



OPEN ACCESS

EDITED BY

Nidhi Mundada,
University of Pennsylvania, United States

REVIEWED BY

Vannessa Duarte,
Catholic University of the North,
Coquimbo, Chile
Madhurananda Pahar,
The University of Sheffield,
United Kingdom

*CORRESPONDENCE

Sergi Valero
✉ svalero@fundacioace.org

†These authors have contributed equally
to this work

RECEIVED 24 February 2026

REVISED 07 April 2026

ACCEPTED 07 April 2026

PUBLISHED 04 May 2026

CITATION

Blazquez-Folch J, Calm B,
Hinojosa-Calleja A, García-Gutiérrez F,
Alegret M, Muñoz N, Cano A,
Fernández MV, Miguel A, Solivar A,
De Rojas I, Valenzuela-Seba A,
García-González P, Puerta R, Olivé C,
Capdevila-Bayo M, Muñoz-Morales Á,
Bayón-Buján P, Montreal L, Orellana A,
Ortega G, Sanz-Cartagena P,
Rosende-Roca M, Cantero-Fortiz Y,
Gurruchaga MJ, Tarraga L, Butler C,
Montalban X, Boada M, Ruiz A,
Marquí M and Valero S (2026) Listening
for Alzheimer's clues: machine learning
analysis of multidomain speech features
for cognitive impairment screening.
Front. Aging Neurosci. 18:1816747.
doi: 10.3389/fnagi.2026.1816747

COPYRIGHT

© 2026 Blazquez-Folch, Calm,
Hinojosa-Calleja, García-Gutiérrez,
Alegret, Muñoz, Cano, Fernández,
Miguel, Solivar, De Rojas,
Valenzuela-Seba, García-González,
Puerta, Olivé, Capdevila-Bayo,
Muñoz-Morales, Bayón-Buján,
Montreal, Orellana, Ortega,
Sanz-Cartagena, Rosende-Roca,
Cantero-Fortiz, Gurruchaga, Tarraga,
Butler, Montalban, Boada, Ruiz, Marquí
and Valero. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Listening for Alzheimer's clues: machine learning analysis of multidomain speech features for cognitive impairment screening

Josep Blazquez-Folch^{1†}, Berta Calm^{1†},
Adrián Hinojosa-Calleja¹, Fernando García-Gutiérrez¹,
Montserrat Alegret^{1,2}, Nathalia Muñoz¹, Amanda Cano¹,
Maria Victoria Fernández¹, Andrea Miguel¹, Ariadna Solivar¹,
Itziar De Rojas^{1,2}, Alejandro Valenzuela-Seba¹,
Pablo García-González^{1,3}, Raquel Puerta^{1,3}, Clàudia Olivé^{1,3},
Maria Capdevila-Bayo^{1,4}, Álvaro Muñoz-Morales¹,
Paula Bayón-Buján¹, Laura Montreal¹, Adelina Orellana^{1,2},
Gemma Ortega^{1,2}, Pilar Sanz-Cartagena¹,
Maitee Rosende-Roca¹, Yahveth Cantero-Fortiz^{1,2},
Miren Jone Gurruchaga¹, Lluís Tarraga^{1,2},
Christopher Butler^{1,5,6}, Xavier Montalban^{1,2,7}, Mercè Boada^{1,2},
Agustín Ruiz^{1,2,8,9}, Marta Marquí^{1,2} and Sergi Valero^{1,2*}

¹Ace Alzheimer Center Barcelona - Universitat Internacional de Catalunya, Barcelona, Spain,

²Networking Research Center on Neurodegenerative Diseases (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain, ³PhD Program in Biotechnology, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain, ⁴Department of Pharmacology, Toxicology and Therapeutic Chemistry, Faculty of Pharmacy and Food Science, University of Barcelona, Barcelona, Spain, ⁵Department of Brain Sciences, Imperial College London, London, United Kingdom, ⁶The George Institute for Global Health, School of Public Health, Imperial College London, London, United Kingdom, ⁷Department of Neurology and Multiple Sclerosis Center of Catalonia (Cemcat), Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, ⁸Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Science Center, San Antonio, TX, United States, ⁹Department of Microbiology, Immunology and Molecular Genetics, Long School of Medicine, University of Texas Health Science Center, San Antonio, TX, United States

Introduction: Early detection of Alzheimer's disease (AD) is critical for timely intervention, particularly during the mild cognitive impairment (MCI) stage. This study aimed to develop and evaluate a multidomain speech analysis framework to support cognitive screening, biomarker prediction within the amyloid, tau and neurodegeneration (ATN) framework, and estimation of cognitive function across the AD continuum.

Methods: This study analyzed speech from 2,320 individuals spanning the cognitive spectrum-including those with subjective cognitive decline (SCD), MCI, and Alzheimer's disease dementia (ADD)-using three spoken tasks (~3 min) and extracted multidomain features including acoustic, lexical, syntactic, and semantic features. Machine learning models were trained to classify cognitive status, predict ATN biomarker positivity, and estimate scores across six neuropsychological domains.

Results: Multidomain speech models achieved high performance in differentiating cognitive stages, with AUC values of up to 0.94 for SCD vs. ADD and 0.82 for SCD vs. MCI classifications. In biomarker prediction, the

models yielded AUCs of 0.71, 0.74, and 0.73 for ATN classification, respectively. Speech-based models also showed strong correlations (up to 0.83) with cognitive function scores. Feature importance analysis revealed that verbal fluency measures were the most predictive. Explainability analyses indicated minimal dependency on age, sex, or education, supporting model fairness.

Discussion: These findings show that multidomain speech features capture clinically and biologically relevant information across the AD continuum, enabling cognitive classification, biomarker prediction, and cognitive estimation. These results underscore the potential of speech analysis as a non-invasive, accessible tool for scalable cognitive screening and early detection of AD.

KEYWORDS

Alzheimer's disease, cerebrospinal fluid, early diagnosis, machine learning, mild cognitive impairment, neuropsychological tests, screening

Introduction

Alzheimer's disease (AD) and other dementias affect more than 55 million individuals worldwide (World Health Organization [WHO], 2024). Approximately 39.5% of dementia cases in older adults remain undiagnosed (Amjad et al., 2018), while only 8% of older Americans living with mild cognitive impairment (MCI) receive a formal diagnosis (Liu et al., 2024). This lack of early detection is concerning, as individuals with MCI are at a significantly higher risk of developing AD, with annual conversion rates ranging from 3 to 15% (Liu et al., 2023). The consequences of underdiagnosis include delayed access to potentially beneficial treatments, reduced opportunities for care planning, and greater emotional, physical, and financial burdens on individuals and their caregivers (Power et al., 2023; World Health Organization [WHO], 2024). Thus, improving screening, detection, and early diagnosis of cognitive impairment is a critical public health priority to mitigate these risks and improve patient and family outcomes.

AD manifests through progressive cognitive decline, typically impacting memory, language, executive functions, and behavior. These impairments are rooted in the accumulation of amyloid- β (A β) plaques and neurofibrillary tangles of phosphorylated tau, detectable through neuroimaging and fluid biomarkers and resulting neurodegeneration (Alzheimer's Association, 2025; Molinuevo et al., 2018; Peña-Casanova et al., 2012). Neuropsychological batteries have traditionally served as the primary assessment tool for detecting cognitive impairment (Alegret et al., 2012; Espinosa et al., 2013). However, these evaluations often require the physical presence of specialized clinicians in memory clinics, making them resource-intensive and time-consuming. In response to these challenges, numerous online and remote cognitive screening protocols have been proposed to expand accessibility (Alegret et al., 2024; Ding et al., 2022; Meier et al., 2021).

In recent years, digital technologies have revolutionized the landscape of detecting cognitive decline, introducing scalable and effective solutions. Automated speech analysis, in particular, has emerged as a promising tool in detecting AD at its earliest stages (García-Gutiérrez et al., 2023; García-Gutiérrez et al., 2024;

Martínez-Nicolás et al., 2021). Language deficits, often evident years before AD progresses to the dementia stage, provide valuable insights into MCI (Sperling et al., 2013; Szatloczki et al., 2015). Speech-based diagnostics leverage advanced machine learning (ML) and natural language processing techniques to analyze linguistic and paralinguistic features from simple speech tasks (Toth et al., 2018). This approach not only facilitates non-invasive and remote longitudinal assessments but also aligns with the growing trend toward e-health technologies and decentralized clinical trials (Tröger et al., 2022).

However, despite the rapid growth of speech-based approaches for cognitive impairment detection, several methodological and translational gaps remain. Prior studies have demonstrated the utility of individual feature domains—including acoustic parameters (Xue et al., 2021), lexical and syntactic measures (Fraser et al., 2016; Pekkala et al., 2013), semantic and graph-based representations (Ambadi et al., 2021), and deep learning-based language embeddings (Amini et al., 2023; Lindsay et al., 2021)—for discriminating cognitive status. Multidomain and multimodal frameworks have also been proposed (Chou et al., 2024; Hajjar et al., 2023; He et al., 2023; Nyongesa et al., 2025), and recent work has explored associations between speech features and amyloid burden in both cognitively unimpaired individuals and those with subjective cognitive decline (Mueller et al., 2021; Verfaillie et al., 2019), as well as longitudinal prediction of AD progression (Amini et al., 2024). Nevertheless, many studies are limited by relatively small or homogeneous cohorts, focus on a restricted set of linguistic domains, lack biomarker-confirmed stratification, or provide limited interpretability of ML models. Furthermore, while amyloid prediction from speech has been explored, comprehensive modeling of the ATN (amyloid, tau, neurodegeneration) framework using multidomain spontaneous speech features remains underdeveloped.

Building on this broader literature, the present study introduces a unified multidomain speech analysis protocol that integrates acoustic, lexical, syntactic, and semantic features within a biomarker-characterized cohort spanning the AD continuum. While previous investigations—including our own—have demonstrated the value of acoustic features for cognitive stage discrimination and amyloid status prediction (García-Gutiérrez

et al., 2023; García-Gutiérrez et al., 2024), the current work extends prior approaches in several meaningful ways. First, it systematically incorporates multiple linguistic domains within a single analytical framework to evaluate their complementary diagnostic value. Second, it evaluates prediction not only of clinical status but also of ATN biomarker profiles, thereby aligning speech-based modeling with contemporary biological definitions of AD. Third, it applies advanced feature-selection and optimization strategies to enhance model robustness and interpretability in a larger and more diverse sample. By integrating multidomain linguistic analysis with biomarker-informed stratification and explainable machine-learning techniques, this study aims to advance speech-based digital biomarkers toward clinically scalable, biologically grounded tools for early AD detection.

