On Automatic Detection of Cognitive Decline

Fabio Tamburini
FICLIT-University of Bologna
via Zamboni, 32, Bologna, Italy
fabio.tamburini@unibo.it

Abstract

Cognitive decline refers to the gradual loss of thinking abilities, including memory, attention, reasoning, and problem-solving. It can be a normal part of ageing or a symptom of conditions like dementia or Alzheimer's disease when it significantly interferes with daily life. Early diagnosis is crucial, as timely intervention can slow progression and improve quality of life. Emerging approaches such as Digital Linguistic Biomarkers, subtle changes in speech and language patterns captured through digital tools, offer a promising, non-invasive way to detect early signs of cognitive decline before more obvious symptoms appear and perform massive population screening. In this position paper, we contend that the prevailing paradigm for the automatic detection of cognitive decline, primarily relying on classifiers that analyse subjects' linguistic productions at a single point in time, is not the most effective approach. Instead, we advocate for a paradigm shift toward longitudinal analyses that track linguistic patterns over decades. To support this perspective, we present an experiment in which we compile and analyse a long-term corpus of spontaneous speech productions from well-known individuals, enabling insights into cognitive changes across extended time spans.

Keywords: Cognitive Decline, Digital Linguistic Biomarkers, Longitudinal Analyses

1. Introduction

Cognitive decline, ranging from mild memory impairment to severe loss of independence, represents a growing public health challenge with profound personal, societal, and economic consequences. Alzheimer's disease (AD), the most common cause of dementia, lies at the centre of this crisis. In 2025, approximately 7.2 million Americans aged 65 and older are estimated to be living with Alzheimer's dementia, and projections suggest this number could rise to nearly 13.8 million by 2060 (Alzheimer's Association, 2025). These figures reflect only the clinically visible portion of the problem and underestimate the true scope of Alzheimer's-related cognitive decline.

Biomarker-based studies reveal a much broader spectrum of disease. Around 5 million older adults may already have Alzheimer's-related dementia detectable through objective brain changes, while an additional 5–7 million may have Mild Cognitive Impairment (MCI) attributable to Alzheimer's pathology. Together, this implies that 10–12 million older Americans could be experiencing Alzheimer's-related cognitive decline even before reaching the stage of diagnosable dementia. MCI itself is common, affecting an estimated 12–18% of people aged 60 or older, and carries a substantial risk of progression: roughly 10–15% of individuals with MCI develop dementia each year, with about one-third progressing within five years (Alzheimer's Association, 2025).

Beyond prevalence, the burden of cognitive decline is immense. In 2025, healthcare and long-term care costs associated with Alzheimer's and other dementias are projected to reach $384 billion. When the value of unpaid caregiving, estimated at over $413 billion, is included, the total societal cost becomes staggering. This burden extends well beyond economics, deeply affecting caregivers, families, and communities. These realities highlight the urgency of addressing cognitive decline not only as a medical issue but also as a major social and policy concern.

A critical challenge lies in shifting the focus from treating advanced dementia to identifying individuals at earlier stages, such as MCI or even preclinical Alzheimer's disease. Pathological brain changes related to AD can begin decades before symptoms become apparent, yet current diagnostic approaches rely heavily on clinical presentation and expensive, specialised biomarker tests. As a result, opportunities for early intervention are often missed. This diagnostic gap underscores the need for feasible, scalable, and cost-effective screening methods that can be deployed at the population level. Early identification would allow clinicians to implement lifestyle interventions, monitor disease progression, and potentially apply emerging disease-modifying therapies when they are most effective. Population-wide screening could also reduce disparities by reaching individuals who lack access to specialised neurological care.

Alzheimer's disease is best understood as a continuous process rather than a set of discrete stages. It begins with a preclinical phase characterised by silent biological changes, such as amyloid and protein tau accumulation, without obvious symptoms. During this phase, individuals may experience Subjective Cognitive Decline (SCD), reporting perceived worsening of memory despite normal performance on standard tests. The next stage is MCI,

Proceedings of the 6th RaPID@MENTAI.ai Workshop @ LREC 2026, pages 41–54
12 May 2026.

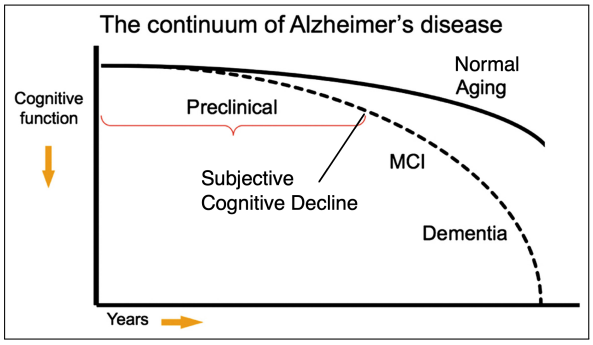

Figure 1: Model of the clinical trajectory of Alzheimer's disease (Picture adapted from Sperling et al. 2011).

often due to Alzheimer's pathology, where mild but noticeable cognitive symptoms emerge without significantly interfering with daily life. As the disease progresses, individuals move through stages of mild, moderate, and severe dementia, with increasing impairment in everyday functioning and eventual loss of independence. Importantly, cognitive decline unfolds gradually over many years, with disease-related trajectories showing steeper declines than normal ageing but remaining continuous in nature, as illustrated by Figure 1.

An important modifier of this trajectory is cognitive reserve (CR), a concept referring to the brain's resilience to age-related and pathological changes. Cognitive reserve helps explain why individuals with similar levels of brain pathology can show very different clinical outcomes. Factors such as education, intellectually demanding occupations, social engagement, and cognitively stimulating activities contribute to higher cognitive reserve, allowing some individuals to compensate more effectively and delay the onset of symptoms. Figure 2 illustrates the point clearly and, comparing this picture with Figure 1, the impact of cognitive reserve on diagnosis and disease progression should be evident.

Although cognitive reserve cannot be measured directly, it is inferred through proxies like educational attainment, occupational history, and lifestyle questionnaires (e.g. "CRIq" from Nucci et al. 2012). Higher cognitive reserve may delay diagnosis, even though underlying disease progression continues.

Within this context, novel screening approaches are gaining attention, particularly Digital Linguistic Biomarkers (DLBs) (Gagliardi et al., 2021). Language is a complex cognitive function supported by widespread brain networks (Catani et al., 2012; Hagoort, 2017; Hertrich et al., 2020), making it sensitive to subtle neural changes. DLBs consist of quantifiable linguistic and speech features, such as lexical diversity, syntactic complexity, fluency, semantic coherence, and acoustic properties, that can be automatically extracted using digital tools. Even mild disruptions in memory or executive function can manifest as detectable changes in everyday speech and writing.

Compared to traditional neuropsychological assessments, DLB-based methods are less resource-intensive and can be administered remotely and repeatedly at scale. Speech samples collected via smartphones or telehealth platforms can be analysed using natural language processing and machine learning techniques, enabling low-cost, ecologically valid monitoring of cognitive function over time. A growing body of evidence (see next section) shows that DLBs can reliably distinguish healthy ageing from MCI and early Alzheimer's disease, often detecting changes before overt behavioural symptoms appear and it has gained traction among researchers and clinicians as a means of obtaining fast, replicable, and objective proxy measures of mental disorders (Gagliardi, 2024).

