Artificial Intelligence for Early Detection of Alzheimer's Disease: A Comparative Review of Speech, Neuroimaging, and EEG Biomarkers

Hudson Barber
Harvard-Westlake School
Los Angeles, California USA

1. ABSTRACT

Alzheimer's disease (AD) is a chronic neurodegenerative disorder and the most common cause of dementia globally. Accurate and early diagnosis of AD is essential to enhance patient care and management, facilitate recruitment for clinical trials, and provide access to novel treatments at a stage when intervention may have greater potential benefit. However, the early detection of AD before obvious clinical symptoms have emerged is challenging.

Firstly, there is no unique, widely available screening test for early diagnosis of AD. Secondly, AD progresses through a series of brain and cognitive alterations that resemble those that occur during the typical aging process, making detection of the disease in its early stages difficult since these subtle signs and symptoms can often be mistaken for aging. This overlap between normal aging and early AD-related changes further complicates efforts to identify individuals at risk before substantial cognitive decline occurs. While a combination of neuropsychological assessment, neuroimaging, and analysis of cerebrospinal fluid (CSF) biomarkers can facilitate the diagnosis of AD once symptoms have manifested, these diagnostic tools are not ideal for widespread pre-symptomatic screening due to their high costs, invasiveness, and/or limited availability.

In recent years, artificial intelligence (AI) techniques have been increasingly employed to pre-symptomatically identify subtle changes in speech and language, neuroimaging, and electrophysiological activity including electroencephalography (EEG) and quantitative EEG (qEEG) that are predictive of AD onset. This comparative review discusses recent original studies that have investigated the potential of AI-powered approaches in these three modalities for early detection of AD and discusses the challenges that need to be overcome for these techniques to translate into routine clinical practice.

2. INTRODUCTION

Alzheimer's disease (AD) is a chronic, irreversible neurodegenerative disorder characterized by a steady degeneration of the brain, leading to problems with memory, cognition, and the ability to carry out daily activities. It is the most common form of dementia among people aged 65 and older, with 55 million people affected worldwide (a number expected to reach almost 140 million by 2050)[1]. As the world population ages, AD not only imposes a great psychological and financial burden on the patient and family but also exerts a huge pressure on the health care system. The estimated global cost for dementia, mainly caused by AD, is nearly $2 trillion by 2030[2].

After considerable research over the years, there is no cure for AD. Currently available treatments can postpone the development of symptoms but do not slow the disease course. In the absence of a disease-modifying treatment, the primary goal of clinical management is to alleviate symptoms and retard functional deterioration. As a result, early intervention is thought to be crucial for maximizing the potential impact of therapies as symptoms can only be apparent after significant pathophysiological changes. Early and accurate diagnosis is one of the greatest opportunities to enhance care, refer patients to clinical trials, and enable long-term care planning in this increasingly prevalent disorder with a poor treatment option and late clinical manifestation[2].

However, accurate early diagnosis of AD has not been possible in clinical settings[2]. This is in part due to the fact that neurological changes begin years before clinical symptoms develop. These early symptoms can often be mistaken with normal aging and other conditions. Furthermore, AD progression and presentation are highly variable across individuals, making an early diagnosis extremely difficult. Although various methods including cognitive screening, neuroimaging (MRI and PET), and CSF biomarkers (amyloid beta and tau) are helpful for the late symptomatic diagnosis of AD[3], they are not cost-effective, non-invasive, scalable, and sensitive enough to monitor minor changes in the early phases of AD, leading to under-diagnosis or late-diagnosis of the disease[2].

However, due to the abovementioned constraints, there has been a surge in the use of AI to develop predictive models for early diagnosis of AD based on large and heterogeneous data, including clinical information, neuroimaging, and biomarkers[4]. ML and DL techniques have been applied to MRI and PET images, CSF biomarkers, and clinical data to differentiate between AD, MCI, and healthy control (HC) subjects[3]. Recently, these efforts have been directed toward more accessible digital biomarkers, including speech and language, electroencephalography (EEG) and quantitative EEG (qEEG), and other behavioral data to identify subtle alterations that occur at early stages of the disease[5].

Though most AI-related studies claim excellent performance, they were developed on small, uniform databases and were tested with different methodologies[1]. This discrepancy in the testing methods makes comparison among modalities challenging and also highlights the limitations of current AI models for real-world clinical applications[1].

This paper presents a comparative review of AI-enabled methods for early AD diagnosis based on three non-invasive modalities, namely, speech and language, structural MRI, and EEG/qEEG. These three modalities were selected because of their non-invasive nature and their ease of scalability for repeated screening. They also capture the complementary aspects of AD changes, including speech and language that mirror early cognitive decline, structural MRI that highlights neurodegeneration that becomes more apparent with AD progression, and EEG/qEEG that reflects functional network disruption. Using recently published journal articles, the paper compares the application of modalities, development and evaluation of models, and existing challenges. Direct comparison enables the paper to find out how effective the AI-enabled methods for early AD diagnosis are and what methodological issues should be addressed to enhance interpretability, generalizability and clinical applicability.

3. BACKGROUND

Alzheimer's disease (AD) is an irreversible, progressive neurodegenerative disease that affects memory and cognitive abilities and the ability to perform daily activities[6]. Pathological hallmarks of AD are characterized by amyloid-β (Aβ) plaques and neurofibrillary tangles of hyperphosphorylated tau[6]. Pathological changes occur years before clinical symptoms, leading to synaptic dysfunction, neuronal loss, and slowly to brain atrophy, especially in the medial temporal region, such as the hippocampus and surrounding cortex[3]. The long preclinical phase is significant because it means that detectable biological changes actually precede symptoms. This creates a window where sensitive biomarkers, and the AI models that are trained on them, could enable earlier detection of AD than symptom-based diagnosis.