Materials and methods

Study participants

This study included data from individuals evaluated at the Memory Clinic of Ace Alzheimer Center Barcelona between March 2022 and February 2024 (Table 1). Patients were referred to the Memory Clinic by their General Health Practitioner due to subjective memory complaints, or attended the Open House Initiative without a prior physician referral (Abdelnour et al., 2017).

All participants underwent a series of neurological, neuropsychological, and social evaluations at Ace. The clinical assessment included the Neuropsychological Battery of Fundació Ace (NBACE) (Alegret et al., 2012), the Spanish version of the Mini-Mental State Examination (MMSE) (Blesa et al., 2001; Folstein et al., 1983), the memory test of the Spanish version of the 7 min screening neurocognitive battery (del Ser et al., 2006; Solomon and Pendlebury, 1998), the short form of the Neuropsychiatric Inventory Questionnaire (NPI-Q) (Boada et al., 2006), the Hachinski's Ischemia Scale (Hachinski et al., 1974), the Blessed Dementia Rating Scale (Blessed et al., 1968), and the Clinical Dementia Rating (CDR) (Morris, 1993). For each participant, the final diagnosis was determined by a multidisciplinary team of neurologists, neuropsychologists, and social workers in a diagnostic consensus conference (Boada et al., 2014).

The cohort included individuals with different degrees of cognitive impairment: (1) SCD (CDR = 0) which refers to the self-perception of cognitive problems, including memory loss, without impairment on standardized cognitive test (Jessen et al., 2014), (2) MCI (CDR = 0.5), which refers to cognitive complaints with impairment on standardized cognitive test according to age and educational level and preserved autonomy (Petersen, 2004), (3) AD dementia (ADD) (CDR > 0.5) (McKhann et al., 2011), and (4) non-AD dementia (CDR > 0.5). Within the MCI group, patients presented an amnesic (aMCI) or a non-amnesic (naMCI) profile (Espinosa et al., 2013). Within the ADD and non-AD dementia groups, patients were divided into mild (CDR = 1), or moderate dementia (CDR = 2). Within the non-AD dementia group, patients presented vascular dementia, Lewy body dementia or frontotemporal dementia behavioral variant. Group differences in continuous variables were assessed using the Kruskal–Wallis test. When significant, pairwise comparisons were performed using

Dunn's post hoc test with Bonferroni correction for multiple testing. Statistical significance was set at $p < 0.05$.

A subset of patients with a diagnosis of MCI underwent lumbar puncture (LP) for the assessment of AD-core biomarkers in CSF. All clinical and biomarker measures were obtained within a 6-month window from the speech protocol administration.

Neuropsychological data

To assess cognitive performance, the neuropsychological tests from the NBACE battery were organized into several cognitive domains (Alegret et al., 2012). Neuropsychological composites were estimated using structural equation modeling, guided by exploratory factor analysis and expert consensus. Seven composites were analyzed: memory, attention, visuospatial/visuoperception, executive functions, language, orientation, and praxis. This study employed the neuropsychological tests, their domain groupings, and composite estimation as previously described elsewhere (Alegret et al., 2024; García-Gutiérrez et al., 2024).

Lumbar puncture and quantification of cerebrospinal fluid ATN biomarkers

LPs were conducted at Ace Alzheimer Center Barcelona by an experienced neurologist under fasting conditions. The procedure adhered to the guidelines outlined by the Alzheimer's Biomarkers Standardization Initiative (Vanderstichele et al., 2012). CSF samples were passively collected using 10 mL polypropylene tubes (Sarstedt Ref 62.610.018) min processed within 2 h of acquisition by centrifugation ($2,000 \times g$ for 10 min at 4°C). Following centrifugation, the samples were divided into aliquots and stored in polypropylene tubes (Sarstedt Ref 72.694.007) at -80°C for later analysis. On the analysis day, a 0.5 mL aliquot was thawed to measure A β 1-42, A β 1-40, P-tau181, and T-tau levels. These protein concentrations were determined using a chemiluminescence enzyme immunoassay with the Lumipulse G 600 II automated platform (Fujirebio Europe, Göteborg, Sweden) (Leitão et al., 2019). Thresholds established by the CSF program at Ace Alzheimer Center Barcelona defined amyloid positivity as an A β 1-42/A β 1-40 ratio below 0.069; and tau (T) positivity as a P-tau181 over 54; and neurodegeneration (N) positivity as a T-tau above 412 pg/mL (Orellana et al., 2022).

Speech data collection and preprocessing

Audio was recorded using the built-in microphone of the 8th-generation iPad at a standard sampling rate (44.1 kHz) and stored in digital format. Recordings were performed in Spanish in a quiet clinical setting under supervision, minimizing environmental noise and ensuring consistent acquisition conditions across participants. The protocol included three tasks: (1) participants were presented with The Cookie Theft Picture (Cummins, 2019) and they were asked to provide a detailed description of the image (IMG task); (2) a semantic verbal fluency test was conducted where participants

TABLE 1 Clinical and sociodemographic characteristics of the sample stratified by diagnosis.

Variable	SCD	naMCI	aMCI	ADD	Non-AD Dementia
Sample size (n, %)	249 (10.7)	465 (20.0)	744 (32.1)	682 (29.4)	180 (7.8)
Age (mean, SD)	66.4 (10.2)	72.4 (10.0)	75.2 (8.5)	80.8 (6.4)	80.7 (7.0)
Sex (% female)	63.1	66.0	60.5	67.9	51.7
Years of formal education (mean, SD)	12.4 (4.1)	9.4 (4.6)	8.4 (4.5)	7.4 (4.5)	6.7 (4.3)
MMSE (mean, SD)	29.3 (0.9)	28.2 (1.7)	26.2 (2.8)	21.7 (3.4)	22.8 (3.5)

SCD, subjective cognitive decline; naMCI, non-amnesic mild cognitive impairment; aMCI, amnesic mild cognitive impairment; ADD, Alzheimer's disease dementia; AD, Alzheimer's disease; MMSE, Mini-Mental State Examination; SD, standard deviation.

were requested to list as many animals as possible within 1-min (SVF task); (3) participants were asked to relate a positive personal experience from the past month (PPE task).

To further characterize the audio quality and recording conditions, several signal-level metrics were computed across all recordings. Summary statistics for recording duration and signal-to-noise ratio, as well as transcription-derived token counts, are reported in [Supplementary Table 1](#).

Subsequently, the audio recordings from these three tasks were resampled to 16 kHz and automatically transcribed using the Whisper large-v2 speech-to-text model ([Radford et al., 2022](#)). Subsequently, the stanza (v1.8.2) library was applied on the resulting transcriptions to perform the part-of-speech tagging (POS-tagging) by assigning different lexical categories to each of the transcription tokens ([Qi et al., 2020](#)). The generated audio, the transcriptions, and their annotations represented the base information used to extract the features described in the following section (section 2.5).

Multidomain speech features extraction

Speech features were extracted from the audio data (described in section 2.4) and grouped into five domain categories: acoustic, lexical, syntactical, semantic, and task-specific features. The following sections detail the computation of the features used for the different modeling tasks.

Acoustic

Acoustic features enable the characterization of alterations in voice production and paralinguistic aspects such as prosody or speech-related traits. In this study, acoustic features were extracted using the Python libraries OpenSmile (v2.5.0) ([Eyben et al., 2010](#)) and librosa (v0.10.2) ([McFee et al., 2015](#)).

First, different statistics (mean, coefficient of variation, quartiles, interquartile range, average of positive and negative slopes, skewness, and kurtosis) were calculated from the low-level descriptors (LLDs) of the eGeMAPs feature set ([Eyben et al., 2016](#)). A symmetric moving average of three frames was used to compute the LLDs using frames of 20 ms. Among the LLDs used were frequency-related parameters such as the frequency and bandwidth of the first three formants, which are directly related to speech production; the logarithm of the fundamental frequency starting at 27.5 Hz (F0), which represents pitch; and jitter, which measures the variability of the period of the F0. In addition, energy/amplitude parameters such as shimmer, loudness, and harmonics-to-noise ratio, all related to speech quality, were considered. Moreover,

spectral parameters such as the relative energy of the first three formants, the alpha ratio, the Hammarberg index, spectral slopes (0–500 Hz and 500–1,500 Hz), harmonic differences (H1–H2 and H1–A3), spectral flux, and the first four Mel-frequency cepstral coefficients were calculated.

Second, using the librosa library, mean, standard deviation, skewness, and kurtosis statistics were computed for several acoustic parameters, using librosa default parameters. The analyzed features included chromagram characteristics, root mean square, zero crossing rate of the signal, tempo, onset envelope, spectral bandwidth, and spectral centroid.

All of the described features were calculated for voiced segments, except for loudness, which was computed across the entire audio. Alpha ratio, spectral slopes (0–500 Hz and 500–1,500 Hz), and Hammarberg index were calculated for both voiced and unvoiced segments. Additionally, temporal parameters such as the mean, standard deviation, and quartiles of the voiced and unvoiced segments were included in the acoustic parameter set. The open-source speaker-diarization-3.1 model from pyannote ([Bredin et al., 2020](#)) was employed to perform voice activity detection.

Lexical

Lexical features characterize the use of linguistic elements related to vocabulary and word usage, two attributes affected by cognitive impairment ([Ahmed et al., 2013](#); [Fraser et al., 2016](#); [Garrard et al., 2005](#)). Collectively, these features quantify the richness of information communicated in speech. In this study, lexical features were calculated using POS-tagging information and word frequency norms ([Speer, 2022](#)).

Based on POS-tagging information, the ratios of different lexical categories were considered by calculating the proportions of nouns, verbs, adjectives, adverbs, auxiliaries, and adpositions relative to the total number of words. Additionally, ratios between specific lexical categories, such as adjectives to nouns, adjectives to verbs, and nouns to verbs, were considered ([Ahmed et al., 2013](#)). Other metrics, such as the measure of textual lexical diversity (MTLD) as presented in ([McCarthy and Jarvis, 2010](#)), the moving-average type-token ratio (MATTR), and the Uber, Guiraud, MassLog, and Herdan indices, were calculated ([Covington and McFall, 2010](#); [Kurdi, 2023](#)). The MTLD was computed using thresholds of 0.65, 0.70, and 0.75, and it was recalculated considering only nouns, verbs, adjectives, and adverbs. The MATTR was calculated using a window size of 10 tokens with a five-token stride, considering only nouns, verbs, adjectives, and adverbs.

Based on word frequency norms, using the Python library wordfreq (v3.1.1) ([Speer, 2022](#)), the mean and standard deviation of

the logarithm of word frequency excluding functional words were determined.