Overall, the escalating burden of cognitive decline demands a paradigm shift toward early, population-level detection. By integrating scalable screening tools such as digital linguistic biomarkers with existing clinical approaches, healthcare systems may better address the personal, societal, and economic costs of Alzheimer's disease and related dementias.

1.1. State-of-the-art on the Automatic Detection of Cognitive Decline

In recent decades, advanced NLP techniques have been increasingly applied to the analysis of written texts, clinically elicited utterances, and spontaneous speech, with the aim of identifying DLBs of psychiatric and neurological disorders and automatically extracting linguistic features for pathology recognition, classification, and characterisation.

Computational methods have already proven effective in detecting linguistic indicators of cerebral functional disorders, including language alterations and disruptions linked to depression (Jiang et al., 2017; Stasak et al., 2019), focal brain lesions (Fergadiotis and Wright, 2011), Parkinson's disease (Benba et al., 2016; Sztahó and Vicsi, 2016; Arias-Vergara et al., 2018; Upadhya et al., 2019; Wang et al., 2022; Xue et al., 2023; Singh and Tripathi, 2024; Anap et al., 2025) and schizophrenia (Nenchev et al., 2024). They have also been successfully employed to detect prodromal dementia (MCI) (Roark et al., 2007, 2011; Satt et al., 2013; Vincze et al., 2016; dos Santos et al., 2017; Matsuda Toledo et al., 2018; Meilán et al., 2018; Tóth et al., 2018; Wang et al., 2019; Meilán et al., 2020; Wang et al., 2021; Gosztolya et al., 2021; Calzà et al., 2021; Ivanova et al., 2022; Egas-López et al., 2022; Moret-Tatay et al., 2023; Yamada et al., 2023;

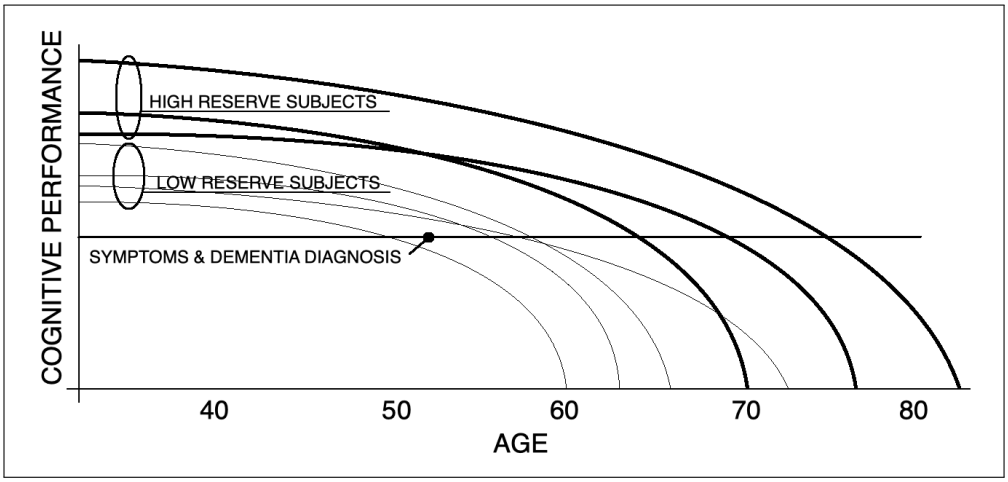

Figure 2: The graphs summarise evidence from observational studies indicating that individuals with higher and lower levels of education tend to differ in cognitive ability in early adulthood and even if, on average, show only small differences in rates of cognitive decline over time, the threshold of symptoms is reached at different ages (Picture adapted from Lövdén et al. 2020).

Favaro et al., 2024; Pourramezan Fard et al., 2024), as well as specific associated pathologies such as Alzheimer's disease (Jarrold et al., 2014; Fraser et al., 2016; Chinaei et al., 2017; López-de-Ipiña et al., 2015; Yancheva and Rudzicz, 2016; Sirts et al., 2017; Eyigoz et al., 2020; Yang et al., 2024; Huang et al., 2024; Kumar et al., 2025; Li et al., 2025), Primary Progressive Aphasia (PPA) (Fraser et al., 2014), and Frontotemporal Dementia (Jarrold et al., 2014; Coppieters et al., 2024).

Recent reviews (de la Fuente Garcia et al., 2020; Pulido et al., 2020; Petti et al., 2020; Ding et al., 2024; Cornacchia et al., 2025; Shankar et al., 2025; Shakeri and Farmanbar, 2025) outlines that, whereas neuropsychological tests and structured assessments often affect the naturalness of a subject's responses, the analysis of spontaneous spoken language offers an ecological and cost-efficient way to detect linguistic alterations in potential patients even within primary care settings.

Considering the cited literature, two main aspects emerge in addressing the problem:

- **Features/DLBs identification**: Many studies focus on defining or proposing a set of linguistic features capable of distinguishing potentially pathological subjects from healthy controls. Some studies manually extract features from speech or written samples, while others develop NLP systems to automate this process. Nearly all works employ statistical significance tests to identify the most promising linguistic indicators associated with the pathology.

- **Automatic classification**: Once relevant features are identified, several studies aim to construct automatic systems for pathology detection. Common machine/deep learning methods are employed, achieving varying levels of performance.

1.2. Past Shared Challenges

"ADReSS/ADReSSo" Challenges targets three difficult automatic prediction problems of societal and medical relevance, namely: detection of Alzheimer's Dementia, inference of cognitive testing scores, and prediction of cognitive decline by providing new and richly annotated English datasets to evaluate automatic systems (Luz et al., 2021b,a).

The recent "Prediction and Recognition Of Cognitive declinE through Spontaneous Speech" (PROCESS)[1] Signal Processing Grand Challenge at ICASSP-2025, proposes signal processing and prediction tasks to detect dementia via speech processing (Tao et al., 2025).

Both challenges introduced new datasets in which subjects' speech samples were classified into two or three classes following clinical judgments. Participants should apply their systems for classifying each sample/subject into one of the proposed classes. The best systems participating to these challenges obtained F1 classification results in the range from about 70% to 80%.

In general, the very large set of works published in the last years and devoted to the automatic detection of cognitive decline (see the survey papers cited before) present similar performance results: when dealing with the binary classification of healthy controls w.r.t. AD subjects, performance often are higher than 90% of correct classification, while, on the most challenging and definitely most

[1] https://processchallenge.github.io/

interesting problem of distinguishing MCI subjects from controls, it drops to around 75/80%.

As previously noted, it is crucial to detect the disease at its earliest stages, ideally when individuals present with MCI, or even earlier, when they experience only subjective and temporary memory difficulties. To be truly effective, the technological approaches described above must be capable of supporting large-scale screening across broad segments of the population. Unfortunately, current state-of-the-art systems do not yet provide sufficient reliability in identifying these early stages of dementia. In our view, this limitation is due more to challenges related to cognitive reserve than to the optimal combination of DLBs or classifier choice. Moreover, when analysing speech productions via DLB extraction to characterise individual speech profiles, subjects' speaking styles and specific accents, including those influenced by immigration from other countries, play a substantial role and can significantly blur the analysis of speech/language features.

These factors lead to considerable overlap between the two or three relevant classes in the feature space, resulting in unsatisfactory performance and limiting the applicability of such approaches for large-scale screening.