As AD progresses, it goes through different stages of cognitive decline[2]. Individuals with amnestic mild cognitive impairment (aMCI) present with memory decline, but they remain functionally independent in their daily lives[2]. A longitudinal study found that aMCI subjects have a high risk of developing Alzheimer's disease dementia, so it is crucial for early detection[2]. From now on, MCI will be used as a general term including aMCI unless otherwise specified. A “preclinical” phase of AD has also been identified where individuals show biomarker evidence of amyloid and tau pathology in the absence of objectively measured cognitive deficits on standardized testing[6]. The preclinical and MCI stages are critical because they offer the greatest opportunity for identifying risk and evaluating interventions before significant irreversible degeneration and functional impairment[7]. With that, because amyloid-targeting therapies have produced a very limited clinical benefit, amyloid positivity alone should be treated more as a risk indicator for AD rather than a pathway to intervention.

While there are various assessments to aid in the diagnosis of AD, they all have their own limitations for early detection[4]. Cognitive assessments and clinical evaluations are still the most common instruments used, but they may only capture cognitive decline after significant decline and may struggle to differentiate between other forms of neurodegenerative diseases with similar symptoms[4]. Cerebrospinal fluid (CSF) biomarkers and positron emission tomography (PET) can directly measure Aβ and tau, but they are invasive and expensive and require additional equipment that most clinics are not equipped with[3]. Magnetic resonance imaging (MRI) and other neuroimaging can measure structural changes in the brain associated with AD, but may only capture significant changes in the later stages and may not accurately capture changes due to other diseases and disorders[8].

Thus, there is a need for non-invasive and accessible biomarkers to capture early signs of the disease and to be assessed multiple times throughout the disease trajectory[9]. In addition to neuroimaging techniques, digital (e.g., speech and language) and physiological (e.g., electroencephalogram (EEG)/quantitative EEG (qEEG)) markers have also been explored for early detection of AD[5]. For example, speech samples have the potential to capture subtle declines in lexical diversity, fluency, and discourse and may reflect disruptions to memory and executive functioning[5]. High-throughput MRI analyses can capture subtle patterns of atrophy and structural connectivity-based changes that occur beyond overall volume reduction[10].

EEG/qEEG assessments can capture changes in brain oscillations and functional connectivity that may occur in the absence of structural changes[11].

These candidate biomarkers often generate high-dimensional and noisy data that are not easily explained by conventional statistical analysis[12]. Acoustic speech features, voxel-wise MRI data, and EEG time-frequency data exhibit subtle non-linear patterns that may differentiate early AD from healthy aging or other pathologies[12]. As AI techniques, particularly machine learning and deep learning, are adept at modeling non-linear relationships, data fusion, and learning from raw data, they are particularly well-suited to this problem[5]. For instance, in the neuroimaging literature, deep learning frameworks have been employed to classify AD, MCI, and cognitively normal individuals using whole-brain MRI or PET data[13]. Likewise, several speech and EEG studies have found that AI frameworks are capable of discriminating prodromal or early AD from controls using signals that may not be perceivable by human observers[1].

Given that each data modality captures a distinct aspect of disease, it is unlikely that any one modality will be universally adequate[12]. Taken together, these studies indicate that AI-based digital biomarkers may aid in the detection of early AD beyond standard neuropsychological testing and invasive measures, especially if trained and tested on larger, more diverse samples[5]. In the subsequent sections, we will overview AI-based methods using three promising non-invasive data modalities for early AD detection: speech and language, structural MRI, and EEG/qEEG.

4. METHODS

4.1. Search strategy

To conduct this literature review, a systematic search was performed to identify peer-reviewed articles that investigated the application of AI to detect AD at an early stage using speech digital biomarkers, neuroimaging, and EEG/qEEG physiological signals.

Three databases (PubMed, Google Scholar, and ScienceDirect) were searched from August to November 2025. Given the dynamic nature of research in this field, we prioritized articles published between 2022 and 2025. However, papers published prior to 2022 were also considered if they had a significant impact, were frequently cited, and/or were key references for more recent studies. The search keywords started with broader terms and became more specific as search patterns emerged. Examples of search strings included the term “Alzheimer's” combined with modality- and technique-specific terms such as “artificial intelligence,” “digital biomarkers,” “speech analysis,” “neuroimaging,” “EEG,” and “multimodal early detection.”

The search included both review articles and original studies. The reference list and “cited by” sections of highly relevant articles were explored to identify any additional articles, especially recently published ones.

4.2. Inclusion and exclusion criteria

Inclusion criteria:

Studies were included if they:

- were published in a peer-reviewed journal.
- were written in English and were accessible in full text.
- used AI or machine learning techniques.
- used at least one of the modalities of interest in this review:
    - speech or language biomarkers.
    - MRI or PET neuroimaging.
    - EEG or quantitative EEG (qEEG).
- targeted AD or MCI with an emphasis on early detection (e.g. risk prediction, early-stage classification, screening).
- reported sufficient methodological information and basic performance results (e.g. accuracy, AUC, sensitivity, specificity).

Exclusion criteria:

Studies were excluded if they:

- did not use AI or machine learning techniques.
- targeted non-AD dementias with no AD comparison group.
- were editorials, commentaries, or case reports instead of full articles.
- did not report sufficient methodological information or performance results.

4.3. Study selection process

The initial search across databases yielded between 430 and 450 records, encompassing both original research articles and reviews with a broader scope. First, the titles and abstracts were screened to filter out studies that were obviously not relevant to this review, such as those that focused on drug development or non-AI diagnostic techniques. After screening, the number of articles to be assessed in full text was reduced to about 120.

Of those, 25 studies were selected for inclusion as they fulfilled the inclusion criteria and provided representative and clear evidence across the three modalities of interest:

- AI-based approaches using speech biomarkers.
- MRI- or PET-based neuroimaging models.
- physiological biomarkers based on EEG or qEEG.

The final selection also included a few review and meta-analysis articles that placed the methodologies into a wider context.

4.4. Categorization and data extraction

Each selected article was classified into one of the three main modality categories:

- speech and language biomarkers.
- neuroimaging (MRI, PET, radiomics).
- EEG, qEEG, and other physiological biomarkers.