Features based on POS-tagging information were calculated only for the IMG and PPE tasks, while those based on word frequency were applied to all tasks.

Syntactical

The syntactical features evaluate the structure and complexity of spoken sentences, providing information on the grammatical organization of the discourse. In this study, scores were extracted from the dependency and constituency trees using stanza (v1.8.2) (Qi et al., 2020) and based on the universal dependencies format (Nivre et al., 2020).

From the constituency-based trees, the Yngve and Frazier indices were calculated, as well as the maximum and average depths of the generated syntactic trees (Voleti et al., 2020). For each metric and task, the average value of the indices for all sentences was considered. Alternatively, for the dependency-based trees, the maximum and average number of root dependencies across all sentences were computed (Voleti et al., 2020). Syntactical features were computed only for the IMG and PPE tasks.

Semantic

This group of features includes metrics that analyze the conceptual information of the words used in speech. Within the group of semantic features, norm-based and data-driven features were considered.

First, based on published Spanish norms of word age of acquisition (Alonso et al., 2015), the mean and standard deviation of word age of acquisition in each patient's speech were computed. These features were extracted for all tasks, providing a baseline measure of linguistic maturity.

Second, for the IMG and SVF tasks, word embeddings were computed using a word2vec model (Mikolov et al., 2013) trained ad hoc with transcription data from our database, excluding functional words. Subsequently, various features were extracted from the generated word embeddings to analyze the richness and complexity of the semantic elements of the discourse. Specifically, the Euclidean distance and cosine similarity of the word embeddings were considered, as well as the mean and standard deviation of the pairwise distances between words. Additionally, spectral clustering was performed on the entire vocabulary, stratified by task, and each word was assigned a cluster membership. This analysis allowed us to examine the number of clusters visited by each subject, the number of transitions between clusters, the number of words within each cluster, the frequency of repetition of previously visited clusters, and the Euclidean distance and cosine similarity between pairs of visited clusters. Overall, these metrics provide insight into the semantic coherence and diversity of the language.

The Python library gensim (v4.3.0) (Rehurek and Sojka, 2011) was used for training the word2vec models. For spectral clustering, Scikit-Learn (v1.5.1) (Pedregosa et al., 2012) was employed.

Task-specific features

In addition, specific features were extracted by exploiting the unique characteristics of each speech task.

For the IMG task, the features proposed in Ambadi et al. (2021), based on spatial semantic graphs, were used. These features aim to characterize the participant's visual narrative during the image description. Specifically, the following metrics were calculated:

the number of content information units (CIUs) mentioned, the distance between different CIUs, the mean and standard deviation of the distance traveled on the vertical and horizontal axes, the mean number of words used to describe each CIU, the proportion of keywords related to the CIUs relative to the total number of words (concreteness), the number of times a participant repeated the same CIU (both consecutively and non-consecutively), the number of times the participant moved between two CIUs within the same quadrant of the image, the proportion of times the participant moved between CIUs in different quadrants versus the same quadrant, and the ratio of CIUs mentioned in each quadrant to the total number of CIUs in the quadrant.

Relative to the SVF task, the following features were considered: the total number of animals mentioned, the number and proportion of unique animals to the total, the proportion of non-animals, the number of repeated animals, and the sum and mean number of animals between repeated animals.

Finally, given the properties of the PPE task, sentiment features were extracted using the model presented in Pérez et al. (2023). The sentiments considered were a positive, negative, or neutral description, and the detection of emotions such as sadness, happiness, anger, disgust, fear, and surprise.

Machine learning modeling

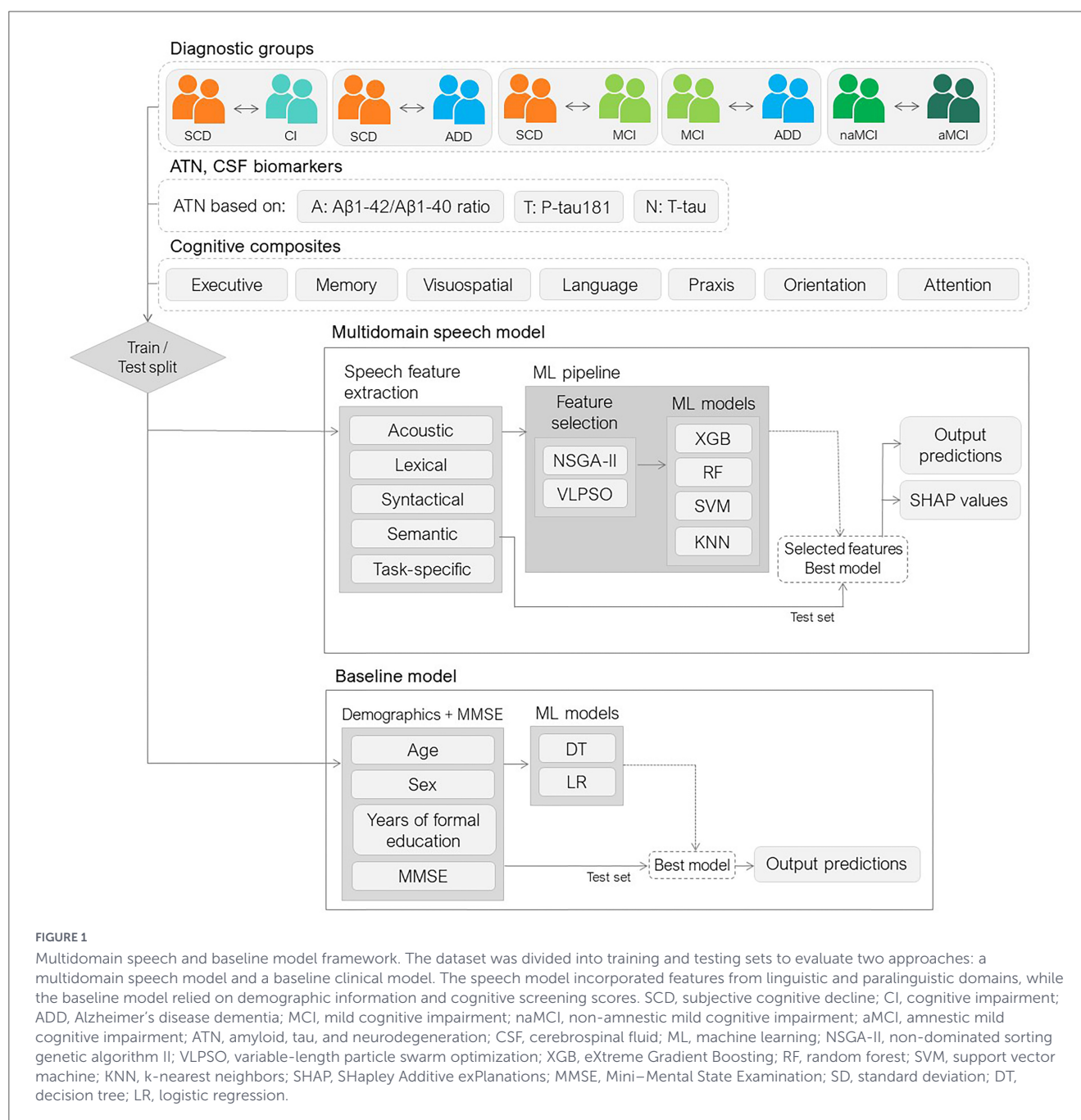
Using the feature set described in section 2.5, ML models were trained for both classification and regression tasks. The framework used is summarized in Figure 1.

Classification tasks focused on diagnosing different cognitive conditions included: cognitive impairment (CI) detection (SCD vs. MCI and dementia), ADD identification (SCD vs. ADD; and MCI vs. ADD), MCI detection (SCD vs. MCI), and MCI subtype discrimination (naMCI vs. aMCI). Also, classification models were used for the prediction of ATN status.

Regression tasks aimed to predict cognitive composite values (as detailed in section 2.2).

The data were divided into training (80%) and testing sets (20%), stratified by diagnostic condition, with subjects matched for age, sex, and education within each set. Differences between training and test sets within each diagnostic group were assessed using Mann-Whitney U tests for continuous variables and chi-square tests for sex. Bonferroni correction was applied to account for multiple comparisons ($p < 0.05$). No significant differences were observed between training and test sets across diagnostic groups for demographic and clinical variables (all Bonferroni-corrected $p > 0.05$), supporting the adequacy of the data split. All model development steps (feature selection and hyperparameter optimization) were performed exclusively on the training set, while the test set was held out and used only for the final evaluation of the models. Furthermore, to evaluate the predictive ability of each speech test, models were trained for each task using each test independently (IMG, SVF, or PPE) and combined (IMG + SVF + PPE).

The models employed for each problem included Random Forest (RF), eXtreme Gradient Boosting (XGB), and Support Vector Machine (SVM). For all tasks and models, an initial feature selection step was conducted based on the training data. The well-established non-dominated sorting genetic algorithm



II (NSGA-II) (Deb et al., 2002) was used for feature selection, optimizing the balanced accuracy (BA) for classification tasks and the mean absolute error (MAE) for regression tasks, along with the number of selected features (Garcia-Gutierrez et al., 2022; Xue et al., 2016). Additionally, the variable-length particle swarm optimization (VLPPO) (Tran et al., 2019) was used for feature selection in the ATN status prediction as it showed the best performance in this task in previous works (García-Gutiérrez et al., 2023). The performance metrics to be optimized were evaluated on the validation set using five-fold cross-validation (CV). An SVM was used as classification/regression model within the NSGA-II and k-nearest neighbors (KNN) were used within the VLPPO.

After feature selection, the hyperparameters of all models were optimized using Bayesian optimization with a tree-structured Parzen estimator as the surrogate model (Bergstra et al., 2011).

During this optimization, the BA and MAE were maximized or minimized respectively, using the training set, and optimizing the metrics using the validation data from a 5-fold CV. A total of 1,000 iterations were performed in the hyperparameter optimization, with the first 500 iterations conducted randomly.

After completing feature selection and hyperparameter optimization, the final models were retrained on the full training set and evaluated on the independent test set. Performance on the test set reflects the generalization ability of the final model on unseen data.

Additionally, ML models trained on multidomain speech features were compared to baseline screening models trained on MMSE and Demo (demographics: age, sex, and years of formal education). The ML models used for baseline models were the decision tree and logistic regression.

All models were developed in Python (v3.11.10). The Scikit-Learn (v1.2.2) (Pedregosa et al., 2012) implementation was used for the RF and SVM models. For the XGB algorithm, XGBoost (v2.1.1) (Chen and Guestrin, 2016) was adopted. The pymoo (v0.6.1.3) (Blank and Deb, 2020) library was used to implement the NSGA-II, the pywinEA2 (available on GitHub) was used to implement the VLPPO and hyperparameter optimization was carried out using optuna (v3.6.1) (Akiba et al., 2019). For more details on the hyperparameters and the search space of each model, refer to [Supplementary Appendix 2](#).