2. A Different Perspective

Building on the considerations outlined above, we argue that achieving the ultimate goal requires a true paradigm shift, in the Kuhnian sense. The cognitive reserve of an individual is extremely difficult to measure, as it is shaped by the entirety of his/her life experiences. Attempting to aggregate data from different subjects, even when they are classified within the same group, whether pathological (MCI/AD) or non-pathological (HC), creates challenges that machine learning classifiers cannot easily resolve with sufficient accuracy to enable large-scale population screening.

We believe a shift in perspective is needed: rather than designing systems that classify individuals "synchronically", at a single point in time, we should consider a "diachronic approach", examining each subject across the course of ageing. An ideal method would involve recording spontaneous speech samples at regular intervals, for example every two years after the age of 50, calculating the DLBs for each session, and assessing whether the individual shows signs of cognitive decline by comparing his/her current productions with his/her own past recordings.

This line of inquiry is not entirely new, as a limited number of studies have attempted to investigate cognitive decline using longitudinal data. For example, Laguarta and Subirana (2021) acknowledged the importance of longitudinal analyses and proposed a complex set of multimodal biomarkers that could, in principle, support such an approach. However, they did not present a longitudinal experiment due to the lack of suitable datasets. Petti et al. (2023) explored a simpler strategy, employing DLBs derived solely from written language (speech transcriptions), and demonstrated the promise of a longitudinal perspective. Gkoumas et al. (2024) introduced a multimodal longitudinal corpus spanning 12 months and conducted a DLB study on it, though the time span was too limited for significant changes to be observed. Comparable observations apply to the studies by Robin et al. (2023); Luz et al. (2021a), which are highly engaging but limited to a relatively short duration of 18 or 24 months. In another study, Petti and Korhonen (2024) created a novel longitudinal corpus by collecting interviews of famous individuals from YouTube, applying the same type of DLB analysis used in Petti et al. (2023). Despite the simplicity of their approach, the corpus itself is highly relevant to our objectives and will be examined in more detail in the following section. Finally, Chang et al. (2025) conducted an in-depth study on applying DLBs to track longitudinal trends, successfully distinguishing different trajectories between healthy controls and subjects with MCI. The main limitation, however, was the restricted number of longitudinal points, with data collected from only two visits/interviews.

The studies reviewed provide valuable groundwork for our perspective, highlighting the importance of longitudinal analyses in detecting cognitive decline. However, they all lack in adopting a fully subject-centred approach, which lies at the core of our proposal.

3. An Experiment to Support our View

To support our perspective, we designed an experiment that, given the nature of the available data, can only be regarded as a pilot study. The core of our proposal focuses on collecting subject data across the ageing process over an extended period of time, beginning, for instance, from the age of 50 onwards.

Unfortunately, no dataset currently exists that covers such an extended time span available for research purposes. There is, however, a notable exception: Petti and Korhonen (2024) introduced a longitudinal corpus - LoSST-AD - spanning a substantial portion of the subjects' lifespans, which would be ideal for our study. They compiled this resource by downloading interviews and other recordings from YouTube for ten well-known English-speaking individuals who had passed away from Alzheimer's disease, along with a matched set of healthy controls selected to reflect similar socio-

demographic profiles. However, for ethical reasons, the authors chose not to distribute the recordings themselves, making this valuable corpus only partially accessible (they released anonymised transcriptions only).

3.1. The μCLSD Dataset

Given the absence of an appropriate dataset to test our research hypothesis, we were compelled to construct a new linguistic resource, drawing inspiration from the work of Petti and Korhonen (2024). In the absence of large-scale longitudinal projects tracking subjects over the last 20–30 years of their lives, the only feasible approach is to rely on publicly available recordings of well-known individuals. Interviews with actors, writers, politicians, and other public figures represent a valuable source of material, potentially spanning decades and thus enabling extensive longitudinal analyses of speech production across the final stages of life.

We created the "Micro Corpus for the Longitudinal Study of Dementia" - μCLSD - by selecting 16 subjects, 8 who died from Alzheimer's disease (the AD group) and 8 who passed away due to other causes (the HC group), such as old age or illnesses not directly associated with cognitive impairment[2]. The sample is gender-balanced, and each AD subject is paired with an HC counterpart of the same gender and with a comparable professional background. Each group was further subdivided into four subjects speaking British English and four subjects speaking American English, in order to evaluate the approach across different varieties of English.

The selection of subjects was also guided by the availability of interviews on YouTube covering a wide time span of their lives, allowing us to reasonably assume that the earliest recordings were produced during periods unaffected by any cognitive disease.

From a technical perspective, we manually extracted audio fragments from these interviews, each lasting between one and one minute fifteen seconds. The interviewer's voice and external noise (e.g., music or applause) were removed to approximate the conditions of a spontaneous monologue in a quiet environment. All audio files were then resampled at 16 kHz, 16 bits and reduced to a single channel.

3.2. Our Pipeline for Extracting DLBs

Natural Language Processing (NLP) techniques and tools are playing an increasingly vital role in the medical field (Wang et al., 2020), supporting a wide spectrum of applications such as patient care, diagnostics, clinical coding, and patient-oriented services (Locke et al., 2021). In particular, there is a rising interest in leveraging automated speech and language analysis as a promising early indicator of pathological processes.

A newly built DLB pipeline (v2.0), based on our previous work (Gagliardi and Tamburini, 2022), processes audio signals to generate DLBs for each sample. It consists of two phases: preprocessing and feature extraction.

During the preprocessing phase, the input speech audio is transcribed relying on OpenAI Whisper-v3 using the "medium-en" model (Radford et al., 2023), then voice activity detection (Bredin et al., 2020), voiced segment identification, vowel-consonant distinction (Li et al., 2020), dependency parsing with UDPipe (Kondratyuk and Straka, 2019) and constituency parsing using STANZA (Qi et al., 2020) are performed on the input speech or its automatic transcription.

The feature extraction phase computes DLBs listed in Table 1 (please, refer to Calzà et al. 2021 for a detailed description) using the information obtained during preprocessing. These DLBs can be categorised into five groups: Acoustic, Rhythmic, Lexical, LIWC based counts (Pennebaker et al., 2015), and Syntactic DLBs, which, taken together, offer a fine-grained representation of the linguistic patterns related to subject cognitive abilities (Gagliardi and Tamburini, 2022).

Figure 3 depicts the overall structure of our pipeline. The tool can process three different kinds of inputs: spoken recordings (as a WAV audio file), raw written texts (TXT) transcriptions, or preprocessed texts in the CoNLL-U format containing morphosyntactic and syntactic analyses. Given a specific input type, either a WAV, TXT, or CoNLL file, the pipeline computes all the DLBs that can be derived from it. The larger set is obtained by providing the speech recording, alone or with the manual transcription (to bypass any mistake produced by the ASR module).

It is relevant to underscore that, for the experiments presented in this paper, we provide only the speech audio (WAV) file to the pipeline, thus any further computation must start from this single information and no manual effort is needed to process interview recordings for feature extraction. We conducted a series of tests to evaluate the effectiveness of the OpenAI Whisper-v3 model in transcribing English utterances, using data from the PROCESS Challenge as a benchmark (see Section 1.2). The model achieved a Word Error Rate (WER) of 8.7%, which appears satisfactory given that the speech is spontaneous and may include pathological traits. Our analysis showed that the primary

[2]Of course, in the absence of other clinical information, the HC group could, in principle, include individuals with undiagnosed cognitive impairment of some kind.