Several studies on multimodal fusion and explainable AI were also included to facilitate cross-modality comparison. The following data were extracted from each study:

- modality and type of data (e.g. audio, transcripts, MRI images, EEG signals).
- source of the dataset (e.g. Alzheimer's Disease Neuroimaging Initiative (ADNI), OASIS, clinical data).
- type of model (e.g. CNN, transformers, support vector machines, ensemble).
- classification problem (e.g. AD vs. controls, MCI vs. healthy aging).
- performance metrics.
- stated limitations and implications for early detection.

These data were entered into a Zotero library using a standardized format to facilitate modality-wise comparison.

Modality	**What it captures**	**Earliest stage it may signal**	**Strength for screening**	**Biggest constraint for clinical adoption**
Speech & Language	Cognitive-linguistic efficiency	Very early	High	Not AD-specific, affected by other conditions
EEG / qEEG	Functional brain rhythms + network coordination	Early functional disruption	Moderate-low	Protocol variability, limited standardization
Structural MRI	Anatomical atrophy and regional patterns	Later / clearer neurodegeneration	Moderate	Cost / access, weaker signal in preclinical stage

This table helps explain why multimodal models often perform well - speech and EEG can present early functional change while MRI provides clear structural confirmation of AD once degeneration is more prominent.

4.5. Synthesis approach

This review is modality-based, enabling an evaluation of speech-based, neuroimaging-based, and EEG-based approaches separately before comparison. Within each modality, the common modeling techniques, emerging performance patterns, and methodological constraints were summarized. Across modalities, the results were compared based on common evaluation criteria, including scalability, invasiveness, cost, generalizability, interpretability, and strength of validation. This synthesis serves the primary purpose of this review, which is to assess how the various AI-enabled biomarkers contribute to the early detection of AD and determine which approaches currently offer the most reliable and accessible screening for early AD.

5. RESULTS

5.1 Speech-Based Approaches

Speech is commonly proposed as the most promising and feasible digital biomarker for early AD among the literature discussed. As a modality, speech can be collected non-invasively and frequently in daily life, making it more suitable for large-scale screening than imaging or CSF. Alterations in fluency, timing, word-finding, and discourse organization are associated with early strain on memory and executive language control even when deficit is not yet apparent on a standard assessment[14]. As such, speech-based models are most commonly positioned as screening or risk-stratification tools rather than diagnostic tools because speech changes are not unique to AD[15].

Most of the existing speech research used one or both of the two main categories of features[14]. The first category is audio and paralinguistic features derived from the acoustic signal. These features include pitch perturbations, speaking rate, spectral features, pause length, and the frequency of fillers[14]. The second category of features is linguistic features derived from speech transcripts, such as lexical richness, syntactic complexity, semantic appropriateness, and discourse structure[14]. While both categories of features have been found to reflect changes in cognition, they are confounded by different sets of factors. Audio and paralinguistic features are influenced by variations in recording environment, whereas linguistic features are influenced by variations in transcription and language use[14].

Traditional machine learning approaches, such as logistic regression, support vector machines, random forest, and XGBoost are still popular, and deep learning approaches are also used, especially when the sample size is larger, such as convolutional neural networks (CNNs), long short-term memory networks (LSTMs) and transformer-based language models[15]. Some studies have reported similar results across speech-based classification tasks. García-Gutiérrez et al. classified spontaneous speech of subjects with subjective cognitive decline (SCD), mild cognitive impairment (MCI) and Alzheimer's dementia[16]. The best results were obtained when classifying SCD vs AD dementia (F1 = 0.92) and the worst when classifying stages that are closer, such as MCI vs AD dementia (F1 = 0.63). The difference between non-adjacent and adjacent stages is a common trend observed in the results of these speech-based studies, which poses a challenge for the early diagnosis[14]. Interestingly, the same study observed that the performance depends on the sociodemographic features. For example, when sociodemographic features are not used as covariates, the performance for the classification of SCD vs AD dementia decreases (F1 = 0.92 to F1 = 0.81), which supports the need for reproducibility and harmonization in reporting sociodemographic variables and non-speech covariates.

Additional studies have shown excellent results in case-control designs, however, these results should be viewed with caution. For instance, Noto et al. demonstrated near perfect case-control classification performance using acoustic features such as the variability of the spectral kurtosis (AUC = 0.98)[17]. These results demonstrate the upper bound performance that can be obtained

on hand-picked case-control datasets rather than a demonstration of readiness for real-world screening, a point emphasized in more comprehensive reviews of speech tasks[14].

Finally, studies that directly compare audio and text pipelines can be very informative for contrasting the relative contributions of acoustic vs. linguistic signal from the transcript. On the ADReSSo benchmark, Priyadarshinee et al. compared a knowledge-based audio embedding to the use of speech transcripts as input to transformers[18]. On this benchmark, the text-based approach performed significantly better (RoBERTa accuracy 88.7%, F1 0.886 vs VGGish BiLSTM accuracy 78.9%, F1 0.79). This implies that when the quality of the transcripts is good, lexical and syntactic information extracted from the transcript may provide the most discriminative information for this task. However, the authors note an additional concern for deployment: for instance, automated transcription may not preserve the timing of pauses and hesitations, which may be of clinical interest. This sensitivity of performance to preprocessing may contribute to the inconsistencies in speech performance across tasks[14].

Indeed, speech is suitable for the first stage of screening and diagnosis for early detection because of its relatively low cost, ease of data collection, and good responsiveness to mild cognitive impairments[15]. Nevertheless, speech is insufficient as a sole diagnostic tool, particularly for mildly affected patients or in a preclinical phase of the disease, and is sensitive to language, accent, and environmental noises[14]. These results imply that speech might be better suited for screening and diagnosis in conjunction with other modalities, such as EEG or neuroimaging[12].