Machine learning explainability

For each task, the best-performing model was selected. Subsequently, explainable AI (XAI) techniques were employed to assess the interpretability of the models.

SHapley Additive exPlanations (SHAP) values were used to gain insights into model predictions and feature importance. SHAP values were computed for the test set to explain individual predictions by attributing importance to each feature highlighting the most influential features (Lundberg and Lee, 2017). To aid interpretation at a higher level, SHAP feature importance values were grouped by feature domain. However, since larger groups can accumulate more importance by virtue of having more features, we normalized the total SHAP importance of each group by dividing by the number of features it contained. This adjustment ensures comparability between groups of different sizes and reduces the bias favoring larger groups (Gregorutti et al., 2015).

Finally, a sex-based analysis examined whether models were prone to sex-based errors. A chi-square test compared sex in the original dataset and misclassified samples. Biases related to age and year of formal education were assessed using a t-test. To explore sex-based feature importance differences, SHAP values were grouped by feature domain, aggregated per sex, and compared using t-tests to assess statistical significance between men and women.

Results

Study participants

This study included 2,320 patients evaluated at the Memory Clinic of Ace Alzheimer Center Barcelona: 249 patients with SCD (CDR = 0), 1,209 patients with MCI (CDR = 0.5) (Petersen, 2004), and 862 patients with dementia (646 with mild dementia, CDR = 1.0; and 216 with moderate dementia, CDR = 2.0). Within the MCI group, 744 were aMCI, while the remaining 465 were naMCI. Among the dementia group, 682 were diagnosed with probable/possible ADD, 117 with vascular dementia, 41 with Lewy body dementia, and 22 with Frontotemporal dementia behavioral variant. A clear age and MMSE gradient were observed within groups, with SCD patients being the youngest (66.4 ± 10.2 years; MMSE 29.3 ± 0.9) and ADD patients the oldest (80.8 ± 6.4 years; MMSE 21.7 ± 3.4), concomitant with a decline in years of formal education (from 12.4 ± 4.1 in SCD to 7.4 ± 4.5 in ADD dementia).

The clinical and sociodemographic characteristics of participants are summarized in [Table 1](#).

Significant differences across diagnostic groups were observed for both age and MMSE (Kruskal-Wallis test, $p < 0.001$ for both variables). *Post hoc* pairwise comparisons with Bonferroni correction revealed significant differences between all groups for MMSE, while age differences were significant between all groups except between ADD and non-AD dementia ([Supplementary Appendix 3](#)).

A subsample of 236 patients with MCI underwent LP for the assessment of AD-core biomarkers in CSF. Their demographic and clinical characteristics are summarized in [Table 2](#) stratified by ATN status. 52.5% of cases were A+, 55.9% were T+, and 46.2% were N+.

Discrimination between the diagnostic groups

Among the multidomain speech models, the highest sensitivity and specificity were achieved in discriminating SCD-ADD, followed by SCD-CI and SCD-MCI. In contrast, the naMCI-aMCI and MCI-ADD comparisons presented the greatest challenges for diagnostic classification using speech data ([Figure 2](#)).

For the SCD-ADD and MCI-ADD discriminations, the baseline model trained on the MMSE+Demo feature set achieved a higher AUC than the multidomain speech models. In contrast, for the SCD-CI, SCD-MCI, and naMCI-aMCI classifications, the multidomain speech models yielded superior AUC. For SCD-ADD discrimination, the baseline model achieved nearly perfect prediction, with an accuracy of 0.99 with an AUC of 0.99. Similarly, speech-based models demonstrated robust performance, with a 0.94 AUC ([Supplementary Appendix 4](#)).

When relying solely on multidomain speech features, combining features from all three tasks consistently outperformed the use of individual task features. Additionally, when evaluating individual tasks, models utilizing SVF-generated features consistently achieved higher BA than those based on IMG and PPE features, with PPE performing worse than both SVF and IMG ([Supplementary Appendix 4](#)).

Cognitive domains prediction

The baseline model trained on the MMSE+Demo feature set outperformed the multidomain speech models in predicting memory, attention, and orientation cognitive composites. Conversely, the multidomain speech models excelled in predicting the executive, visuospatial, and language cognitive composites. Correlation between the predicted values and cognitive composites further revealed that the best results were achieved for the executive cognitive composite, followed by memory, visuospatial, and language composites ([Figure 3](#)).

Models that were trained using a combination of multidomain speech features from all three tasks exhibited higher correlations and greater coefficient of determination (R^2) than models trained on individual tasks. Within the individual tasks, SVF-generated features generally outperformed those derived from IMG and PPE, except in predicting the visuospatial neuropsychological cognitive composite, where IMG task features yielded better results. PPE

TABLE 2 Mean comparison of clinical and sociodemographic features stratified by ATN status.

Variable	A-	A+	T-	T+	N-	N+
Sample size (n, %)	112 (47.5)	124 (52.5)	104 (44.1)	132 (55.9)	127 (53.8)	109 (46.2)
APOE ε4 carriers (%)	13.8	50.0	18.3	43.6	22.1	45.0
Age (mean, SD)	69.7 (9.0)	76.2 (5.1)	69.4 (8.9)	76.0 (5.4)	70.9 (8.7)	75.7 (5.9)
Sex (% female)	51.7	63.7	51.9	62.8	50.3	66.9
Years of formal education (mean, SD)	8.7 (4.2)	8.3 (4.8)	9.0 (4.5)	8.1 (4.5)	8.8 (4.6)	8.2 (4.4)
MMSE (mean, SD)	27.3 (2.6)	25.6 (2.8)	27.2 (2.7)	25.8 (2.8)	27.1 (2.7)	25.7 (2.8)

A, Aβ1-42/Aβ1-40 ratio; T, P-tau181; N, T-tau; MMSE, Mini-Mental State Examination; SD, standard deviation.

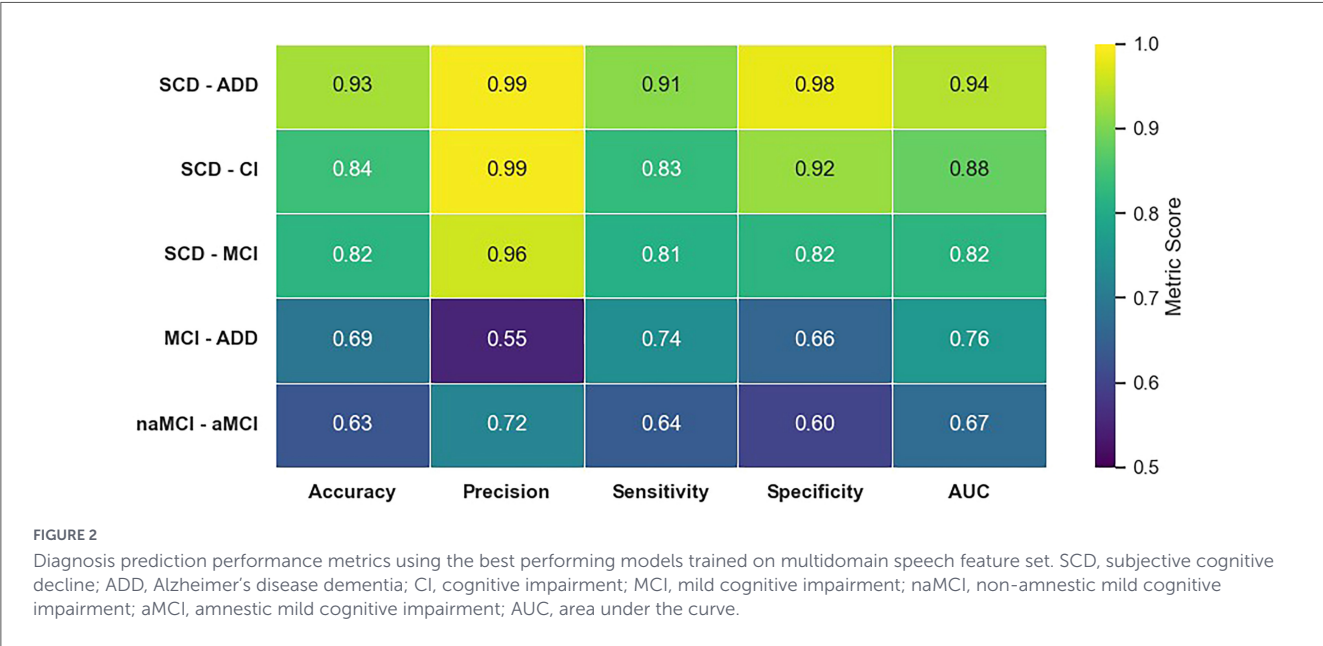


FIGURE 2
Diagnosis prediction performance metrics using the best performing models trained on multidomain speech feature set. SCD, subjective cognitive decline; ADD, Alzheimer’s disease dementia; CI, cognitive impairment; MCI, mild cognitive impairment; naMCI, non-amnesic mild cognitive impairment; aMCI, amnesic mild cognitive impairment; AUC, area under the curve.