Acoustic DLBs (SPE)
Silence segments duration (M, MD, SD)
Speech segments duration (M, MD, SD)
Temporal regularity of voiced segments
Verbal Rate
Transformed Phonation Rate
Standardised Phonation Time
Standardised Pause Rate
Root Mean Square energy (M, SD)
Pitch (M, SD)
Spectral Centroid (M, SD)
Higuchi Fractal Dimension (M, SD)
Rhythmic DLBs (RHY)
Percentage of vocalic intervals - %V
SD of vocalic, ΔV, and cons., ΔC, interval durations
Pairwise Variability Index, raw, rPVI, and norm., nPVI
Variation coefficient for ΔV and ΔC
Lexical DLBs (LEX)
Content Density
Part-of-Speech rate
Reference Rate to Reality
Personal, Spatial and Temporal Deixis rate
Relative pronouns and negative adverbs rate
Lexical Richness: TTR, Brunet's and Honoré's Indexes
Action Verbs rate
Frequency-of-use tagging
Propositional Idea Density
Mean Number of words in utterances
Linguistic Inquiry and Word Count DLBs (LWC)
Language Metrics (e.g., words per sentence
Function Words (e.g., pronouns, articles, ...)
Affect Words (e.g., positive/negative emotion)
Cognitive Processes (e.g., insight, certainty, ...)
Perceptual processes (e.g., seeing, hearing, feeling)
Biological processes (e.g., body, health/illness, ...)
Personal concerns (e.g., work, leisure, money, ...)
Social Words (e.g., family, friends)
Punctuation (e.g., periods, commas, colons, ...)
Syntactic DLBs (SYN)
Number of dependent elements of the nouns (M, SD)
Global Dependency Distance (M, SD)
Syntactic complexity
Syntactic embeddedness: maximum tree depth (M, SD)
Utterance length (M, SD)

Table 1: The list of Digital Linguistic Biomarkers extracted by the pipeline. Some of these features are computed as means (M), medians (MD), and standard deviations (SD). Please refer to Calzà et al. (2021) for extended descriptions and computation details.

source of errors stems from the hyper-normalization tendency of ASR systems, which often remove or correct disfluencies, restarts, and repeated words to produce cleaner transcriptions. While this could pose a significant issue if disfluency counts were used as input features, we deliberately chose not to rely on such information. As a result, our system is only minimally affected by this limitation.

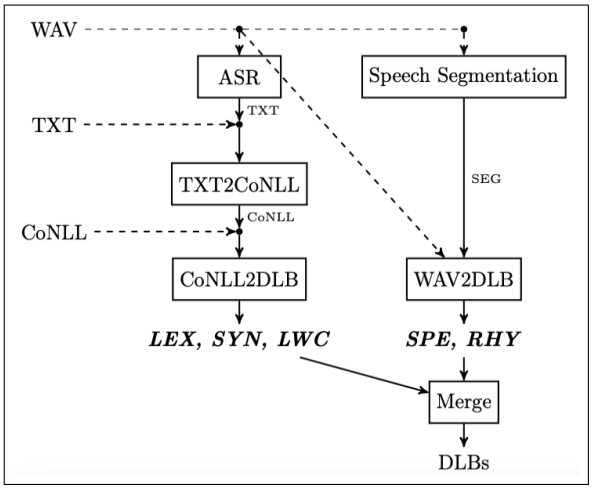

Figure 3: The whole structure of the Pipeline. Inputs can be provided as WAV, raw text, or CoNLL files. The modules are described in detail in Calzà et al. (2021) and Gagliardi and Tamburini (2022). In the experiments presented in this paper, we use only the audio WAV signal; all computations required to extract DLBs are carried out automatically.

3.3. Feature Processing & Selection

The Explainable Boosting Classifier (EBC)[3] (Lou et al., 2012) is an interpretable machine learning model from the family of Generalised Additive Models (GAMs). It combines the predictive power of boosting with the transparency of GAMs. Instead of learning a single complex function, it learns feature shape functions (one per feature, plus interactions, if allowed), which describe how each feature contributes to the prediction.

For feature ranking, the EBC provides global "importance scores" by measuring how much each feature contributes to the model's predictions across the dataset. This is typically done by (a) evaluating the magnitude of each feature's shape function (larger deviations indicate stronger impact) and (b) comparing across features to rank them by influence on the target outcome. This makes EBC especially useful in domains where both accuracy and interpretability matter, since it produces rankings along with human-readable explanations of how features affect predictions.

To determine the most relevant features for this task, we relied on the dataset provided for the PROCESS Challenge. The pipeline system described in the previous section took part in the competition and achieved strong performance, ranking first in the three-way classification of HC vs. MCI vs. AD (Zhang et al., 2025). Using the training and validation sets provided by this challenge, we carried out the following steps:

[3] https://interpret.ml

- extracted all DLBs listed in Table 1 using our software pipeline;
- normalised each feature using z-scores;
- applied the previously described EBC algorithm, retaining only DLBs with an importance score ≥ 0.01, resulting in the selection of 109 out of 126 features.

3.4. Drawing Cognitive Decline Profiles

Novelty detection with Local Outlier Factor (LOF) is a technique used to identify new data points that differ from the training distribution.

The LOF algorithm (Breunig et al., 2000) measures the local density deviation of a data point compared to its neighbours. The key idea is: points in dense regions are considered normal, while points in sparse regions, especially if their density is much lower than that of their neighbours, are considered novelties or outliers. For novelty detection (as opposed to outlier detection in training data), LOF is trained on "normal" examples only. New incoming samples are then scored: a score close to 1 means "normal", while larger scores indicate potential novelties/anomalies. This makes LOF useful in applications like fraud detection, fault monitoring, or detecting rare events in streaming data.

A short mathematical introduction for LOF in novelty detection could be described as:

- **k-distance and neighbors**. For each point x, find its k nearest neighbours $N_k(x)$ using distance $d(x, y)$. The *k-distance* is the distance to the k^{th} nearest neighbour.
- **Reachability distance**. For a point x and a neighbour y, the *reachability distance* is defined as:

$$\text{reach-dist}_k(x, y) = \max\big(d(x, y),\, k\text{-distance}(y)\big).$$

- **Local reachability density (LRD)**. The local reachability density of x is the inverse of the average reachability distance to its neighbours:

$$\text{lrd}_k(x) = \frac{|N_k(x)|}{\sum_{y \in N_k(x)} \text{reach-dist}_k(x, y)}.$$

- **LOF score**. The Local Outlier Factor compares the density of x with that of its neighbors:

$$\text{LOF}_k(x) = \frac{1}{|N_k(x)|} \sum_{y \in N_k(x)} \frac{\text{lrd}_k(y)}{\text{lrd}_k(x)}.$$

$\text{LOF}_k(x)$ can be easily interpreted as:

- $\text{LOF}_k(x) \approx 1$: x has similar density to neighbors $\Rightarrow$ x is in line with the training set (normal condition).
- $\text{LOF}_k(x) > 1$: x has much lower density $\Rightarrow$ x is an outlier or is different from points in the training set (novelty, thus a potentially pathologic condition).