Overall, this evidence suggests that speech is best understood as a high-sensitivity, low-specificity signal. While it can flag the subtle cognitive inefficiency earlier and more cost-effectively than other clinical tools, it does struggle in other areas that clinical methods excel in - adjacent-stage discrimination and heterogeneous real-world populations. The repeated gap implies that many models are in fact capturing broad impairment rather than AD-specific patterns. As a result, the strongest near-term use case is triage/risk stratification, rather than stand-alone diagnosis.

5.2 Neuroimaging Approaches

Structural MRI, in particular T1-weighted MRI, is the neuroimaging modality most commonly adopted in AI-based Alzheimer's disease classification pipelines[10]. The majority of these studies utilize CNNs trained on whole-brain or region-based MRI images to learn spatial features related to Alzheimer's disease neurodegeneration[13]. With end-to-end DL, the model takes voxel-level MRI volumes, whole brain or ROI crops, as input and learns discriminative spatial features directly from the input, rather than relying on measurements. Aside from end-to-end DL, the next two most prevalent approaches are radiomics frameworks that use conventional ML algorithms with handcrafted image features and multimodal frameworks that incorporate MRI with PET or blood biomarkers[19]. This is needed because structural MRI primarily reflects atrophy and cannot directly measure amyloid burden. So, combining these modalities can link anatomical loss to either earlier or more specific biological markers.

These three approaches illustrate the tradeoff between performance, interpretability, and clinical feasibility. End-to-end DL is most accurate in large and standardized datasets because it can exploit high-dimensional spatial patterns. However, it is most vulnerable to overfitting and limited interpretability when trained on smaller samples. Radiomics, on the other hand, is generally more interpretable and usable for smaller samples than end-to-end DL. Still, it often underperforms when engineered features don't capture the subtle atrophy patterns and when feature selection isn't controlled. Multimodal frameworks are the most grounded, especially when paired with PET or blood biomarkers. They integrate structural change, however, they are harder to use because of their cost. Taking this into account, the best approach often depends on the goal, sample size, and monetary resources available.

For the late-stage problems, many groups have found success using AI models on MRI data, especially on well-vetted research cohorts[10]. The meta-analysis of T1 MRI models by Basanta-Torres et al. found pooled performance close to ceiling when classifying Alzheimer's disease versus healthy controls (SROC AUC 0.985, sensitivity 0.97, specificity 0.95)[10]. But in the same meta-analysis, the performance was lower for each successively earlier problem. For example, the classification of MCI vs. controls is less accurate than the classification of AD vs. controls, reflecting the fact that distinguishing subtle changes in structure in the prodrome from normal aging is a difficult problem. This is an observation of many neuroimaging studies[20]. Structural MRI is very useful after a disease is well established, but earlier-stage classification is difficult, because of both the subtlety of the change and its heterogeneity.

The main limitation of these results is their generalizability[20]. Liu et al. found that training on ADNI and testing on NACC resulted in poorer performance on the latter data set, especially for the MCI-related tasks[13]. The authors of a more general review of machine learning applied to MRI for AD diagnosis and classification say that "external validation is less frequently used in comparison to cross-validation" and that "different scanner types, acquisition protocols, patient demographics and clinical diagnoses could potentially influence the performance of machine learning models"[8]. In practice, independent testing on data from multiple centers has a much larger effect on model reliability than the choice of algorithm[20].

Furthermore, to detect the disease at an earlier stage, various studies used multimodal data fusion where PET and MRI data were fused. PET imaging provided the information of various molecular markers such as glucose metabolism, amyloid beta, and tau whereas, MRI provided the information of the structural changes in the brain[3]. Reviews of PET/MR DL approaches suggest that the fusion of PET and MRI images can improve the classification and staging results of AD by incorporating the information of the molecular markers and anatomical structure[21]. However, the cost and the availability of PET imaging are very high and low, respectively. Moreover, the sample size and data completeness of multimodal datasets are low compared to the unimodal MRI datasets[3,21]. Therefore, the reported improvements in the PET/MRI approach should be cautiously considered by taking into consideration the cost and availability of PET imaging and the quality of external validation[3,21].

Another approach used was the prediction of tau-PET pathology from clinical and other easily accessible biomarkers. This approach can reduce the number of PET imaging acquisitions[22].

The tau-PET prediction results were promising (highest R2 = 0.66–0.69). Furthermore, the results were externally validated. Although the prediction of tau-PET pathology is not equal to AD diagnosis, it can help in the early detection of AD by predicting the disease pathology without using PET imaging for all the subjects[22].

The other approach used was the radiomics approach where the features were handcrafted from either MRI or PET images and machine learning was used as a classifier. Although the results of the radiomics approach can be competitive in a controlled environment, the sensitivity of the approach to preprocessing, segmentation, and harmonization can hinder the reproducibility of the results in different sites. Moreover, for clinical implementation, standardization and multi-site validation of the results are necessary[19].

In summary, neuroimaging-based studies suggest that the structural MRI-based AI models are more accurate in distinguishing AD from healthy controls, whereas the results are less consistent for the separation of MCI from healthy controls and preclinical AD detection[10]. In the studies, the evaluation design plays a significant role in the credibility of the neuroimaging-based results, in particular when the AI models were validated on external, independent, diverse and multi-site cohorts[20].

5.3 EEG-Based Approaches

EEG and qEEG are popular and non-invasive methods of detecting early functional changes associated with AD[11]. While MRI primarily reflects structural changes that accumulate over time, EEG assesses the current state of neural activity and functional connectivity[23]. As a result, EEG and qEEG are often interpreted as potential indicators of early functional network disruption - based on recurring findings such as spectral slowing and altered connectivity - that may emerge before pronounced anatomical atrophy is visible in neuroimaging[11].

Spectral and functional connectivity characteristics are commonly used as features for EEG-based AI models[24]. Band-specific power and phase-locking values are common inputs, as well as graph-theoretical properties such as clustering, path length, and small world index, all derived from EEG FC analyses[25]. Spectral slowing, including increased power in the lower frequency bands (e.g., delta and theta) and decreased power in the higher frequency bands (e.g., alpha and beta), as well as decreased FC, has been consistently reported in AD and MCI patients[11,24,25]. However, since these EEG changes often overlap across AD and MCI, AD-MCI classification is typically less consistent than AD-control discrimination. Further, reported separation is heavily reliant on cohort composition and design.