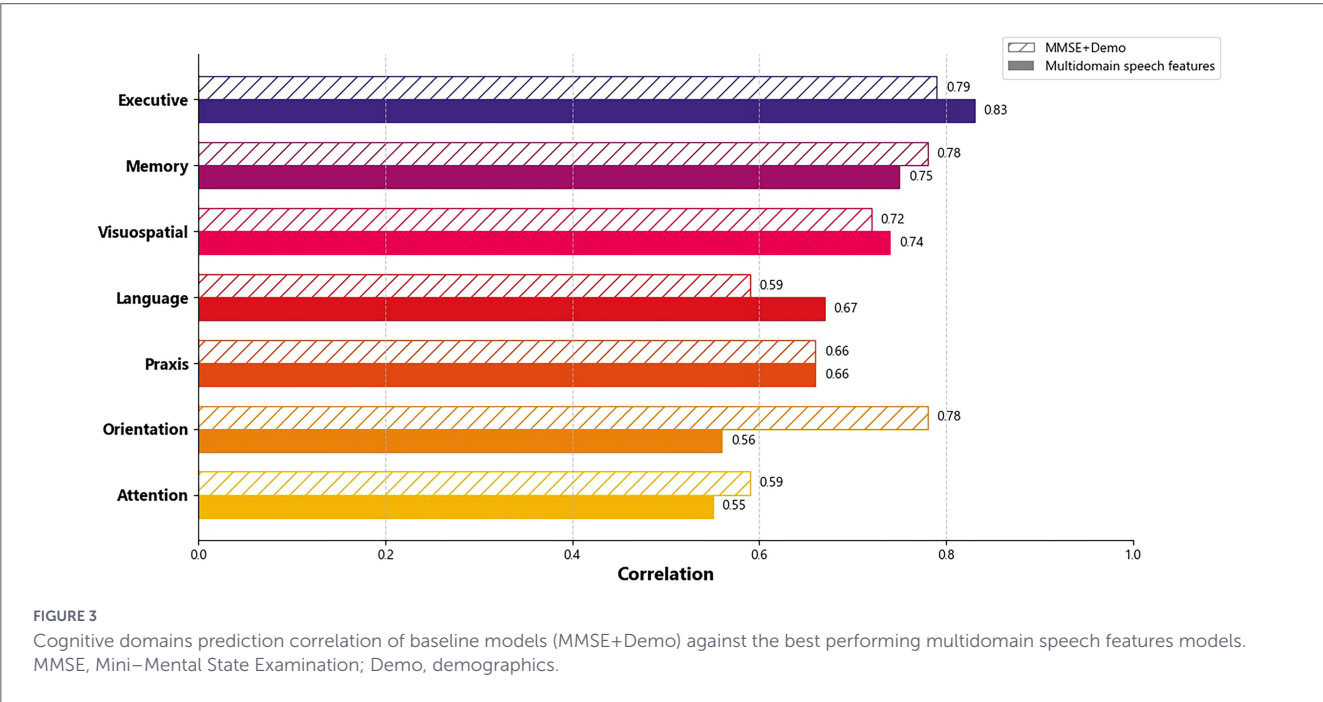


FIGURE 3
Cognitive domains prediction correlation of baseline models (MMSE+Demo) against the best performing multidomain speech features models. MMSE, Mini-Mental State Examination; Demo, demographics.

TABLE 3 Discrimination of ATN status.

ATN status	Feature set	Model	Acc	Pre	Sen	Spe	AUC
A+/A-	MMSE+Demo	LR	0.77	0.81	0.81	0.7	0.76
	Multidomain speech (SVF)	VLPSO+XGB	0.69	0.72	0.81	0.5	0.71
T+/T-	MMSE+Demo	DT	0.79	0.79	0.91	0.56	0.73
	Multidomain speech (IMG)	VLPSO+KNN	0.77	0.75	0.97	0.39	0.74
N+/N-	MMSE+Demo	DT	0.65	0.63	0.86	0.42	0.64
	Multidomain speech (IMG)	VLPSO+KNN	0.69	0.8	0.57	0.83	0.73

A, Aβ1-42/Aβ1-40 ratio; T, P-tau181; N, T-tau; Acc, accuracy; Pre, precision; Sen, sensitivity; Spe, specificity; AUC, area under the curve; MMSE, Mini-Mental State Examination; Demo, demographics; LR, logistic regression; DT, decision tree; VLPSO, variable-length particle swarm optimization; XGB, eXtreme gradient boosting; KNN, K-nearest neighbors; SVF, semantic verbal fluency test; IMG, image description task.

consistently produced the worst performance among the three tasks (Supplementary Appendix 4).

ATN detection

The detection of CSF biomarker positivity yielded AUC values ranging from 0.71 to 0.74 when using the multidomain speech feature set. The MMSE+Demo feature set only outperformed the speech models in predicting A+ cases (Table 3).

The most effective models employed VLPSO feature selection rather than NSGA-II. In contrast to the findings on previous sections, for the prediction of ATN status, models using multidomain speech features from a single task achieved higher BAs compared to those combining features across tasks. Specifically, the model trained on SVF features was the best performing for the A+/A- discrimination, while the IMG feature set yielded the best results for TN detection (Table 3).

Machine learning explainability results

Features from multiple domains generally exhibited balanced importance across most diagnostic classification tasks, with the notable exception of MCI-ADD discrimination. In this case, task-specific features—particularly the number of unique animals named in the SVF task—accounted for the majority of predictive importance. For cognitive domain prediction, task-specific features remained the most salient features within the speech domain. Notably, the number of unique animals in the SVF task emerged as the most influential feature across all neuropsychological composite predictions. In contrast, for ATN status classification, acoustic features played a more prominent role in determining the predicted label. Nevertheless, linguistic features still contributed the largest share of feature importance (Figure 4). Notably, for ATN detection, no individual feature stands out.

Most models were not prone to bias related to sex, age, or years of formal education. A sex-based analysis revealed no significant tendency toward sex-related misclassification, with the exception of the MCI-ADD model, which was prone to sex-based error, showing a higher proportion of women in misclassified samples (75.4%) compared to the original dataset (64.9%). Age-related bias was observed in three models: SCD- CI and SCD-MCI were prone

to age-related errors with misclassified samples having a lower mean age than the original dataset, while T-tau model showed the opposite trend. Only the SCD-CI model was prone to bias based on years of formal education.

When comparing SHAP features aggregated by domain, none of the *t*-tests showed a statistically significant difference between sexes (all *p* > 0.05).

Discussion

This study investigated the potential of multidomain speech features for the detection of cognitive impairment and AD. The findings demonstrate that speech-based models, leveraging acoustic, lexical, syntactic, and semantic features, can effectively differentiate between cognitive stages and predict the presence of CSF biomarkers associated with AD pathology.

The applied ML models showed a good discriminatory capacity of speech parameters for distinguishing among the diagnostic groups (Figure 2) and predicting ATN status in CSF (Table 3). Also, ML models were able to estimate cognitive performance using speech features (Figure 3). Furthermore, by incorporating XAI techniques, we gained valuable insights into how different features influenced the decisions made by the models (Figure 4).

Our study confirms prior evidence suggesting that speech impairment manifests early in AD progression and can be systematically captured through automated speech analysis. In our previous works (García-Gutiérrez et al., 2023; García-Gutiérrez et al., 2024), we relied solely on a restricted set of paralinguistic features. Previous research has shown that linguistic deficits, particularly in lexical retrieval and fluency, correlate with early cognitive decline, reinforcing the utility of linguistic biomarkers in cognitive impairment screening (Gomez and White, 2006; Pekkala et al., 2013). As demonstrated by other researchers, using a more diverse feature set derived from speech analysis could substantially enhance the performance of predictive models (Amini et al., 2023; Mahajan and Baths, 2021; Xue et al., 2021).

The multidomain speech model used in our work achieved competitive performance in distinguishing diagnosis stages of cognitive impairment. On the identification of the ADD diagnosis (SCD-ADD and MCI-ADD), the baseline model using MMSE and demographics outperformed the speech model. Nevertheless,

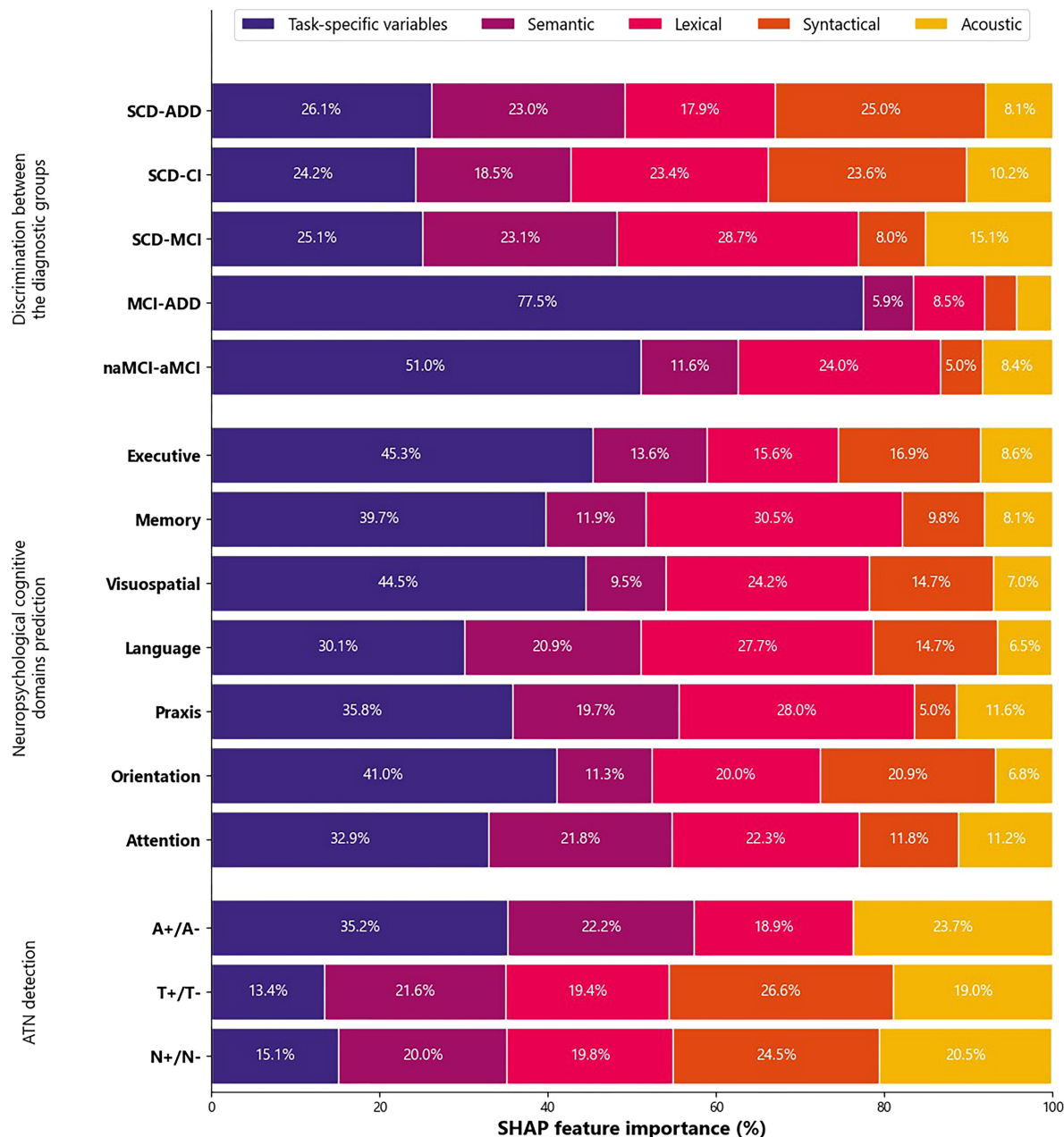


FIGURE 4

SHAP feature importance by feature domain for the different tasks. SCD, subjective cognitive decline; ADD, Alzheimer's disease dementia; CI, cognitive impairment; MCI, mild cognitive impairment; naMCI, non-amnesic mild cognitive impairment; aMCI, amnesic mild cognitive impairment; A, A β 1-42/A β 1-40 ratio; T, P-tau181; N, T-tau.

on all the other discrimination pairs, the speech model was superior to the proposed baseline model. The classification metrics obtained were consistent with previous studies (Amini et al., 2023; Asgari et al., 2017; Chen et al., 2021; Gosztolya et al., 2021; He et al., 2023; Lindsay et al., 2021; Mahajan and Baths, 2021; Themistocleous et al., 2018; Xue et al., 2021). Therefore, our findings provide new evidence supporting the potential of automated speech analysis for the identification of different stages of AD. Also, it was shown that the combination of speech features from multiple speech tasks improves prediction performance.

