Reading Dynamics and Comprehension in Cognitive Aging: A Multimodal Language Resource

Claudia Marzi*, **Noemi Boni***, **Alice Todesco***, **Andrea Nadalini***
Giorgia Albertin**, **Cristina Dolciotti**†, **Paolo Bongioanni**†, **Marcello Ferro***
Fabio Tamburini**, **Gloria Gagliardi****, **Vito Pirrelli***
*Italian National Research Council - Institute for Computational Linguistics "A. Zampolli"
**University of Bologna - FICLIT, †Pisa University Hospital
claudia.marzi@cnr.it

Abstract

We introduce a novel Italian language resource for the study of reading and comprehension in aging populations, combining behavioural and linguistic data from healthy controls (HC), individuals with subjective cognitive decline (SCI), participants with Mild Cognitive Impairment (MCI), and patients with mild dementia (CDR1). Reading performance was recorded through a *finger-tracking* based application during both silent and oral reading, enabling fine-grained temporal analyses at the text, token and character level. Comprehension was assessed via multiple question types (*wh*-, inferential, referential, and lexical). Descriptive and non-linear regression analyses informed a feature selection process, yielding temporal and comprehension-based measures that capture individual reading dynamics. These features were explored through unsupervised clustering and supervised classification to investigate their discriminative and predictive potential across cognitive profiles. The resource supports research on reading and cognitive decline, offers a reproducible protocol for large-scale data collection, and provides a foundation for developing early cognitive screening and monitoring tools for aging populations.

Keywords: reading behaviour, cognitive aging, finger-tracking, language resources for assistive technologies, cluster analysis, automatic classification.

1. Introduction

This paper presents *ReMind*, a novel language resource designed to support the study of reading and comprehension processes in Italian aging populations, with a particular focus on the early detection of cognitive decline. The dataset includes reading behavioural data from older adults belonging to four clinically-assessed groups – healthy controls (HC), individuals with subjective cognitive decline (SCI), participants with Mild Cognitive Impairment (MCI), and patients with mild dementia (CDR1) – engaged in both silent and oral reading tasks.

Reading behaviour was recorded with the *ReadLet* application, which administers reading tasks on a common tablet with finger movements and audio productions being tracked. This method enables fine-grained temporal measurements – from sentences in a connected text to individual characters in word tokens – providing a detailed profile of reading dynamics. Data are encrypted and pseudonymised before being processed in a secure cloud environment that integrates text, audio, and haptic streams (Marzi et al., 2025; Ferro et al., 2024; Nadalini et al., 2023; Crepaldi et al., 2022; Taxitari et al., 2021; Marzi et al., 2020; Ferro et al., 2018). In addition to reading performance and related temporal indices, the dataset integrates comprehension accuracy scores across different question types, covering *wh*-questions, inferential reasoning, referential resolution and lexical knowledge, which provide a multidimensional view of reading proficiency and text understanding. These multimodal behavioural indices provide a rich source of information for identifying subtle differences in reading efficiency and understanding. Importantly, such measures may help detect early, prodromal signs of cognitive decline, offering opportunities for timely interventions such as cognitive stimulation and lifestyle-based strategies. Preliminary descriptive and non-linear regression analyses guided the selection of key features, which were then used to explore group-level differences through unsupervised clustering and supervised classification. By focusing on subject-level aggregation of reading speed and comprehension accuracy, this approach provides a reproducible framework for analysing cognitive profiles based on natural reading behaviour.

Beyond its immediate utility, the dataset may serve as a benchmark for multimodal language research in aging populations, complementing traditional neuro-psychological assessments with non-invasive, ecologically valid measures. The framework is readily extendable to additional modalities, such as acoustic features and longitudinal data, enabling more comprehensive investigations of cognitive trajectories over time. Taken together, this resource bridges methodological and applied goals, by offering a protocol for high-resolution reading assessment, supporting theoretical research on language processing and aging, and providing the basis for developing practical, language-driven cognitive screening tools.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 2608–2618
11–16 May 2026.