ML and DL algorithms have been applied to EEG features, with highly variable performance reported in the literature[26]. Acharya et al. found that, in a review of DL studies employing EEG signals, the classification performance varied greatly depending on the classification problem and dataset, with higher performance reported for AD vs. control classification as compared to MCI vs. control[26]. Similarly, in a review of qEEG AI studies, Rezaei et al. reported high accuracy and AUC values in controlled settings, but also emphasized the influence of dataset quality, preprocessing and design on reported performance[24].

One major issue highlighted in the EEG literature is the heterogeneity of acquisition and analysis protocols[23]. The electrode montage, sampling rate, filtering and artifact removal and connectivity measures vary across studies[23,25]. Papaliagkas and Paitel et al. noted that this limits the comparability of studies and contributes to the large performance variability[3,25]. Therefore, the accuracy of EEG-based models should be interpreted with caution, especially when the model was trained and tested on data from a single site[24].

Recently, advanced methodologies accounting for early changes in neural activity have been proposed. Amato et al. proposed a digital-twin framework combining EEG recordings with subject-specific models to infer parameters related to synaptic activity[27]. In that study, the authors found that the digital-twin approach outperformed conventional EEG feature-based classification within the employed dataset[27]. While the results suggest that modeling parameters related to the latent activity of the brain network improves the detection of AD/MCI, the study was performed on a small sample and needs to be repeated and validated in independent datasets before any clinical conclusion can be drawn.

Taken together, EEG-based AI models have demonstrated group differences and moderate to high performance in research settings[24]. The main advantage of EEG recordings is that they can capture changes in network activity that are not evident in structural neuroimaging[11]. However, the variability in acquisition, preprocessing and validation protocols is a major obstacle for the translation of EEG-based models to clinical practice[23]. Standardization of protocols and external validation across multiple centers is thus required to establish whether EEG-based AI models can be translated to clinical practice[24].

6. DISCUSSION

The three modalities presented in this paper focus on different aspects of AD: while speech has the potential to track subtle language changes, EEG can capture functional changes in the coordination of brain networks and MRI structural neurodegeneration[12]. None of the modalities alone provides a complete picture of the underlying disease mechanisms[12]. The changes captured by the three modalities relate to different phases of the disease: whereas language decline can track subtle language changes, EEG functional changes can capture network disruption and MRI structural neurodegeneration[10,11,14]. These phases do not occur at the same pace for all individuals[12].

Capturing this modality-effect interaction is essential for contextualizing the results of each AI model. The accuracy of each modality cannot be compared without acknowledging which stage of the disease each modality has the best predictive ability for and under what circumstances the data was collected[20]. Speech-based models are ideal for early screening due to their low cost and scalability; however, results are contingent upon language, culture, recording conditions, and task design[15]. They also have limited diagnostic specificity, because similar speech changes can and do occur in other dementias, depression, stroke, and normal aging. Due to this, speech is better used as a risk-screening signal rather than an AD discriminator. EEG-based models capture real-time neurological activity; however, accuracy results vary depending on preprocessing and validation techniques[23]. MRI-based models are most accurate

for late-stage disease but have high costs and limited access[10]. Although each modality is most effective under certain conditions, no modality is consistently the most accurate across the board[12]. This pattern supports a stage-aware and triage-based approach in which speech tools are used in broad screening. EEG is leveraged when early functional disruption is suspected, but MRI is reserved for confirmation when clinical concern is elevated.

Since no single modality has the capability of meeting all diagnostic needs, AI multimodal models have been shown to be more accurate than single modality models in ideal research settings[12]. Multimodal AI models are capable of incorporating data from multiple biomarkers and thus capture different features of the disease progression[12]. Speech biomarkers can indicate early changes in cognition, EEG biomarkers capture disruptions in functional networks, and MRI biomarkers capture neurodegenerative changes which generally become more prominent in later disease stages. EEG is valuable in patient context because it can reveal functional impairment even if MRI imaging is still normal. This helps bridge early behavioral screening signals from speech with later anatomical confirmation from MRI. This perspective aligns with the way in which clinicians currently incorporate this data into their practice and mirrors the themes from the digital biomarker literature which emphasize the importance of multimodal modeling rather than single modality classification[9].

While multimodal models have been shown to have high levels of accuracy, it is important to acknowledge that the extent of the gain may be exaggerated in certain settings[12]. Many of the multimodal gains reported are based on highly curated multimodal data sets that have been collected from the same device, time, and place with few missing values, but these may not accurately represent clinical data which is often missing, heterogeneous, and collected across multiple sites[20]. In clinical practice, multimodal AI is not only about obtaining the highest possible accuracy but also about developing systems that can cope with missing modalities and multimodal heterogeneity and return interpretable results[28].

The results from this paper indicate that early detection of AD is unlikely to be accomplished with a single “best” biomarker[5]. Rather, future improvements will require further research into what each digital biomarker can capture and how they can be used together in biologically valid, clinically relevant, and generalizable ways[12]. Speech, EEG, and MRI biomarkers should be viewed as interdependent and complementary tools rather than in competition with one another in the pursuit of early detection of AD[14].

7. LIMITATIONS

In spite of the promising developments of AI-powered solutions for the early diagnosis of AD, there are a number of issues in the existing literature which affect the interpretation of the results. These issues are different across speech, MRI, and EEG and pertain mostly to the availability and representativeness of the data, as well as the lack of external validation, or consistency across studies. These issues may explain why results that are considered excellent for research studies may not be good enough for clinical applications.

7.1 Size and Representativeness of the Data

A number of the highest results reported in the literature are based on relatively small, well-defined, and not so representative clinical datasets[20]. Most of the neuroimaging studies are based on the ADNI dataset which is not representative of real clinical conditions due to differences in the demography of patients, comorbidities, or scanning parameters[13,20]. Most of the EEG and qEEG studies are based on single center studies with a small number of subjects, which may cause the risk of overfitting and insufficient statistical power[23,25]. Similarly, speech datasets such as ADReSS and DementiaBank are limited in terms of language, tasks, and recording conditions[14]. Therefore, high performance may be due to dataset specific patterns rather than patterns that are more generally representative of the clinical population[20].