We employed a LOF-based novelty detection approach to visualise the temporal evolution of subjects' cognitive abilities. Features were first normalised using z-scores, and the resulting values were then smoothed by computing a weighted moving average with a window of three samples, where weights reflected the temporal distance between sample pairs, to reduce the influence of transient spikes on neighbourhood contributions.

For each subject, the first four recordings were used as the training set for the LOF algorithm[4] assuming that these early samples represent speech unaffected by disease and thus serve as a reference for that individual. Subsequent recordings from the same subject were then compared against this reference LOF model, producing a score that reflects the degree of deviation from the reference. These scores were arranged to generate plots depicting each subject's cognitive functions trajectory over time in a way similar to Figure 1.

4. Results and Discussion

Figure 4 presents the computed cognitive function profiles. Comparing the first eight profiles of subjects diagnosed with AD to the eight profiles of healthy controls (HC) reveals notable differences: cognitive function profiles of AD subjects show significant deviations from the reference samples (the first four points in each profile) well before the corresponding markers of the official diagnosis. In contrast, the profiles of HC subjects remain stable until very advanced ages, reflecting the typical pattern of cognitive decline associated with normal ageing. The extracted DLBs and our proposed method for tracing their evolution over time appear to be sensitive to the differing cognitive trajectories of the two subject groups, allowing for precise detection of subtle speech variations linked to cognitive decline.

Our approach departs fundamentally from previous studies in the literature: rather than applying a classification process to determine a subject's cognitive status at a specific point in time comparing its DLBs with other subjects, we generate a continuous cognitive function profile that evolves across the ageing process considering only the DLBs of a single subject across time. This allows us to simulate an individual's cognitive trajectory and identify

[4] We relied on the Scikit-Learn LOF module. (https://scikit-learn.org/stable/auto_examples/neighbors/plot_lof_novelty_detection.html).

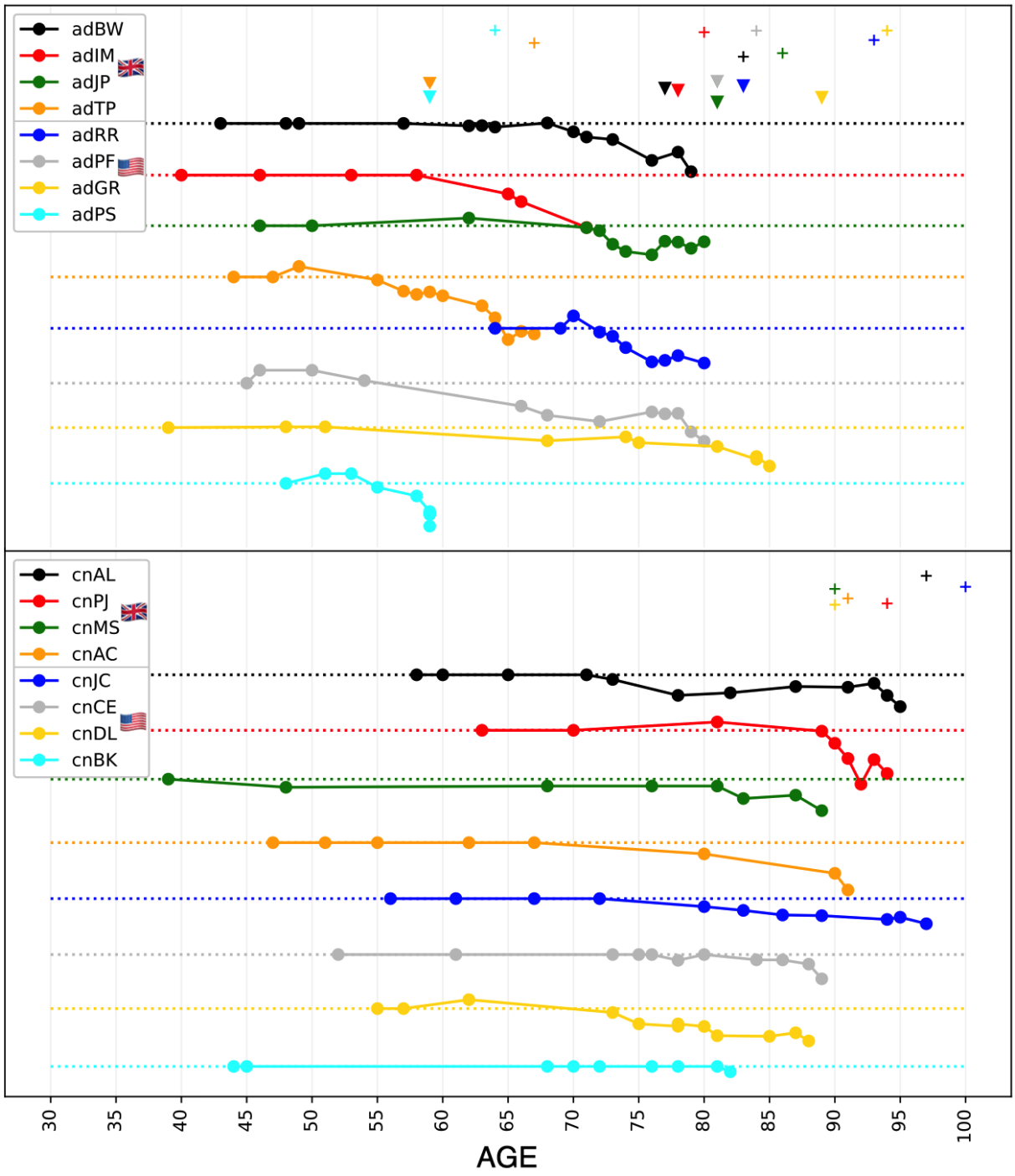

Figure 4: Evolution of cognitive functions over time for the sixteen subjects forming the μCLSD corpus. On top of each picture crosses mark the age of death, while triangles the year of first diagnosis for AD subjects. The top picture presents the profiles for the AD subjects (marked with 'ad' before name initials) while the bottom picture for HC (marked with 'cn'). Flags contained into the legenda indicate the English varieties spoken by the corresponding subjects.

significant deviations from his/her personal reference baseline, thereby detecting alterations relative to his/her normal cognitive status.

We do not envision this method being applied within specialised clinical practice. Instead, we propose it as a pilot approach to help define protocols for large-scale screening of ageing populations, aimed at detecting the earliest signs of cognitive decline. In this framework, general practitioners could use the method as a simple first-level tool and, when needed, refer individuals for specialist-administered neuropsychological assessments. Indeed, observed changes in a subject's cognitive profile do not necessarily indicate a true impair-

ment; rather, they may simply highlight a condition that warrants further evaluation by an expert to clarify the nature of these variations in cognitive functioning.

We do not claim that the proposed method can reliably screen entire populations or accurately distinguish individuals with cognitive impairment from healthy subjects. Rather, its purpose could be to support general practitioners in identifying potential concerns and referring these individuals to more specific and reliable assessments. In the absence of widely applicable large-scale screening tools, this approach could represent a practical solution, enabling general practitioners to serve as an initial filter for detecting individuals who may be at risk of cognitive impairment, even before any symptoms become apparent. For example, the method may capture within-person changes in speech, such as those related to general health variations or voice alterations due to other conditions, rather than patterns specifically associated with Alzheimer's disease. However, a general practitioner, being familiar with the individual's overall health status, may judge these signals as non-relevant and decide not to refer the person for further evaluation.