2. Background

A growing body of research highlights the central role of language and reading processes in reflecting cognitive aging and early neuro-degenerative decline (Kokje et al., 2022; Burke and Shafto, 2004). Because reading integrates multiple linguistic and executive components, subtle changes in reading behaviour can reveal early alterations in cognitive processing efficiency, sometimes preceding overt clinical symptoms. Reading performance and text comprehension rely on the coordinated operation of phonological and visual decoding, working memory, attention and executive control (Reuter-Lorenz et al., 2021; Peng et al., 2018; Burke and Shafto, 2004). Evidence from cognitive aging research indicates that these processes are differentially affected by normative aging and by prodromal stages of cognitive decline (Wilson et al., 2020; Spaan, 2016). In normal aging, readers often exhibit slower decoding and strategic adaptations, including increased reliance on contextual prediction (Zhang et al., 2022), while attentional control and working memory limitations may reduce the efficient integration of information across sentences (Gordon et al., 2016). While some degree of slowing and strategic adaptation is expected in healthy aging, quantitative or qualitatively different alterations in the coordination of decoding, attention, and memory processes may signal the emergence of pathological cognitive decline. By capturing subtle variations in the temporal and comprehension profiles associated with these interrelated mechanisms, fine-grained behavioural measures can help distinguish normative aging from early cognitive impairment, potentially providing indicators of decline before overt clinical symptoms become evident.

In this context, coordination does not merely refer to overall processing speed, but also to the ability to flexibly modulate reading dynamics across modalities and task demands. Skilled readers typically exhibit differentiated temporal profiles in silent and oral reading, reflecting adaptive allocation of cognitive resources, as evidenced by task-dependent modulation of eye-movement dynamics and strategic prediction mechanisms (Zhang et al., 2022; Gordon et al., 2016; Rayner, 1998). Such differentiation is consistent with models of compensatory recruitment and executive control in aging (Reuter-Lorenz et al., 2021). A reduction in this differentiation may signal diminished processing flexibility or reduced efficiency in coordinating linguistic and executive systems.

These considerations suggest that detailed reading profiles may provide sensitive markers of incipient cognitive decline. Previous work on developmental and proficiency-related differences in reading dynamics has shown that readers coordinate eye, finger, voice temporal processes differently depending on their level of linguistic and cognitive maturity (Marzi et al., 2025; Nadalini et al., 2023; Crepaldi et al., 2022). Building on this evidence, it can be hypothesised that similar multimodal approaches may also be sensitive to early disruptions in the coordination of reading processes associated with prodromal stages of cognitive decline.

Despite the potential of reading as a cognitive marker, fine-grained behavioural datasets on older adults remain scarce, especially for languages other than English. Existing resources are often limited to a single modality – eye-tracking or speech analyses – without integrating comprehension outcomes or combining temporal and linguistic information. The ReMind dataset addresses this gap by providing a multimodal behavioural resource for Italian, combining finger-tracking, derived acoustic measures, and structured comprehension tasks, offering a unified, multimodal perspective on reading behaviour across clinical profiles. This approach allows for the extraction of behavioural signatures that can differentiate healthy aging from early cognitive decline, providing a foundation for predictive modelling.

Recent advances in non-invasive behavioural tracking have made high-resolution data collection feasible in both laboratory and ecological settings. Finger-tracking on touchscreens provides a cost-effective and scalable alternative to eye-tracking, enabling longitudinal assessments and large-scale studies without compromising temporal precision (Marzi et al., 2025; Nadalini et al., 2023; Crepaldi et al., 2022). In addition, by aggregating multiple features – temporal, comprehension-based, and modality-specific – the ReMind dataset supports both exploratory analyses (e.g., clustering) and supervised predictive modelling, capturing subtle individual differences in cognitive functioning.

Our approach aligns with a growing trend in language resource development that bridges theoretical research and practical applications. By providing ecologically valid, multimodal data, this resource may support early detection of cognitive decline, development of assistive technologies, and implementation of cognitive interventions.

3. Dataset and Experimental Protocol

The dataset includes reading behavioural data from older adults, all native speakers of Italian ($N = 89$, 44 females, 45 males, age range $57 - 88$, education range $4 - 21$, mean 12.97, sd 4.45), assigned to four groups: healthy controls (HC, $N = 38$), subjective cognitive decline (SCI, $N = 9$), individuals with Mild Cognitive Impairment (MCI, $N = 29$), and participants diagnosed with mild dementia (CDR1, $N = 13$). Group assignment was based on standard neuro-psychological evaluation and Clinical Dementia Rating (CDR) scores. All participants took part on a voluntary basis, after providing written informed consent, and were clinically assessed at the Pisa University Hospital.