Cognitive tests from a standardized battery of neuropsychological measures were grouped into seven composite scores representing the cognitive domains of attention, executive functions, language, memory, visuospatial functions, orientation, and praxis (Alegret et al., 2012). Using these cognitive composites as the target variables of the ML models, the correlations obtained with speech parameters were higher than in previous works in all cognitive domains (García-Gutiérrez et al., 2024). Notably, executive function and memory demonstrated particularly strong associations with speech parameters. In contrast, the attention domain (though affected in AD) showed weaker correlations with speech

biomarkers, consistent with prior findings (García-Gutiérrez et al., 2024).

During the preclinical stage of AD, before symptom onset, the pathophysiological course of the disease is characterized by first the formation of A plaques and later p-tau protein aggregates, which accumulate in the brain and disrupt normal neuronal function (Sperling et al., 2013). Subsequently, at the MCI stage, the accumulation of these proteins in the brain reaches a critical threshold, leading to neuronal injury and pathological changes in the volumes of different brain regions (Sperling et al., 2013). The main idea here is that the analysis of speech and language could provide relevant information about the underlying pathophysiological process of AD (Voleti et al., 2020). Our results align with studies using AD-related biomarkers conducted by Hajjar et al. (2023) in a longitudinal cohort of cognitively unimpaired individuals and MCI patients, García-Gutiérrez et al. (2023) in a cross-sectional study of MCI subjects, Mueller et al. (2021) in a longitudinal study involving healthy individuals and early-stage MCI patients, and Verfaillie et al. (2019) in a cross-sectional study of individuals with cognitive decline. These studies demonstrated that amyloid burden is associated with several speech parameters. Notably, the studies of Mueller et al. (2021) and Verfaillie et al. (2019) were based on the lexico-syntactic content of the speech, while García-Gutiérrez et al. (2023) focuses on the properties of the sound generated when describing a picture. We extended the previous studies to the detection of ATN status providing a new landmark for the use of speech biomarkers in neurodegenerative disorders (Table 3).

Feature importance analysis in our study illustrates that all the speech domains played a dominant role in diagnostic classification. The number of unique animals listed in the SVF task emerged as the strongest predictor across some tasks, underscoring the role of semantic richness. Importantly, our explainability analyses revealed that the models exhibited minimal dependency on sociodemographic variables such as age, sex, or years of formal education. With the exception of a few tasks, no statistically significant biases were observed, and the SHAP-based feature domain importance did not differ between sexes. This suggests that our speech-based models possess strong potential for generalizability across diverse populations. The low reliance on demographic variables also supports the notion of democratization in cognitive assessment—highlighting that such models could be deployed in broader, more heterogeneous settings without exacerbating existing health disparities.

A key advantage of our approach is its integration of multiple linguistic and acoustic dimensions, surpassing prior studies focused on isolated speech domains. While some existing research has explored the diagnostic potential of single-domain speech features (Ambadi et al., 2021; García-Gutiérrez et al., 2024; Kurdi, 2023), our findings suggest that a multidimensional approach enhances predictive accuracy and robustness. This comprehensive framework allows for a more holistic assessment of cognitive function and neurodegeneration. A key strength of this approach is the systematic assessment of model biases, a critical requirement for ensuring the reliability and fairness of ML applications in healthcare.

Despite these promising results, several limitations must be acknowledged. Although our cohort is relatively large and diverse, external validation using independent datasets is necessary to confirm the generalizability of our findings. Additionally, the study is based on cross-sectional observations, which may not fully capture the dynamic and multifaceted relationship between dementia and multidomain speech features. Longitudinal studies are essential to understand how speech patterns evolve over time in parallel with amyloid and tau pathology (Amini et al., 2024). Another limitation is the exclusive focus on Spanish-speaking participants; therefore, future research should explore the applicability of these models across different languages and cultural contexts (Lindsay et al., 2021). This would help ensure that speech-based diagnostic tools are robust, generalizable, and inclusive. While our models achieved good discrimination for many tasks, there is still room for improved precision in distinguishing MCI subtypes and predicting biomarker positivity. Further refinement of feature selection methods and model architectures could enhance predictive performance (Gosztolya et al., 2021).

Building on the current framework, future studies could benefit from incorporating neuroimaging data into the modeling process. Early alterations in brain metabolism and functional connectivity—particularly in language-related areas, have been linked to subtle speech changes in preclinical AD, suggesting that multimodal approaches may offer improved diagnostic precision (Scheltens et al., 2021). Moving beyond a cross-sectional design is essential, and longitudinal analyses of speech-based biomarkers could provide a clearer view of how language features evolve over time, support tracking of disease progression across the AD spectrum, and help distinguish AD pathology from other types of dementia (Amini et al., 2024). Another crucial aspect is the potential of remote and passive speech-based screening and monitoring for neurodegenerative diseases (Popp et al., 2024). Given the feasibility of deploying these models in telemedicine settings, speech biomarkers could bridge accessibility gaps in early AD detection.

Given the advancements in AI and ML, speech biomarkers could become a key tool in early intervention strategies for AD. Future work should explore the integration of AI-driven speech analysis with other digital health solutions to develop a standardized framework for cognitive screening and disease monitoring. The potential to deploy such methods in telehealth applications could revolutionize early detection and patient management strategies, paving the way for more personalized and timely interventions in neurodegenerative disorders.

Conclusion

In conclusion, this study highlights the significant potential of multidomain speech analysis as a non-invasive and accessible method for detecting and monitoring AD and cognitive impairment. By integrating acoustic, lexical, syntactic, and semantic features, our approach offers a robust predictive framework. The strong correlation between speech-derived

biomarkers and neurophysiological functions underscores the viability of speech analysis as a scalable screening tool.

Furthermore, the effectiveness of this approach in detecting cognitive changes across the AD continuum, as well as its ability to identify early signs of amyloidosis and tauopathy, supports its potential for early diagnosis and intervention. While our findings are promising, further research is needed to validate these results in diverse population cohorts and to refine predictive models for greater accuracy. Future studies should also explore the integration of speech-based assessments into telemedicine platforms to enhance accessibility and incorporate AI-driven solutions for automated cognitive screening.

The convergence of ML, natural language processing, and digital health presents a transformative opportunity to improve early detection and monitoring of AD. By addressing current challenges and optimizing speech-based methodologies, this research contributes to the development of innovative and widely applicable diagnostic solutions for neurodegenerative diseases, ultimately paving the way for more personalized interventions and improved patient outcomes.

Data availability statement

The datasets generated and analyzed during the current study are not publicly available as they contain human privacy-sensitive data. The code used in this study will also not be shared. However, the datasets are available from the corresponding author upon reasonable request.

Ethics statement

This study was approved by the ethics committees of the Hospital Universitari de Bellvitge (Barcelona) (ref. PR007/22) in compliance with Spanish biomedical laws (Law 14/2007, July 3, regarding biomedical research; Royal Decree 1716/2011, November 18) on 10 March 2022 and adhered to the principles of the Declaration of Helsinki. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JB-F: Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. BC: Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. AH-C: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. FG-G: Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. MA: Investigation, Writing – review & editing. NM: Investigation, Writing – review & editing. AC:

Investigation, Writing – review & editing. MF: Investigation, Writing – review & editing. AM: Investigation, Writing – review & editing. AS: Investigation, Writing – review & editing. ID: Investigation, Writing – review & editing. AV-S: Investigation, Writing – review & editing. PG-G: Investigation, Writing – review & editing. RP: Investigation, Writing – review & editing. CO: Investigation, Writing – review & editing. MC-B: Investigation, Writing – review & editing. ÁM-M: Investigation, Writing – review & editing. PB-B: Investigation, Writing – review & editing. LM: Investigation, Writing – review & editing. AO: Investigation, Writing – review & editing. GO: Investigation, Writing – review & editing. PS-C: Investigation, Writing – review & editing. MR-R: Investigation, Writing – review & editing. YC-F: Writing – review & editing. MG: Conceptualization, Writing – review & editing. LT: Writing – review & editing. CB: Writing – review & editing. XM: Writing – review & editing. MB: Conceptualization, Writing – review & editing. AR: Conceptualization, Project administration, Writing – review & editing. MM: Conceptualization, Investigation, Project administration, Writing – review & editing. SV: Conceptualization, Project administration, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This project has received funding from R&D Missions in the Artificial Intelligence program, which is part of the Spain Digital 2025 Agenda and the National Artificial Intelligence Strategy and financed by the European Union through Next Generation EU funds (project TARTAGLIA, exp.MIA.2021.M02.0005). We appreciate the financial support of the GMV company, and its continued commitment to research and development. This project has also received funding from the Instituto de Salud Carlos III (ISCIII) Acción Estratégica en Salud, integrated in the Spanish National RDCI Plan and financed by ISCIII Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER-Una manera de hacer Europa) grant PI19/00335 awarded to MM, grant PI17/01474 awarded to MB, grants AC17/00100, PI19/01301 and PI22/01403 awarded to AR and by the European Union Joint Programme-Neurodegenerative Disease Research (JPND) Multinational research projects on Personalized Medicine for Neurodegenerative Diseases/Instituto de Salud Carlos III grant AC19/00097 awarded to AR and grant FI20/00215 from the Instituto de Salud Carlos III (ISCIII) awarded to ID. MM received funding support from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement N° 796706. AR and MB received support from the European Union/EFPIA Innovative Medicines Initiative Joint undertaking ADAPTED and MOPEAD projects (grant nos. 115975 and 115985, respectively). AC received support from the Instituto de Salud Carlos III (ISCIII) under the grant Sara Borrell (CD22/00125) and the Spanish Ministry of Science and Innovation, Proyectos de Generación de Conocimiento grant PID2021-122473OA-I00.

Acknowledgments

The authors are grateful to all patients who consented to participate for their significant contribution to making this project possible. The authors are grateful to AcceXible who developed the digital platform for speech protocol administration.