In the near future, we plan to evaluate whether the cognitive profile extraction method described in this paper remains an effective detection tool across different varieties of English (e.g. Australian English) and for other typologically distinct languages.

5. Limitations and Ethical Considerations

This pilot study is primarily intended to support the paradigm shift we propose for the early identification of cognitive decline. While the μCSLD dataset we collected is too limited in size and restricted to a single language to allow for broad generalisation, we contend that our approach provides a solid basis for devising new large-scale population screening methods based on DLBs, thereby addressing the limitations of the classification-based methods currently prevalent in the literature, and it favours the development of methods based on longitudinal analyses.

Regarding the corpus, subjects' data were anonymised in this paper but not in the dataset, as all recordings were obtained from publicly accessible sources on the Internet, primarily Wikipedia and YouTube. While voices may be recognisable and interview topics could potentially reveal identity, making full anonymisation of the dataset impossible, we believe that sharing such data is crucial to enable further research in this direction. Unfortunately, given that we selected dead subjects, it was not possible to collect any kind of consent from them.

For these reasons, the dataset will be available only upon request. It will include the audio recordings of the interviews, along with all references to the original sources (primarily URLs), and a clear listing of the subjects' names.

6. Acknowledgements

This study was funded by the European Union–NextGenerationEU programme through the Italian National Recovery and Resilience Plan– NRRP (Mission 4–Education and research), as a part of the project ReMind: an ecological, costeffective AI platform for early detection of prodromal stages of cognitive impairment (PRIN 2022, 2022YKJ8FP– CUP J53D23008380006).

7. Bibliographical References

Alzheimer's Association. 2025. 2025 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 21(4):e70235.

Sachin Anap, Satish Jondhale, Balasaheb Agarkar, and Sachin Chaudhari. 2025. Parkinson's disease detection from speech using combination of empirical wavelet transform and Hilbert transform. *International Journal of Speech Technology*, 28(1):185–194.

Tomas Arias-Vergara, Juan Camilo Vasquez Correa, Juan Rafael Orozco-Arroyave, and Elmar Nöth. 2018. Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions. *Speech Communication*, 101(101):11–25.

Achraf Benba, Abdelilah Jilbab, and Ahmed Hammouch. 2016. Voice Assessments for Detecting Patients with Parkinson's Diseases Using PCA and NPCA. *International Journal of Speech Technology*, 19(4):743–754.

Hervé Bredin, Ruiqing Yin, Juan Manuel Coria, Gregory Gelly, Pavel Korshunov, Marvin Lavechin, et al. 2020. Pyannote. audio: neural building blocks for speaker diarization. In *Proc. ICASSP*, pages 7124–7128.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, page 93–104, New York, NY, USA. Association for Computing Machinery.

L. Calzà, G. Gagliardi, Rema Rossini Favretti, and F. Tamburini. 2021. Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia. *Computer Speech & Language*, 65:101–113.

Marco Catani, Flavio Dell'Acqua, Alberto Bizzi, Stephanie J. Forkel, Steve C. Williams, Andrew Simmons, Declan G. Murphy, and Michel Thiebaut de Schotten. 2012. Beyond cortical localization in clinico-anatomical correlation. *Cortex*, 48(10):1262–1287.

Ho-Ling Chang, Thiri Wai, Yu-Shan Liao, Sheng-Ya Lin, Yu-Ling Chang, and Li-Chen Fu. 2025. A dual-modal fusion framework for detection of mild cognitive impairment based on autobiographical memory. *IEEE Journal of Biomedical and Health Informatics*, 29(6):4474–4485.

Hamidreza Chinaei, Leila Chan Currie, Andrew Danks, Hubert Lin, Tejas Mehta, and Frank Rudzicz. 2017. Identifying and Avoiding Confusion in Dialogue with People with Alzheimer's Disease. *Computational Linguistics*, 43(2):377–406.

Rosie Coppieters, Arabella Bouzigues, Lize Jiskoot, Maxime Montembeault, Boon Lead Tee, Jonathan D. Rohrer, Rose Bruffaerts, and et al. 2024. A systematic review of the quantitative markers of speech and language of the frontotemporal degeneration spectrum and their potential for cross-linguistic implementation. *Neuroscience & Biobehavioral Reviews*, 167:105909.

Ester Cornacchia, Aurora Bonvino, Giorgia Francesca Scaramuzzi, Daphne Gasparre, Roberta Simeoli, Davide Marocco, and Paolo Taurisano. 2025. Digital Screening for Early Identification of Cognitive Impairment: A Narrative Review. *WIREs Cognitive Science*, 16(4):e70009.

Sofia de la Fuente Garcia, Craig W. Ritchie, and Saturnino Luz. 2020. Artificial Intelligence, Speech, and Language Processing Approaches to Monitoring Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease*, 78(4):1547–1574.

Kewen Ding, Madhu Chetty, Azadeh Noori Hoshyar, Tanusri Bhattacharya, and Britt Klein. 2024. Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges. *Artificial Intelligence Review*, 57:325.

Leandro B. dos Santos, Edilson Anselmo Corrêa Jr., Osvaldo N. Oliveira Jr, Diego R. Amancio, L. Mansur, and Sandra M. Aluísio. 2017. Enriching Complex Networks with Word Embeddings for Detecting Mild Cognitive Impairment from Speech Transcripts. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, volume Volume 1: Long Papers, pages 1284–1296. Association for Computational Linguistics.

José Vicente Egas-López, Réka Balogh, Nóra Imre, Ildikó Hoffmann, Martina Katalin Szabó, László Tóth, Magdolna Pákáski, János Kálmán, and Gábor Gosztolya. 2022. Automatic screening of mild cognitive impairment and Alzheimer's disease by means of posterior-thresholding hesitation representation. *Computer Speech & Language*, 75:101377.

Elif Eyigoz, Sachin Mathur, Mar Santamaria, Guillermo Cecchi, and Melissa Naylor. 2020. Linguistic markers predict onset of Alzheimer's disease. *EClinicalMedicine*, 28:100583.

Anna Favaro, Tianyu Cao, Najim Dehak, and Laureano Moro-Velázquez. 2024. Leveraging Universal Speech Representations for Detecting and Assessing the Severity of Mild Cognitive Impairment Across Languages. In *Proc. Interspeech 2024*, Kos, Greece.

Gerasimos Fergadiotis and Heather Harris Wright. 2011. Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, 25:1414–1430.

Kathleen C. Fraser, Jed A. Meltzer, Naida L. Graham, Carol Leonard, Graeme Hirst, Sandra E. Black, and Elizabeth Rochon. 2014. Automated classification of Primary Progressive Aphasia subtypes from narrative speech transcripts. *Cortex*, 55:43–60.

Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2016. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's Disease*, 49:407–422.

Gloria Gagliardi. 2024. Natural language processing techniques for studying language in pathological ageing: A scoping review. *Int J Lang Commun Disord*, 59:110–122.

Gloria Gagliardi, Dimitrios Kokkinakis, and Jon Andoni Duñabeitia. 2021. Editorial: Digital linguistic biomarkers: Beyond paper and pencil tests. *Frontiers in Psychology*, 12:752238.

Gloria Gagliardi and Fabio Tamburini. 2022. The automatic extraction of linguistic biomarkers as a viable solution for the early diagnosis of mental disorders. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5234–5242, Marseille, France. European Language Resources Association.