7.2 External Validation and Variability of Pipelines

Only a few external validation studies are reported in the literature for speech, EEG and neuroimaging. Most of the models were validated on a test set or using a cross-validation technique on the same data[8]. When the models were applied to external datasets, the performance decreased, especially for the diagnosis of the MCI state and other early-stage conditions[13].

Finally, we identify several caveats in the literature. First, methodological decisions regarding EEG preprocessing (e.g., filter settings, artifact removal, electrodes and connectivity metrics used) can impact model performance[25,26]. For speech, factors such as device and task influence results[14]. Greater standardization in these procedures would facilitate between-study comparisons and would provide a clearer understanding of which procedures are most generalizable across datasets and settings[23].

Second, between-study comparisons are hindered by the use of inconsistent diagnostic labels. For example, criteria for MCI, early AD and disease progression vary across studies and cohorts, and even within the same datasets, criteria and cutoff values have evolved over time[2]. This could impact model development and testing, as multimodal studies with inconsistent labels across data types may introduce noise into model training and testing.

7.3 Limited Longitudinal and Preclinical Evidence

Many studies describe their models as instruments for early or preclinical detection, but longitudinal evidence is lacking across data types[5]. The majority of speech, EEG and neuroimaging studies employ a case-control design, comparing individuals with different levels of cognitive impairment at a single time point[1]. Fewer studies use longitudinal follow-up to assess whether biomarkers identified at baseline predict subsequent progression from NC to MCI or AD. Even more novel analyses, such as those using digital-twin models of EEG data to model parameters of underlying neural circuits, are based on short follow-up durations and modest sample sizes[27]. Without longitudinal follow-up, it is challenging to differentiate biomarkers that are stable over time from those reflecting within-person variability.

7.4 Demographic and Device-Related Biases

Similar challenges compromise the external validity of digital biomarkers. Speech and EEG data, for instance, are typically dominated by native English speakers or highly educated participants[11,14]. In neuroimaging studies, this issue is further exacerbated by the typical over-reliance on a handful of large cohorts[20,23]. In all modalities, these issues suggest that, for real-world detection, far fewer AI-identified signatures may generalize across populations, environments and acquisition parameters.

7.5. Lack of evidence on clinical deployment protocols

Reviews have consistently pointed out that AI detection of AD is often assessed outside of clinical protocols[4]. Data quality, recording time, and device quality are not accounted for in most studies. Though MRI consistently performs well on dementia, it is expensive and not feasible for screening. EEG requires an expert for recording and a standard procedure is required. Speech is a viable alternative, but it requires further validation in unstructured recordings. As such, reported results should be taken as an upper-bound until AI models are tested in a real-world setting.

7.6. Sensitivity of multimodal fusion designs

Multimodal designs are often shown to outperform unimodal models [12]. However, results are often reported on small datasets or datasets with very limited heterogeneity[12]. EEG–MRI or speech–EEG datasets are still few, and when available, missing data are imputed[12]. Multimodal AI is a promising alternative for early detection, but it requires further evidence from large, multicenter studies.

8. FUTURE DIRECTIONS

To be deployed in a real-world scenario, the performances reported in AI-based studies for the early detection of AD need to be narrowed to the real-world setting[4,20]. Larger and more diverse datasets are required for all modalities[20]. For speech-based applications, larger datasets with diverse linguistic and cultural backgrounds, as well as more naturalistic recording conditions are needed[14]. In the case of EEG, multicenter acquisition with standardized protocols are needed, whereas neuroimaging studies should move beyond the few large cohorts available[20,23].

Longitudinal designs are still missing in the field[5]. Most studies are cross-sectional, and as such, they do not address whether speech, EEG or imaging features are markers of disease progression[1]. Longitudinal studies are also necessary to disentangle disease markers from features capturing within-subject variability or compensation mechanisms[4].

Standardization is a necessary step toward the translation of AI-based tools to the clinics[23]. Across EEG, speech and neuroimaging studies, acquisition protocols, preprocessing steps and reporting practices are highly heterogeneous and hamper reproducibility across studies, and shared protocols and reporting guidelines should be adopted[12].

Multimodal models are a promising tool for early detection, but they should be designed accounting for missing or partially available modalities and should be able to accept partial data without losing interpretability[12].

This would increase their applicability in the clinics and would reduce their dependency on imputation strategies, which may mask uncertainty. As AI-based detection methods mature, interpretability should also remain a priority[28]. Models providing insights into which linguistic features, neural patterns or brain regions support their predictions will facilitate trust, adoption and clinical use[28]. For clinical translation, explainable AI (XAI) in speech, EEG and neuroimaging can be used to link the output of AI models with interpretable and clinically meaningful outcomes[28].

9. CONCLUSION

The early diagnosis of AD is a challenging scientific and clinical problem that cannot be solved by a single modality of digital biomarkers[4,20]. Each modality of speech, EEG, and neuroimaging represents different levels of pathology. Speech reveals impairments in fluency, naming, and discourse, which are detectable during daily conversations[14,16]. EEG indicates functional alterations in the neural network, which can be detected before any structural changes[24]. Neuroimaging shows the topological patterns of neurodegeneration that can be observed in the later stages of AD[3]. Taken together, these findings suggest that a hierarchy of early diagnosis from cognitive inefficiency to functional and then structural degeneration is supported by each modality of digital biomarkers[9].

In addition, most recent studies have focused on the combination of different levels of information from the cognitive, functional, and structural levels[4,12]. Speech-based approaches provide high accessibility for massive screening, EEG-based approaches are more sensitive to functional alterations in the early stages, and MRI-based approaches are more specific to detect the disease once structural degeneration occurs[12]. A multimodal approach is more in line with the clinical diagnosis of the disease but still needs large-scale and diverse datasets, common acquisition protocols, longitudinal validation, and explainability[12,28].