Conflict of interest

AR is member of the scientific advisory board of Landsteiner Genmed and Grifols SA. He has stocks of Landsteiner Genmed. MB has consulted for Araclon, Avid, Grifols, Lilly, Nutricia, Roche, Eisai and Servier. She received fees from lectures and funds for research from Araclon, Biogen, Grifols, Nutricia, Roche and Servier. She reports grants/research funding from Abbvie, Araclon, Biogen Research Limited, Bioiberica, Grifols, Lilly, S.A, Merck Sharp & Dohme, Kyowa Hakko Kirin, Laboratorios Servier, Nutricia SRL, Oryzon Genomics, Piramal Imaging Limited, Roche Pharma SA, and Schwabe Farma Iberica SLU, all outside the submitted work. She has not received personal compensations from these organizations. MM has consulted for F. Hoffmann-La Roche Ltd and is a member of the Scientific Advisory Board of Biomarkers of Araclon.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2026.1816747/full#supplementary-material>

References

- Abdelnour, C., Rodríguez-Gómez, O., Alegret, M., Valero, S., Moreno-Grau, S., Sanabria, Á, et al. (2017). Impact of recruitment methods in subjective cognitive decline. *J. Alzheimers Dis.* 57, 625–632. doi: 10.3233/JAD-160915
- Ahmed, S., Haigh, A. M., de Jager, C. A., and Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain* 136(Pt 12), 3727–3737. doi: 10.1093/brain/awt269
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: A Next-generation Hyperparameter Optimization Framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (ACM: New York, NY).
- Alegret, M., Blazquez-Folch, J., Pérez, A., Ortega, G., Espinosa, A., Muñoz, N., et al. (2024). FACEmemory®, an innovative self-administered online memory assessment tool. *J. Clin. Med.* 13:7274. doi: 10.3390/jcm13237274
- Alegret, M., Espinosa, A., Vinyes-Junqué, G., Valero, S., Hernández, I., Tárraga, L., et al. (2012). Normative data of a brief neuropsychological battery for Spanish individuals older than 49. *J. Clin. Exp. Neuropsychol.* 34, 209–219. doi: 10.1080/13803395.2011.630652
- Alonso, M. A., Fernandez, A., and Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behav. Res. Methods* 47, 268–274. doi: 10.3758/s13428-014-0454-2
- Alzheimer's Association (2025). 2025 Alzheimer's disease facts and figures. *Alzheimers Dement.* 21:e70235. doi: 10.1002/alz.70235
- Ambadi, P. S., Basche, K., Kosciak, R. L., Berisha, V., Liss, J. M., and Mueller, K. D. (2021). Spatio-Semantic graphs from picture description: Applications to detection of cognitive impairment. *Front. Neurol.* 12:795374. doi: 10.3389/fneur.2021.795374
- Amini, S., Hao, B., Yang, J., Karjadi, C., Kolachalama, V. B., Au, R., et al. (2024). Prediction of Alzheimer's disease progression within 6 years using speech: A novel approach leveraging language models. *Alzheimers Dement.* 20, 5262–5270. doi: 10.1002/alz.13886
- Amini, S., Hao, B., Zhang, L., Song, M., Gupta, A., Karjadi, C., et al. (2023). Automated detection of mild cognitive impairment and dementia from voice recordings: A natural language processing approach. *Alzheimers Dement.* 19, 946–955. doi: 10.1002/alz.12721
- Amjad, H., Roth, D. L., Sheehan, O. C., Lyketsos, C. G., Wolff, J. L., and Samus, Q. M. (2018). Underdiagnosis of dementia: An observational study of patterns in diagnosis and awareness in US older adults. *J. Gen. Intern. Med.* 33, 1131–1138. doi: 10.1007/s11606-018-4377-y
- Asgari, M., Kaye, J., and Dodge, H. (2017). Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimers Dement.* 3, 219–228. doi: 10.1016/j.trci.2017.01.006
- Bergstra, J., Bardenet, R., Kégl, B., and Bengio, Y. (2011). "Algorithms for Hyper-Parameter Optimization," in *Advances in Neural Information Processing Systems*, eds J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger (Red Hook, NY: Curran Associates, Inc).
- Blank, J., and Deb, K. (2020). Pymoo: Multi-objective optimization in python. *IEEE Access* 8, 89497–89509. doi: 10.1109/ACCESS.2020.2990567
- Blesa, R., Pujol, M., Aguilar, M., Santacruz, P., Bertran-Serra, I., Hernández, G., et al. (2001). Clinical validity of the 'mini-mental state' for Spanish speaking communities. *Neuropsychologia* 39, 1150–1157. doi: 10.1016/s0028-3932(01)00055-0
- Blessed, G., Tomlinson, B. E., and Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br. J. Psychiatry* 114, 797–811. doi: 10.1192/bjp.114.512.797
- Boada, M., Tárraga, L., Hernández, I., Valero, S., Alegret, M., Ruiz, A., et al. (2014). Design of a comprehensive Alzheimer's disease clinic and research center in Spain to meet critical patient and family needs. *Alzheimers Dement.* 10, 409–415. doi: 10.1016/j.jalz.2013.03.006
- Boada, M., Tárraga, L., Modinos, G., López, O. L., and Cummings, J. L. (2006). [Neuropsychiatric Inventory-Nursing Home version (NPI-NH): Spanish validation]. *Neurología* 20, 665–673.
- Bredin, H., Yin, R., Coria, J., Gelly, G., Korshunov, P., Lavechin, M., et al. (2020). "Pyannote.Audio: Neural Building Blocks for Speaker Diarization," in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Piscataway, NJ: IEEE), 7124–7128.