Dimitris Gkoumas, Bo Wang, Adam Tsakalidis, Maria Wolters, Matthew Purver, Arkaitz Zubiaga1, and Maria Liakata. 2024. A longitudinal multi-modal dataset for dementia monitoring and diagnosis. *Language Resources and Evaluation*, 58:883–902.

Gábor Gosztolya, Réka Balogh, Nóra Imre, José Vicente Egas-López, Ildikó Hoffmann, Veronika Vincze, László Tóth, Davangere P. Devanand, Magdolna Pákáski, and János Kálmán. 2021. Cross-lingual detection of mild cognitive impairment based on temporal parameters of spontaneous speech. *Computer Speech & Language*, 69:101215.

P. Hagoort. 2017. The core and beyond in the language-ready brain. *Neuroscience and biobehavioral reviews*, 81(Pt B):194–204.

Ingo Hertrich, Susanne Dietrich, and Hermann Ackermann. 2020. The margins of the language network in the brain. *Frontiers in Communication*, 5:93.

Lihe Huang, Hao Yang, Yiran Che, and Jingjing Yang. 2024. Automatic speech analysis for detecting cognitive decline of older adults. *Frontiers in Public Health*, 12.

Olga Ivanova, Juan José G. Meilán, Francisco Martínez-Sánchez, Israel Martínez-Nicolás, Thide E. Llorente, and Nuria Carcavilla González. 2022. Discriminating speech traits of Alzheimer's disease assessed through a corpus of reading task for Spanish language. *Computer Speech & Language*, 73:101341.

William L. Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar. 2014. Aided Diagnosis of Dementia Type through Computer-Based Analysis of Spontaneous Speech. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 27–37. ACL - Association for Computational Linguistics.

Haihua Jiang, Bin Hu, Zhenyu Li, Lihua Yan, Tianyang Wang, Fei Liu, Huanyu Kang, and Xiaoyu Li. 2017. Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, 90:39–46.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing universal dependencies universally. *arXiv preprint arXiv:1904.02099*.

M. Kumar, Sushant, and A. Yadav. 2025. Speech signal's phase information based Alzheimer's disease detection using deep learning. *International Journal of Speech Technology*, 28:397–410.

Jordi Laguarta and Brian Subirana. 2021. Longitudinal Speech Biomarkers for Automated Alzheimer's Detection. *Frontiers in Computer Science*, 3.

Chuyuan Li, Raymond Li, Thalia S. Field, and Giuseppe Carenini. 2025. Delta-KNN: Improving demonstration selection in in-context learning for Alzheimer's disease detection. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25807–25826, Vienna, Austria. Association for Computational Linguistics.

Xinjian Li, Siddharth Dalmia, Juncheng Li, Matthew Lee, Patrick Littell, Jiali Yao, Antonios Anastasopoulos, et al. 2020. Universal phone recognition with a multilingual allophone system. In *Proc. ICASSP*, pages 8249–8253.

Saskia Locke, Anthony Bashall, Sarah Al-Adely, John Moore, Anthony Wilson, and Gareth B. Kitchen. 2021. Natural language processing in medicine: A review. *Trends in Anaesthesia and Critical Care*, 38:4–9.

Karmele López-de-Ipiña, Jordi Solé-Casals, Harkaitz Eguiraun, J.B. Alonso, C.M. Travieso, Aitzol Ezeiza, Nora Barroso, Miriam Ecay-Torres, Pablo Martinez-Lage, and Blanca Beitia. 2015. Feature selection for spontaneous speech analysis to aid in alzheimer's disease diagnosis: A fractal dimension approach. *Computer Speech & Language*, 30(1):43 – 60.

Yin Lou, Rich Caruana, and Johannes Gehrke. 2012. Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page 150–158, New York, NY, USA. Association for Computing Machinery.

Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, and Brian MacWhinney. 2021a. Detecting Cognitive Decline Using Speech Only: The ADReSSo Challenge. In *Interspeech 2021*, pages 3780–3784.

Saturnino Luz, Fasih Haider, Sofia de la Fuente Garcia, Davida Fromm, and Brian MacWhinney. 2021b. Editorial: Alzheimer's Dementia Recognition through Spontaneous Speech. *Frontiers in Computer Science*, Volume 3 - 2021.

Martin Lövdén, Laura Fratiglioni, M. Maria Glymour, Ulman Lindenberger, and Elliot M. Tucker-Drob. 2020. Education and Cognitive Functioning Across the Life Span. *Psychological Science in the Public Interest*, 21(1):6–41.

Cintia Matsuda Toledo, Sandra Maria Aluisio, Leandro Borges dos Santos, Sonia Maria Dozzi Brucki, Eduardo Sturzeneker Trés, Maira Okada de Oliveira, and Letícia Lessa Mansur. 2018. Analysis of macrolinguistic aspects of narratives from individuals with Alzheimer's disease, mild cognitive impairment, and no cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10:31–40.

J.J.G. Meilán, F. Martínez-Sánchez, J. Carro, N. Carcavilla, and O. Ivanova. 2018. Voice Markers of Lexical Access in Mild Cognitive Impairment and Alzheimer's Disease. *Current Alzheimer Research*, 15:111–119.

Juan J. G. Meilán, Francisco Martínez-Sánchez, Israel Martínez-Nicolás, Thide E. Llorente, and Juan Carro. 2020. Changes in the Rhythm of Speech Difference between People with Nondegenerative Mild Cognitive Impairment and with Preclinical Dementia. *Behavioural Neurology*, 2020(1):4683573.

Carmen Moret-Tatay, Isabel Iborra-Marmolejo, María José Jorques-Infante, Gloria Bernabé-Valero, María José Beneyto-Arrojo, and Tatiana Quarti Irigaray. 2023. A pilot screening for cognitive impairment through voice technology (WAY2AGE). *BMC Psychology*, 11(1):170.

Ivan Nenchev, Tatjana Scheffler, Marie de la Fuente, Heiner Stuke, Benjamin Wilck, Sandra Anna Just, and Christiane Montag. 2024. Linguistic markers of schizophrenia: a case study of Robert Walser. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)*, pages 41–60, St. Julians, Malta. Association for Computational Linguistics.

M. Nucci, D. Mapelli, and S. Mondini. 2012. The cognitive Reserve Questionnaire (CRIq): a new instrument for measuring the cognitive reserve. *Aging clinical and experimental research*, 24:218–226.

James Pennebaker, Roger Booth, Ryan Boyd, and Martha Francis. 2015. Linguistic inquiry and word count: Liwc2015. Technical report.

U. Petti, S. Baker, and A. Korhonen. 2020. A systematic literature review of automatic alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association*, 27(11):1784–1797.

Ulla Petti, Simon Baker, Anna Korhonen, and Jessica Robin. 2023. The Generalizability of Longitudinal Changes in Speech Before Alzheimer's Disease Diagnosis. *Journal of Alzheimer's Disease*, 92(2):547–564.

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

Ali Pourramezan Fard, Mohammad H. Mahoor, Muath Alsuhaibani, and Hiroko H. Dodge. 2024. Linguistic-based Mild Cognitive Impairment detection using Informative Loss. *Computers in Biology and Medicine*, 176:108606.