The experimental protocol consisted of multiple language tasks, administered across three testing sessions spaced six months apart, allowing the investigation of longitudinal changes in reading and cognitive performance. Each participant completed two reading tasks (silent and oral) and two speech-elicitation tasks: a procedural planning task designed to assess discourse organisation and executive planning abilities, and two picture descriptions targeting spontaneous narrative production and referential language use. Reading texts were two short excerpts adapted and rearranged from original articles published in the Italian science communication magazine *Focus*, selected to provide age-appropriate content with sufficient lexical and syntactic complexity, ensuring ecological validity and engagement (Albertin et al., 2024). Together, these tasks enable a multimodal assessment of reading, comprehension, and speech production, providing complementary perspectives on cognitive functioning in aging populations.

Data collection was conducted using the ReadLet tablet-based application, which integrates reading, elicited spontaneous speech, and comprehension modules within a unified digital environment. For the reading component, we adopted the *finger-tracking* methodology, which records the participant's finger movements along a text as a proxy for eye-movement patterns. This approach provides high-resolution temporal data suitable for estimating processing times and decoding dynamics during reading (Nguyen et al., 2024; Crepaldi et al., 2022; Marzi et al., 2020; Lio et al., 2019). Each reading task was followed by comprehension questions targeting different levels of understanding: *wh*-, inferential, referential, and lexical items, thereby enabling a multi-layered assessment of comprehension efficiency. Audio data from oral reading and speech tasks were collected to be subsequently transcribed and analysed using an *Automatic Speech Recognition* approach, which allows the extraction of a wide range of acoustic and linguistic voice features (Badal et al., 2024; Hajjar et al., 2023; Gagliardi and Tamburini, 2021). A growing body of research supports the use of ASR-based acoustic markers as potential indicators for early detection of Mild Cognitive Impairment and dementia (Huang et al., 2024; Li et al., 2024).

The resulting multimodal dataset thus includes temporal traces (finger trajectories), and comprehension accuracy scores. Together, these measures provide a quantitative basis for inter-group comparison and for the integration of reading and speech data, enabling a fine-grained analysis of behavioural and cognitive profiles in aging populations. All statistical analyses and data visualisations were performed using R (version R-4.5.1, R Core Team, 2025).

4. Reading Profiles

We first present preliminary analyses of reading profiles across the four participant groups (HC, SCI, MCI, CDR1), focusing on temporal and accuracy measures, as well as the influence of lexical variables. The analyses reported here refer to data collected during the first test session of a longitudinal study, where each subject has been tested three times, with a six month interval between consecutive test sessions. Detailed results are reported in Appendix A.

Reading comprehension accuracy exhibited a clear pattern across groups. Accuracy, in fact, declined across groups from HC and SCI to MCI and CDR1. As shown in Figure 1, median accuracy values were high in HC and SCI participants, with comparable comprehension performance in these groups. Performance decreased in MCI participants, and further in CDR1, reflecting a progressive decline in comprehension associated with increasing cognitive impairment. Interestingly, comprehension accuracy was largely unaffected by reading modality, with median accuracy values overlapping in oral and silent reading modalities.

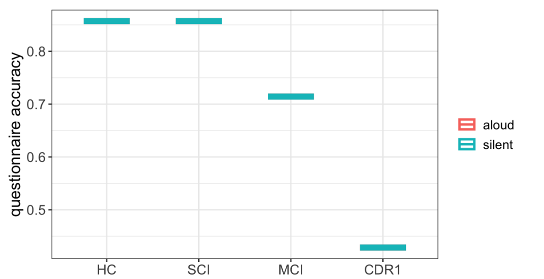

Figure 1: Median comprehension accuracy for four participant groups (Healthy Controls (HC), Subjective Cognitive Decline (SCI), Mild Cognitive Impairment (MCI), mild dementia (CDR1)) and two reading modalities (oral: red, silent: cyan). Median values for each group overlap, resulting in superimposed crossbars.