- Chen, J., Ye, J., Tang, F., and Zhou, J. (2021). Automatic detection of Alzheimer's disease using spontaneous speech only. *Interspeech* 2021, 3830–3834. doi: 10.21437/interspeech.2021-2002
- Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System," in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, NY).
- Chou, C. J., Chang, C. T., Chang, Y. N., Lee, C. Y., Chuang, Y. F., Chiu, Y. L., et al. (2024). Screening for early Alzheimer's disease: Enhancing diagnosis with linguistic features and biomarkers. *Front. Aging Neurosci.* 16:1451326. doi: 10.3389/fnagi.2024.1451326
- Covington, M. A., and McFall, J. D. (2010). Cutting the gordian knot: The moving-average type-token ratio (MATTR). *J. Quant. Ling.* 17, 94–100. doi: 10.1080/09296171003643098
- Cummings, L. (2019). Describing the cookie theft picture: Sources of breakdown in Alzheimer's dementia. *Prag. Soc.* 10, 151–174. doi: 10.1075/ps.17011.cum
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evol. Comp. IEEE Trans.* 6, 182–197. doi: 10.1109/4235.996017
- del Ser, T., Sánchez-Sánchez, F., García de Yébenes, M. J., Otero, A., and Munoz, D. G. (2006). Validation of the seven-minute screen neurocognitive battery for the diagnosis of dementia in a Spanish population-based sample. *Dement Geriatr. Cogn. Disord.* 22, 454–464. doi: 10.1159/000095858
- Ding, Z., Lee, T. L., and Chan, A. S. (2022). Digital cognitive biomarker for mild cognitive impairments and dementia: A systematic review. *J. Clin. Med.* 11:4191. doi: 10.3390/jcm11144191
- Espinosa, A., Alegret, M., Valero, S., Vinyes-Junqué, G., Hernández, I., Mauleón, A., et al. (2013). A longitudinal follow-up of 550 mild cognitive impairment patients: Evidence for large conversion to dementia rates and detection of major risk factors involved. *J. Alzheimers Dis.* 34, 769–780. doi: 10.3233/JAD-12-2002
- Eyben, F., Scherer, K., Schuller, B., Sundberg, J., Andre, E., Busso, C., et al. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Trans. Affect. Comp.* 7, 190–202. doi: 10.1109/TAFFC.2015.2457417
- Eyben, F., Wöllmer, M., and Schuller, B. (2010). "openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor," in *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference*, (New York, NY: ACM), 1459–1462.
- Folstein, M. F., Robins, L. N., and Helzer, J. E. (1983). The mini-mental state examination. *Arch. Gen. Psychiatry* 40:812. doi: 10.1001/archpsyc.1983.01790060110016
- Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *J. Alzheimers Dis.* 49, 407–422. doi: 10.3233/JAD-150520
- García-Gutiérrez, F., Alegret, M., Marquie, M., Muñoz, N., Ortega, G., Cano, A., et al. (2024). Unveiling the sound of the cognitive status: Machine Learning-based speech analysis in the Alzheimer's disease spectrum. *Alzheimers Res. Ther.* 16:26. doi: 10.1186/s13195-024-01394-y
- García-Gutiérrez, F., Delgado-Alvarez, A., Delgado-Alonso, C., Díaz-Álvarez, J., Pytel, V., Valles-Salgado, M., et al. (2022). Diagnosis of Alzheimer's disease and behavioural variant frontotemporal dementia with machine learning-aided neuropsychological assessment using feature engineering and genetic algorithms. *Int. J. Geriatr. Psychiatry* 37. doi: 10.1002/gps.5667
- García-Gutiérrez, F., Marquie, M., Muñoz, N., Alegret, M., Cano, A., de Rojas, I., et al. (2023). Harnessing acoustic speech parameters to decipher amyloid status in individuals with mild cognitive impairment. *Front. Neurosci.* 17:1221401. doi: 10.3389/fnins.2023.1221401
- Garrard, P., Maloney, L. M., Hodges, J. R., and Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain* 128(Pt 2), 250–260. doi: 10.1093/brain/awh341
- Gomez, R. G., and White, D. A. (2006). Using verbal fluency to detect very mild dementia of the Alzheimer type. *Arch. Clin. Neuropsychol.* 21, 771–775. doi: 10.1016/j.acn.2006.06.012
- Gosztolya, G., Balogh, R., Imre, N., Egas-López, J. V., Hoffmann, I., Vincze, V., et al. (2021). Cross-lingual detection of mild cognitive impairment based on temporal parameters of spontaneous speech. *Comp. Speech Lang.* 69:101215. doi: 10.1016/j.csl.2021.101215
- Gregorutti, B., Michel, B., and Saint-Pierre, P. (2015). Grouped variable importance with random forests and application to multiple functional data analysis. *Comput. Stat. Data Anal.* 90, 15–35. doi: 10.1016/j.csda.2015.04.002
- Hachinski, V. C., Lassen, N. A., and Marshall, J. (1974). Multi-infarct dementia. A cause of mental deterioration in the elderly. *Lancet* 2, 207–210. doi: 10.1016/s0140-6736(74)91496-2
- Hajjar, I., Okafor, M., Choi, J. D., Moore, E., Abrol, A., Calhoun, V. D., et al. (2023). Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. *Alzheimers Dement.* 15:e12393. doi: 10.1002/dad2.12393
- He, R., Chapin, K., Al-Tamimi, J., Bel, N., Marquie, M., Rosende-Roca, M., et al. (2023). Automated classification of cognitive decline and probable Alzheimer's dementia across multiple speech and language domains. *Am. J. Speech Lang. Pathol.* 32, 2075–2086. doi: 10.1044/2023_AJSLP-22-00403
- Jessen, F., Amariglio, R. E., van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., et al. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* 10, 844–852. doi: 10.1016/j.jalz.2014.01.001
- Kurdi, M. Z. (2023). Automatic identification of Alzheimer's disease using lexical features extracted from language samples. *arXiv [Preprint]*. doi: 10.48550/arXiv.2307.08070
- Leitão, M. J., Silva-Spínola, A., Santana, I., Olmedo, V., Nadal, A., Le Bastard, N., et al. (2019). Clinical validation of the Lumipulse G cerebrospinal fluid assays for routine diagnosis of Alzheimer's disease. *Alzheimers Res. Ther.* 11:91. doi: 10.1186/s13195-019-0550-8
- Lindsay, H., Tröger, J., and König, A. (2021). Language impairment in Alzheimer's disease: robust and explainable evidence for AD-Related deterioration of spontaneous speech through multilingual machine learning. *Front. Aging Neurosci.* 13:642033. doi: 10.3389/fnagi.2021.642033
- Liu, S., Cao, Y., Liu, J., Ding, X., and Coyle, D. (2023). A novelty detection approach to effectively predict conversion from mild cognitive impairment to Alzheimer's disease. *Int. J. Mach. Learn. Cyber.* 14, 213–228. doi: 10.1007/s13042-022-01570-2
- Liu, Y., Jun, H., Becker, A., Wallick, C., and Matke, S. (2024). Detection rates of mild cognitive impairment in primary care for the united states medicare population. *J. Prev. Alzheimers Dis.* 11, 7–12. doi: 10.14283/jpad.2023.131
- Lundberg, S., and Lee, S. I. (2017). "A Unified Approach to Interpreting Model Predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (Long Beach, CA), 4766–4777.
- Mahajan, P., and Baths, V. (2021). Acoustic and language based deep learning approaches for Alzheimer's dementia detection from spontaneous speech. *Front. Aging Neurosci.* 13:623607. doi: 10.3389/fnagi.2021.623607
- Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., and Meilán, J. J. G. (2021). Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: A systematic review article. *Front. Psychol.* 12:620251. doi: 10.3389/fpsyg.2021.620251
- McCarthy, P. M., and Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behav. Res. Methods* 42, 381–392. doi: 10.3758/BRM.42.2.381
- McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., et al. (2015). "librosa: Audio and Music Signal Analysis in Python," in *Proceedings of the 14th Python in Science Conference*, (London: Mendeley), 18–24. doi: 10.1159/000526209
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005
- Meier, I. B., Buegler, M., Harms, R., Seixas, A., Cöltekin, A., and Tarnanas, I. (2021). Using a digital neuro signature to measure longitudinal individual-level change in Alzheimer's disease: The Altoidea large cohort study. *NPJ Digit. Med.* 4:101. doi: 10.1038/s41746-021-00470-z
- Mikolov, T., Chen, K., Corrado, G. S., and Dean, J. (2013). "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of the International Conference on Learning Representations*, (Scottsdale), doi: 10.3390/s19225043
- Molinueto, J. L., Ayton, S., Batrla, R., Bednar, M. M., Bittner, T., Cummings, J., et al. (2018). Current state of Alzheimer's fluid biomarkers. *Acta Neuropathol.* 136, 821–853. doi: 10.1007/s00401-018-1932-x
- Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* 43, 2412–2414. doi: 10.1212/wnl.43.11.2412-a
- Mueller, K. D., Van Hulle, C. A., Kosciak, R. L., Jonaitis, E., Peters, C. C., Betthausen, T. J., et al. (2021). Amyloid beta associations with connected speech in cognitively unimpaired adults. *Alzheimers Dement.* 13:e12203. doi: 10.1002/dad2.12203
- Nivre, J., de Marneffe, M. C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., et al. (2020). "Universal Dependencies v2: An ever-growing multilingual treebank collection," in *Proceedings of the 12th Language Resources and Evaluation Conference*, eds N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, et al. (Marseille: European Language Resources Association), 4034–4043.
- Nyongesa, C. A., Hogarth, M., and Pa, J. (2025). Artificial intelligence-driven natural language processing for identifying linguistic patterns in Alzheimer's disease and mild cognitive impairment: A study of lexical, syntactic, and cohesive features of speech through picture description tasks. *J. Alzheimers Dis.* 106, 120–138. doi: 10.1177/13872877251339756
- Orellana, A., García-González, P., Valero, S., Montreal, L., de Rojas, I., Hernández, I., et al. (2022). Establishing in-house cutoffs of CSF Alzheimer's disease biomarkers for the AT(N) stratification of the Alzheimer center barcelona cohort. *Int. J. Mol. Sci.* 23:6891. doi: 10.3390/ijms23136891
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2012). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

- Pekkala, S., Wiener, D., Himali, J. J., Beiser, A. S., Obler, L. K., Liu, Y., et al. (2013). Lexical retrieval in discourse: An early indicator of Alzheimer's dementia. *Clin. Ling. Phon.* 27, 905–921. doi: 10.3109/02699206.2013.815278
- Peña-Casanova, J., Sánchez-Benavides, G., de Sola, S., Manero-Borrás, R. M., and Casals-Coll, M. (2012). Neuropsychology of Alzheimer's disease. *Arch. Med. Res.* 43, 686–893. doi: 10.1016/j.arcmed.2012.08.015
- Pérez, J., Rajngewerc, M., Giudici, J., Furman, D., Luque, F., Alonso Alemany, L., et al. (2023). pysentimiento: A python toolkit for opinion mining and social NLP tasks. *arXiv [Preprint]*. doi: 10.48550/arXiv.2106.09462
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–194. doi: 10.1111/j.1365-2796.2004.01388.x
- Popp, Z., Low, S., Igwe, A., Rahman, M. S., Kim, M., Khan, R., et al. (2024). Shifting from active to passive monitoring of Alzheimer disease: The state of the research. *J. Am. Heart Assoc.* 13:e031247. doi: 10.1161/JAHA.123.031247
- Power, M. C., Willens, V., Prather, C., Moghtaderi, A., Chen, Y., Gianattasio, K. Z., et al. (2023). Risks and benefits of clinical diagnosis around the time of dementia onset. *Gerontol. Geriatr. Med.* 9:23337214231213185. doi: 10.1177/23337214231213185
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). “Stanza: A Python Natural Language Processing Toolkit for Many Human Languages,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, eds Celikyilmaz and AaWTH (Stroudsburg, PA: Association for Computational Linguistics).
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. *arXiv [Preprint]*. doi: 10.48550/arXiv.2212.04356
- Rehurek, R., and Sojka, P. (2011). *Gensim-python Framework for Vector Space Modelling*. Brno: NLP Centre, Faculty of Informatics, Masaryk University.
- Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., et al. (2021). Alzheimer's disease. *Lancet* 397, 1577–1590. doi: 10.1016/S0140-6736(20)32205-4
- Solomon, P. R., and Pendlebury, W. W. (1998). Recognition of Alzheimer's disease: The 7 minute screen. *Fam. Med.* 30, 265–271.
- Speer, R. (2022). *rspeer/wordfreq: v3.0*.
- Sperling, R. A., Karlawish, J., and Johnson, K. A. (2013). Preclinical Alzheimer disease—the challenges ahead. *Nat. Rev. Neurol.* 9, 54–58. doi: 10.1038/nrneurol.2012.241
- Szatloczki, G., Hoffmann, I., Vincze, V., Kalman, J., and Pakaski, M. (2015). Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Front. Aging Neurosci.* 7:195. doi: 10.3389/fnagi.2015.0195
- Themistocleous, C., Eckerström, M., and Kokkinakis, D. (2018). Identification of mild cognitive impairment from speech in Swedish using deep sequential neural networks. *Front. Neurol.* 9:975. doi: 10.3389/fneur.2018.00975
- Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., et al. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr. Alzheimer Res.* 15, 130–138. doi: 10.2174/1567205014666171121114930
- Tran, B., Xue, B., and Zhang, M. (2019). Variable-Length particle swarm optimization for feature selection on high-dimensional classification. *IEEE Trans. Evol. Comp.* 23, 473–487. doi: 10.1109/TEVC.2018.2869405
- Tröger, J., Baykara, E., Zhao, J., Ter Huurne, D., Possemis, N., Mallick, E., et al. (2022). Validation of the remote automated kie speech biomarker for cognition in mild cognitive impairment: Verification and validation following DiME V3 framework. *Digit. Biomark.* 6, 107–116. doi: 10.1159/000526471
- Vanderstichele, H., Bibl, M., Engelborghs, S., Le Bastard, N., Lewczuk, P., Molinuevo, J. L., et al. (2012). Standardization of preanalytical aspects of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: A consensus paper from the Alzheimer's Biomarkers Standardization Initiative. *Alzheimers Dement.* 8, 65–73. doi: 10.1016/j.jalz.2011.07.004
- Verfaillie, S. C. J., Witteman, J., Slot, R. E. R., Pruis, I. J., Vermaat, L. E. W., Prins, N. D., et al. (2019). High amyloid burden is associated with fewer specific words during spontaneous speech in individuals with subjective cognitive decline. *Neuropsychologia* 131, 184–192. doi: 10.1016/j.neuropsychologia.2019.05.006
- Voleti, R., Liss, J. M., and Berisha, V. (2020). A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE J. Sel. Top. Signal Process* 14, 282–298. doi: 10.1109/jstsp.2019.2952087
- World Health Organization [WHO] (2024). *Dementia: Key facts*. Geneva: WHO.
- Xue, B., Zhang, M., Browne, W., and Yao, X. A. (2016). Survey on evolutionary computation approaches to feature selection. *IEEE Trans. Evol. Comp.* 20, 606–626. doi: 10.1109/TEVC.2015.2504420
- Xue, C., Karjadi, C., Paschalidis, I. C., Au, R., and Kolachalama, V. B. (2021). Detection of dementia on voice recordings using deep learning: A Framingham Heart Study. *Alzheimers Res. Ther.* 13:146. doi: 10.1186/s13195-021-00888-3