AI technologies have extended the scope of non-invasive signals for the early diagnosis of AD from the brain signals of speech and EEG to structural neuroimaging signals[20]. A combination of three levels of digital biomarkers (i.e., cognitive, functional, and structural levels) provides a promising solution for developing more accurate, accessible, and clinically reliable diagnostic and screening tools for AD in the near future[14].

References

1. M.M Alam, S. Latifi. A systematic review of techniques for early-stage Alzheimer's disease diagnosis using machine learning and deep learning. *Journal of Data Science and Intelligent Systems*. 2025, https://doi.org/10.47852/bonviewJDSIS52025037.

2. B. Babu, G. Parvathy, F. S. Mohideen Bawa, G. S. Gill, J. Patel, D. S. Sibia, J. Sureddi, V. Patel. Comparing the artificial intelligence detection models to standard diagnostic methods and alternative models in identifying Alzheimer's disease in at-risk or early symptomatic individuals: a scoping review. *Cureus*. Vol. **16**, pg. e75389, 2024, https://doi.org/10.7759/cureus.75389.

3. R. C. Christodoulou, A. Woodward, R. Pitsillos, R. Ibrahim, M. F. Georgiou. Artificial intelligence in Alzheimer's disease diagnosis and prognosis using PET-MRI: a narrative review of high-impact literature Post-Tauvid approval. *Journal of Clinical Medicine*. Vol. **14**, pg. 5913, 2025, https://doi.org/10.3390/jcm14165913.

4. C. Fabrizio, A. Termine, C. Caltagirone, G. Sancesario. Artificial intelligence for Alzheimer's disease: promise or challenge? *Diagnostics*. Vol. **11**, pg. 1473, 2021, https://doi.org/10.3390/diagnostics11081473.

5. Z. Gu, B. Ge, Y. Wang, Y. Gong, M. Qi. Artificial intelligence technologies for enhancing neurofunctionalities: a comprehensive review with applications in Alzheimer's disease research. *Frontiers in Aging Neuroscience*. Vol. **17**, pg. 1609063, 2025, https://doi.org/10.3389/fnagi.2025.1609063.

6. G. Zhang, W. Zhao, S. Lv, Z. Wang, Y. Yi, H. Ma, Y. Lu, W. Yan, J. Teng. Emerging trends in Alzheimer's disease diagnosis and prediction using artificial intelligence: a bibliometric analysis

of the top cited 100 articles. *Digital Health*. Vol. **11**, pg. 20552076251362098, 2025, https://doi.org/10.1177/20552076251362098.

7. A. G. Vrahatis, K. Skolariki, M. G. Krokidis, K. Lazaros, T. P. Exarchos, P. Vlamos. Revolutionizing the early detection of Alzheimer's disease through non-invasive biomarkers: the role of artificial intelligence and deep learning. *Sensors (Basel, Switzerland)*. Vol. **23**, pg. 4184, 2023, https://doi.org/10.3390/s23094184.

8. Zia-Ur-Rehman, M. K. Awang, G. Ali, M. Faheem. Recent advancements in neuroimaging-based Alzheimer's disease prediction using deep learning approaches in e-health: a systematic review. *Health Science Reports*. Vol. **8**, pg. e70802, 2025, https://doi.org/10.1002/hsr2.70802.

9. W. Qi, X. Zhu, B. Wang, Y. Shi, C. Dong, S. Shen, J. Li, K. Zhang, Y. He, M. Zhao, S. Yao, Y. Dong, H. Shen, J. Kang, X. Lu, G. Jiang, L. M. M. Boots, H. Fu, L. Pan, H. Chen, Z. Yan, G. Xing, S. Cao. Alzheimer's disease digital biomarkers multidimensional landscape and AI model scoping review. *NPJ Digital Medicine*. Vol. **8**, pg. 366, 2025, https://doi.org/10.1038/s41746-025-01640-z.

10. S. Basanta-Torres, M. Á. Rivas-Fernández, S. Galdo-Alvarez. Artificial intelligence for Alzheimer's disease diagnosis through T1-weighted MRI: a systematic review. *Computers in Biology and Medicine*. Vol. **197**, pg. 111028, 2025, https://doi.org/10.1016/j.compbiomed.2025.111028.

11. Y. Yuan, Y. Zhao. The role of quantitative EEG biomarkers in Alzheimer's disease and mild cognitive impairment: applications and insights. *Frontiers in Aging Neuroscience*. Vol. **17**, pg. 1522552, 2025, https://doi.org/10.3389/fnagi.2025.1522552.

12. X. Xu, J. Li, Z. Zhu, L. Zhao, H. Wang, C. Song, Y. Chen, Q. Zhao, J. Yang, Y. Pei. A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis. *Bioengineering*. Vol. **11**, pg. 219, 2024, https://doi.org/10.3390/bioengineering11030219.

13. S. Liu, A. V. Masurkar, H. Rusinek, J. Chen, B. Zhang, W. Zhu, C. Fernandez-Granda, N. Razavian. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. *Scientific Reports*. Vol. **12**, pg. 17106, 2022, https://doi.org/10.1038/s41598-022-20674-x.

14. X. Qi, Q. Zhou, J. Dong, W. Bao. Noninvasive automatic detection of Alzheimer's disease from spontaneous speech: a review. *Frontiers in Aging Neuroscience*. Vol. **15**, pg. 1224723, 2023, https://doi.org/10.3389/fnagi.2023.1224723.

15. A. Sharafeldeen, J. Keowen, A. Shaffie. Machine learning approaches for speech-based Alzheimer's detection: a comprehensive survey. *Computers*. Vol. **14**, pg. 36, 2025, https://doi.org/10.3390/computers14020036.

