J. Recognizing dementia from neuropsychological tests with State space models. In: *Automatic Speech Recognition and Understanding Workshop (ASRU)*; 2025.
- Liu Y., Tian Y., Zhao Y., Yu H., Xie L., Wang Y., Ye Q., Liu Y. Vmamba: Visual state space model. *arXiv [Preprint]*. 2024 Jan 23;arXiv:2401.10166.
- Shams S., Dindar S.S., Jiang X., Mesgarani N. Samba: Self-supervised audio representation learning with Mamba state space model. *arXiv [Preprint]*. 2024 May 19;arXiv:2405.11831.
- Zhang X, Zhang Q, Liu H, Xiao T, Qian X, Ahmed B, Ambikairajah E, Li H, Epps J. Mamba in speech: towards an alternative to self-attention. *IEEE Trans Audio Speech Lang Process* 2025;33:1933–48.
- Jiang X, Li YA, Florea AN, Han C, Mesgarani N. Speech Slycerin: examining the performance and efficiency of Mamba for speech separation, recognition, and synthesis. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2025. p. 1–5.
- Touvron H, Lavril T, Izacard G, Martinet X, Lanchaux M.A., Lacroix T., Rozière B., Goyal N., Hambro E., Azhar F., Rodriguez A., Joulin A., Grave E., Lample G. LLaMA: Open and efficient foundation language models. *arXiv [Preprint]*. 2023 Feb 27;arXiv:2302.13971.
- Yang A., et al. Qwen2 technical report. Technical Report. Qwen Team, Alibaba Group; 2024. *arXiv [Preprint]*. 2024 Jun 23.
- Devlin J, Chang MW, Lee K, Toutanova KBERT. Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*; 2019. p. 4171–86.
- Fernández G, Mandolesi P, Rotstein NP, Colombo O, Agamennoni O, Politi LE. Eye movement alterations during reading in patients with early Alzheimer disease. *Invest Ophthalmol Vis Sci* 2013 Dec 30;54(13):8345–52. <https://doi.org/10.1167/iovs.13-12877>. PMID: 24282223.
- Fernández G, Manes F, Rotstein NP, Colombo O, Mandolesi P, Politi LE, Agamennoni O. Lack of contextual-word predictability during reading in patients with mild Alzheimer disease. *Neuropsychologia* 2014;62:143–51. <https://doi.org/10.1016/j.neuropsychologia.2014.07.023>. Sep; Epub 2014 Jul 28. PMID: 25080188.
- Fernández G, Schumacher M, Castro L, Orozco D, Agamennoni O. Patients with mild Alzheimer's disease produced shorter outgoing saccades when reading

- sentences. *Psychiatry Res* 2015 Sep 30;229(1–2):470–8. <https://doi.org/10.1016/j.psychres.2015.06.028>. Epub 2015 Jun 27. PMID: 26228165. Biondi 2018- arXiv: 1702.00837v3.
- [42] Biondi J., Fernandez J., Castro S., Agamennoni O., Eye-movement behavior identification for AD diagnosis. arXiv:1702.00837.
- [43] The Raticati Group. Mobile Statistics Report, 2020-2024. [https://www.radicati.com/wp/wp-content/uploads/2021/Mobile\\_Statistics\\_Report\\_2021-2025\\_Executive\\_Summary.pdf](https://www.radicati.com/wp/wp-content/uploads/2021/Mobile_Statistics_Report_2021-2025_Executive_Summary.pdf).
- [44] Satomi E, Apolinário D, Magaldi RM, Busse AL, Vieira Gomes GC, Ribeiro E, Genta PR, Piovezan RD, Poyares D, Jacob-Filho W, Suemoto CK. Beyond sleep: rest and activity rhythm as a marker of preclinical and mild dementia in older adults with less education. *Neurobiol Sleep Circadian Rhythms* 2024 Dec 25;18:100110. <https://doi.org/10.1016/j.nbscr.2024.100110>. PMID: 39834590; PMCID: PMC11745811.
- [45] Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, Jimison HB, Wild K, Boise L, Zitzelberger TA. Intelligent Systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. *J Gerontol B Psychol Sci Soc Sci* 2011;66 Suppl 1(Suppl 1):i180–90. <https://doi.org/10.1093/geronb/gbq095>. Jul; PMID: 21743050; PMCID: PMC3132763.
- [46] Gordon ML, Gatys LA, Guestrin C, Bigham JP, Trister A, Patel K. App usage predicts cognitive ability in older adults. In: *Proceedings of the 2019 ACM CHI Conference on Human Factors in Computing Systems (CHI '19)*, Glasgow, Scotland, UK; 2019. p. 12. <https://doi.org/10.1145/3290605.3300398>. 4–9 May.
- [47] Alfalahi H, Khandoker AH, Chowdhury N, et al. Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. *Sci Rep* 2022;12:7690.
- [48] Kourtis LC, Regele OB, Wright JM, et al. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *npj Digit Med* 2019;2:9.
- [49] Vasudevan S, Saha A, Tarver ME, et al. Digital biomarkers: convergence of digital health technologies and biomarkers. *npj Digit Med* 2022;5:36. <https://doi.org/10.1038/s41746-022-00583-z>.
- [50] Committee on Policy Issues in the Clinical Development and Use of Biomarkers for Molecularly Targeted Therapies; Board on Health Care Services; Institute of Medicine; National Academies of Sciences, Engineering, and Medicine Graig LA, Phillips JK, Moses HL. Biomarker Tests For Molecularly Targeted Therapies: Key To Unlocking Precision Medicine. Washington (DC): National Academies Press (US); 2016 Jun 30. p. 5. Processes to Improve Patient Care. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK379329/>.
- [51] Salloway SP, Sevingy J, Budur K, Pederson JT, DeMattos RB, Von Rosenstiel P, Paez A, Evans R, Weber CJ, Hendrix JA, Worley S, Bain LJ, Carrillo MC. Advancing combination therapy for Alzheimer's disease. *Alzheimers Dement* 2020 Oct 7;6(1):e12073. <https://doi.org/10.1002/trc2.12073>. PMID: 33043108; PMCID: PMC7539671.

## Glossary of Key Terms

**Automatic Dementia Classification (ADC):** Computational systems that detect dementia from digital signals such as speech recordings.

**Automatic Speech Recognition (ASR):** Technology that converts spoken language into text. **Paralinguistic Features:** Non-verbal aspects of speech such as pitch, pauses, hesitations, or intonation.

**Linguistic Features:** Verbal aspects of speech, including vocabulary, syntax, coherence, and semantics.

**Deep Learning:** Neural network-based machine learning models used for automated feature extraction and classification.

**Convolutional Neural Network (CNN):** A deep learning architecture effective for analyzing acoustic speech signals.

**Recurrent Neural Network (RNN):** A sequence-processing deep learning model used for temporal data like speech

**State-Space Models (SSMs):** Architectures designed for efficient modeling of long sequences, applied here for long speech recordings.

**Language Models (LMs):** Neural models trained on text data to capture linguistic patterns and meaning.

**Area Under the Curve (AUC):** A measure of a model's ability to distinguish between classes.