María Luisa Barragán Pulido, Jesús Bernardino Alonso Hernández, Miguel Ángel Ferrer Ballester, Carlos Manuel Travieso González, Jiří Mekyska, and Zdeněk Smékal. 2020. Alzheimer's disease and automatic speech analysis: A review. *Expert Systems with Applications*, 150:113213.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 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*. ACL, Stroudsburg (PA).

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

Brian Roark, Margaret Mitchell, and Kristy Hollingshead. 2007. Syntactic complexity measures for detecting Mild Cognitive Impairment. In Kevin Bretonnel Cohen, Dina Demner-Fushman, Carol Frieman, Lynette Hirschman, and John Pestian, editors, *Proceedings of the Workshop BioNLP 2007: Biological, translational, and clinical language processing*, pages 1–8. ACL, Stroudsburg (PA).

Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead, and Jeffrey A. Kaye. 2011. Spoken Language Derived Measures for Detecting Mild Cognitive Impairment. *IEEE Transactions on Audio Speech, and Language Processing*, 19(7):2081–2090.

Jessica Robin, Mengdan Xu, Aparna Balagopalan, Jekaterina Novikova, Laura Kahn, Abdi Oday, Mohsen Hejrati, Somaye Hashemifar, Mohammadreza Negahdar, William Simpson, and Edmond Teng. 2023. Automated detection of progressive speech changes in early alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(2):e12445.

Aharon Satt, Alexander Sorin, Orith Toledo-Ronen, Oren Barkan, Ioannis Kompatsiaris, Athina Kokonozi, and Magda Tsolaki. 2013. Evaluation of Speech-Based Protocol for Detection of Early-Stage Dementia. In *Proceedings of Interspeech 2013*, pages 1692–1696. ISCA, Grenoble.

Arezo Shakeri and Mina Farmanbar. 2025. Natural language processing in alzheimer's disease research: Systematic review of methods, data, and efficacy. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 17(1):e70082.

Ravi Shankar, Anjali Bundele, and Amartya Mukhopadhyay. 2025. A systematic review of natural language processing techniques for early detection of cognitive impairment. *Mayo Clinic Proceedings: Digital Health*, 3(2):100205.

Nutan Singh and Priyanka Tripathi. 2024. An ensemble technique to predict Parkinson's disease using machine learning algorithms. *Speech Communication*, 159:103067.

Kairit Sirts, Olivier Piguet, and Mark Johnson. 2017. Idea density for predicting Alzheimer's disease from transcribed speech.

Reisa A. Sperling, Paul S. Aisen, Laurel A. Beckett, David A. Bennett, Suzanne Craft, Anne M. Fan, Takeshi Iwatsubo, Clifford R. Jack Jr., Jeffrey Kaye, Thomas J. Montine, Denise C. Park, Eric M. Reiman, Christopher C. Rowe, Eric Siemers, Yaakov Stern, Kristine Yaffe, Maria C. Carrillo, Bill Thies, Marcelle Morrison-Bogorad, Molly V. Wagster, and Creighton H. Phelps. 2011. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):280–292.

Brian Stasak, Julien Epps, and Roland Goecke. 2019. Automatic depression classification based on affective read sentences: Opportunities for text-dependent analysis. *Speech Communication*, 115:1–14.

Dávid Sztahó and Klára Vicsi. 2016. Estimating the Severity of Parkinson's Disease Using Voiced Ratio and Nonlinear Parameters. In Pavel Král and Carlos Martín-Vide, editors, *Statistical Language and Speech Processing*, pages 96–107. Springer International Publishing, Cham.

Fuxiang Tao, Bahman Mirheidari, Madhurananda Pahar, Sophie Young, Yao Xiao, Hend Elghazaly, Fritz Peters, Caitlin Illingworth, Dorota Braun, Ronan O'Malley, Simon Bell, Daniel Blackburn, Fasih Haider, Saturnino Luz, and Heidi Christensen. 2025. Early Dementia Detection Using Multiple Spontaneous Speech Prompts: The PROCESS Challenge. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

László Tóth, Ildiko Hoffmann, Gábor Gosztolya, Veronika Vincze, Gréta Szatlóczki, Zoltán Bánréti, Magdolna Pákáski, and János Kálmán. 2018. A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. *Current Alzheimer Research*, 15:1–10.

Savitha S. Upadhya, A. N. Cheeran, and J. H. Nirmal. 2019. Discriminating parkinson diseased and healthy people using modified mfcc filter bank approach. *International Journal of Speech Technology*, 22(4):1021–1029.

Veronika Vincze, Gábor Gosztolya, László Tóth, Ildikó Hoffmann, Gréta Szatlóczki, Zoltán Bánréti, Magdolna Pákáski, and János Kálmán. 2016. Detecting Mild Cognitive Impairment by Exploiting Linguistic Information from Transcripts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 181–187. Association for Computational Linguistics.

Jing Wang, Huan Deng, Bangtao Liu, Anbin Hu, Jun Liang, Lingye Fan, Xu Zheng, Tong Wang, and Jianbo Lei. 2020. Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: Bibliometric study on pubmed. *Journal of Medical Internet Research*, 22(1):e16816.

Meng Wang, Yanxia Wen, Shicong Mo, Liqiong Yang, Xiaqing Chen, Man Luo, Hongdian Yu, Fan Xu, and Xianwei Zou. 2022. Distinctive acoustic changes in speech in Parkinson's disease. *Computer Speech & Language*, 75:101384.

Tianqi Wang, Yin Hong, Quanyi Wang, Rongfeng Su, Manwa Lawrence Ng, Jun Xu, Lan Wang, and Nan Yan. 2021. Identification of Mild Cognitive Impairment Among Chinese Based on Multiple Spoken Tasks. *Journal of Alzheimer's Disease*, 82(1):185–204.

Tianqi Wang, Chongyuan Lian, Jingshen Pan, Quanlei Yan, Feiqi Zhu, Manwa L. Ng, Lan Wang, and Nan Yan. 2019. Towards the Speech Features of Mild Cognitive Impairment: Universal Evidence from Structured and Unstructured Connected Speech of Chinese. In *Proc. Interspeech 2019*, pages 3880–3884.

Zaifa Xue, Huibin Lu, Tao Zhang, Jiahui Xu, and Xiaonan Guo. 2023. A local dynamic feature

selection fusion method for voice diagnosis of parkinson's disease. *Computer Speech & Language*, 82:101536.

Yasunori Yamada, Kaoru Shinkawa, Miyuki Nemoto, Kiyotaka Nemoto, and Tetsuaki Arai. 2023. A mobile application using automatic speech analysis for classifying Alzheimer's disease and mild cognitive impairment. *Computer Speech & Language*, 81:101514.

Maria Yancheva and Frank Rudzicz. 2016. Vector-space topic models for detecting Alzheimer's disease. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 2337–2346. Association for Computational Linguistics.

X. Yang, K. Hong, D. Zhang, and K. Wang. 2024. Early diagnosis of Alzheimer's Disease based on multi-attention mechanism. *PLoS ONE*, 19(9):e0310966.

Shibingfeng Zhang, Nadia Khlif, Marcello Ferro, Gloria Gagliardi, and Fabio Tamburini. 2025. Cognitive Decline Detection using DLB Extraction Pipelines. In *Proceedings of Prediction and Recognition Of Cognitive declinE through Spontaneous Speech (PROCESS) Signal Processing Grand Challenge, ICASSP-25*, Hyderabad, India. IEEE.