Reading speed per token was consistently higher during silent reading compared to oral reading across all four groups (HC, SCI, MCI, CDR1), as expected. However, the difference between silent and oral reading progressively decreased from HC to CDR1, as shown in Figure 2. While HC and SCI participants showed the largest speed advantage for silent reading, this advantage was smaller in MCI, and minimal in CDR1. A non-parametric *Kruskal-Wallis* test on the ratio between silent and oral reading speed – which provides a straightforward index of lexical automatisation – revealed a significant effect of group ($\chi^2(3) = 56.66$, $p < .001$). Post-hoc pairwise comparisons using *Dunn*'s test with *Bonferroni* correction indicated that the reduction in silent-aloud ratio was significantly larger in MCI and CDR1 participants compared to healthy controls (HC vs MCI: $p < .001$; HC vs CDR1: $p < .001$), and SCI participants also differed from MCI ($p < 0.01$). The progressive reduction of the speed advantage for silent reading – as compared to the oral one – may reflect a diminishing benefit from the absence of articulatory constraints and a decline in reading automaticity (Kim, 2015). This suggests that cognitive decline affects not only comprehension accuracy but also the efficiency of underlying reading processes, limiting the ability to flexibly exploit the usual temporal benefits of silent reading. Such subtle alterations in reading dynamics may therefore provide sensitive behavioural markers of early cognitive decline, complementing accuracy measures and lexical processing profiles.

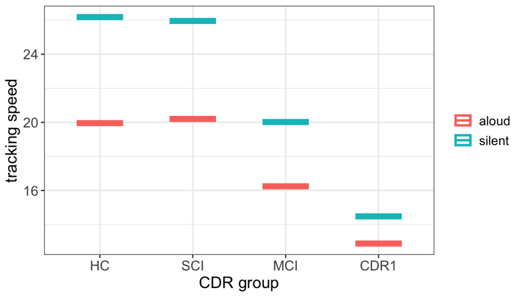

Figure 2: Median tracking speed for our four participant groups and two reading modalities (oral: red, silent: cyan).

Mean reading speed as a function of word token length showed a progressive decrease across groups, from HC and SCI to MCI, and CDR1 participants, both in silent and oral reading tasks (see Figure 3). Importantly, although absolute reading speeds decreased progressively from HC to CDR1, the overall profile remained consistent: the relative impact of length on reading speed did not qualitatively change across groups, with longer words associated with slower reading across all groups. This suggests that while cognitive decline reduces overall processing speed, the mechanisms underlying lexical access and word-level decoding remain largely preserved.

We also considered the role of cognitive reserve, operationalised as participants' education level. Cognitive reserve refers to the ability to maintain cognitive performance despite age-related or pathological decline, supported by life-long intellectual

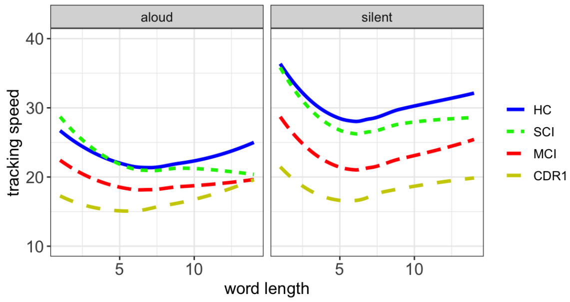

Figure 3: Non-linear regression fits for tracking speed as a function of word token length across our four participant groups, for oral (left) and silent reading (right).

enrichment and education (Stern, 2002; Arenaza-Urquijo and Vemuri, 2018; Cabeza et al., 2018).

For silent reading, a generalised additive model (GAM) was fitted with reading speed as the dependent variable, group and education level (grade level) as fixed effects, and participant as a random smooth. Results revealed a progressive reduction in reading speed across groups: while SCI participants did not significantly differ from HC, both MCI (–$4.56, p > .05$) and CDR1 (–$9.70, p < .001$) showed significantly slower reading. Education level had a positive effect ($0.95, p < .001$), indicating that higher cognitive reserve was associated with faster reading. The model accounted for 32% of deviance explained ($R^2adj = .32$). See Table 6.