16. F. García-Gutiérrez, M. Alegret, M. Marquié, N. Muñoz, G. Ortega, A. Cano, I. De Rojas, P. García-González, C. Olivé, R. Puerta, A. García-Sanchez, M. Capdevila-Bayo, L. Montrreal, V. Pytel, M. Rosende-Roca, C. Zaldua, P. Gabirondo, L. Tárraga, A. Ruiz, M. Boada, S. Valero. Unveiling the sound of the cognitive status: machine learning-based speech analysis in the Alzheimer's disease spectrum. *Alzheimer's Research & Therapy*. Vol. **16**, pg. 26, 2024, https://doi.org/10.1186/s13195-024-01394-y.

17. S. Noto, Y. Sekiyama, R. Nagata, G. Yamamoto, T. Tamura. Analysis of speech features in Alzheimer's disease with machine learning: a case-control study. *Healthcare*. Vol. **12**, pg. 2194, 2024, https://doi.org/10.3390/healthcare12212194.

18. P. Priyadarshinee, C. J. Clarke, J. Melechovsky, C. M. Y. Lin, B. T. Balamurali, J.-M. Chen. Alzheimer's dementia speech (audio vs. text): multi-modal machine learning at high vs. low resolution. *Applied Sciences*. Vol. **13**, pg. 4244, 2023, https://doi.org/10.3390/app13074244.

19. R. Bevilacqua, F. Barbarossa, L. Fantechi, D. Fornarelli, E. Paci, S. Bolognini, C. Giammarchi, F. Lattanzio, L. Paciaroni, G. R. Riccardi, G. Pelliccioni, L. Biscetti, E. Maranesi. Radiomics and artificial intelligence for the diagnosis and monitoring of Alzheimer's disease: a systematic review of studies in the field. *Journal of Clinical Medicine*. Vol. **12**, pg. 5432, 2023, https://doi.org/10.3390/jcm12165432.

20. R. J. Borchert, T. Azevedo, A. Badhwar, J. Bernal, M. Betts, R. Bruffaerts, M. C. Burkhart, I. Dewachter, H. M. Gellersen, A. Low, I. Lourida, L. Machado, C. R. Madan, M. Malpetti, J. Mejia, S. Michopoulou, C. Muñoz-Neira, J. Pepys, M. Peres, V. Phillips, S. Ramanan, S. Tamburin, H. M. Tantiangco, L. Thakur, A. Tomassini, A. Vipin, E. Tang, D. Newby, The Deep Dementia Phenotyping (DEMON) Network, J. M. Ranson, D. J. Llewellyn, M. Veldsman, T. Rittman. Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: a systematic review. *Alzheimer's & Dementia*. Vol. **19**, pg. 5885–5904, 2023, https://doi.org/10.1002/alz.13412.

21. M. L. Raza, S. T. Hassan, S. Jamil, N. Hyder, K. Batool, S. Walji, M. K. Abbas. Advancements in deep learning for early diagnosis of Alzheimer's disease using multimodal neuroimaging: challenges and future directions. *Frontiers in Neuroinformatics*. Vol. **19**, pg. 1557177, 2025, https://doi.org/10.3389/fninf.2025.1557177.

22. L. Karlsson, J. Vogel, I. Arvidsson, K. Åström, O. Strandberg, J. Seidlitz, R. A. I. Bethlehem, E. Stomrud, R. Ossenkoppele, N. J. Ashton, H. Zetterberg, K. Blennow, S. Palmqvist, R. Smith, S. Janelidze, R. La Joie, G. D. Rabinovici, A. P. Binette, N. Mattsson-Carlgren, O. Hansson.

Machine learning prediction of tau-PET in Alzheimer's disease using plasma, MRI, and clinical data. *Alzheimer's & Dementia*. Vol. **21**, pg. e14600, 2025, https://doi.org/10.1002/alz.14600.

23. V. Papaliagkas. The role of quantitative EEG in the diagnosis of Alzheimer's disease. *Diagnostics*. Vol. **15**, pg. 1965, 2025, https://doi.org/10.3390/diagnostics15151965.

24. S. Rezaei, F. Asadirad, A. Motamedi, M. Kamran, F. Parsa, H. Samimi, P. Ghannadikhosh, M. Zahmatyar, S. A. Hosseinzadeh, H. Arabi. Future of Alzheimer's detection: advancing diagnostic accuracy through the integration of qEEG and artificial intelligence. *NeuroImage*. Vol. **317**, pg. 121373, 2025, https://doi.org/10.1016/j.neuroimage.2025.121373.

25. E. R. Paitel, C. B. D. Otteman, M. C. Polking, H. J. Licht, K. A. Nielson. Functional and effective EEG connectivity patterns in Alzheimer's disease and mild cognitive impairment: a systematic review. *Frontiers in Aging Neuroscience*. Vol. **17**, pg. 1496235, 2025, https://doi.org/10.3389/fnagi.2025.1496235.

26. M. Acharya, R. C. Deo, X. Tao, P. D. Barua, A. Devi, A. Atmakuru, R.-S. Tan. Deep learning techniques for automated Alzheimer's and mild cognitive impairment disease using EEG signals: a comprehensive review of the last decade (2013 - 2024). *Computer Methods and Programs in Biomedicine*. Vol. **259**, pg. 108506, 2025, https://doi.org/10.1016/j.cmpb.2024.108506.

27. L. G. Amato, M. Lassi, A. A. Vergani, J. Carpaneto, S. Mazzeo, V. Moschini, R. Burali, G. Salvestrini, C. Fabbiani, G. Giacomucci, G. Galdo, C. Morinelli, F. Emiliani, M. Scarpino, S. Padiglioni, B. Nacmias, S. Sorbi, A. Grippo, V. Bessi, A. Mazzoni. Digital twins and non-invasive recordings enable early diagnosis of Alzheimer's disease. *Alzheimer's Research & Therapy*. Vol. **17**, pg. 125, 2025, https://doi.org/10.1186/s13195-025-01765-z.

28. V. Vimbi, N. Shaffi, M. Mahmud. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics*. Vol. **11**, pg. 10, 2024, https://doi.org/10.1186/s40708-024-00222-1.