For oral reading, the same GAM structure was applied. Results showed a weaker group effect: only CDR1 participants were significantly slower than controls (–$6.08, p < .05$), whereas SCI and MCI did not significantly differ from HC. Education level again showed a positive association with reading speed ($0.41, p < .05$), though smaller than in silent reading. The model, however, explained only 15% of deviance, suggesting less systematic modulation by cognitive status compared to silent reading (Table 7).

Taken together, these preliminary results provide a multidimensional characterisation of reading dynamics in aging populations, highlighting both the gradual quantitative decline in speed and comprehension and the relative preservation of reading profiles and lexical sensitivity. They also suggest that subtle changes in temporal dynamics – such as the decreasing advantage of silent reading – may be indicative of reduced automaticity and increased cognitive effort in text processing, offering potential markers for early detection of cognitive decline. This interpretation is supported by the GAM results, showing that cognitive decline and education level modulate reading speed more strongly in silent than in oral reading. This asymmetry suggests that silent reading, which relies more heavily on automatic decoding and self-paced comprehension processes, is more sensitive to subtle cognitive differences than aloud reading, where articulation constraints may mask fine-grained variability. The observed patterns are consistent with the notion that reading performance relies on the coordinated operation of phonological and visual decoding, working memory, and executive control (Reuter-Lorenz et al., 2021; Peng et al., 2018; Burke and Shafto, 2004). The progressive reduction in the speed advantage for silent reading, together with the modulation by cognitive reserve (Kremen et al., 2025; Arenaza-Urquijo and Vemuri, 2018; Cabeza et al., 2018; Stern, 2002), suggests that subtle changes in these interrelated processes may precede overt comprehension deficits, offering early behavioural markers of cognitive decline.

5. Feature Selection and Multivariate Analysis

5.1. Feature Selection and Preprocessing

Based on preliminary descriptive and non-linear/additive regression analyses, we selected a set of features capturing behavioural aspects of reading performance, as well as relevant cognitive and demographic factors. Temporal indices derived from the finger-tracking data (reading speed per token), the reading modality (silent vs. oral), and comprehension measures consisted of accuracy rates across question types formed the feature set for subsequent analyses.

Features were normalised to comparable scales to prevent variables with larger magnitudes from dominating subsequent multivariate analyses.

5.2. Unsupervised Clustering

To explore whether these features could reveal natural groupings corresponding to clinical profiles, we conducted a preliminary k-means clustering analysis on subject-level data from the first semester. Mean reading speed and comprehension accuracy were aggregated per subject and reading modality. The clustering revealed partial separation between clinical profiles, with CDR0 subjects (HC and SCI) largely grouped together and MCI and CDR1 subjects forming distinguishable clusters. PCA visualisation provided a clear overview of subject distributions and cluster centroids, as shown in Figure 4. The average silhouette score (0.503) indicated a moderate degree of separation among clusters, suggesting partially distinct groupings. Some clusters underrepresented certain clinical groups, likely due to the limited dataset available from the first session only. Inclusion of data from subsequent sessions, is expected to improve cluster stability and separation. These results illustrate the potential of reading behaviour features for early cognitive profiling and provide a foundation for more comprehensive multimodal clustering analyses in subsequent stages of the project.

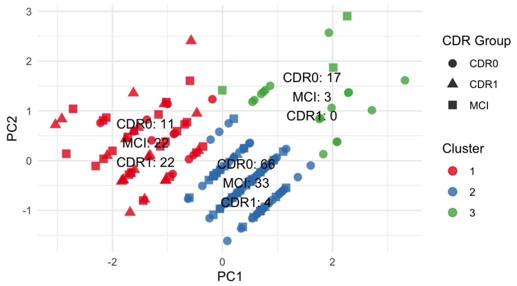

Figure 4: K-means clustering results based on subject-level reading features (mean speed per token, reading modality and comprehension accuracy). Colours indicate cluster membership and shapes represent CDR groups (CDR0 for HC and SCI, MCI, CDR1).

5.3. Supervised Classification

Building on the feature selection and clustering results, we implemented a supervised classification using an *Extreme Gradient Boosting classifier* (XGBoost) to evaluate the discriminative power of reading-derived features in predicting cognitive status. This class of boosted tree ensembles were chosen due to their ability to capture non-linear relationships and interactions between behavioural features, their robustness to small-to-moderate sample sizes, and the availability of feature importance measures to interpret the contribution of individual predictors (Bentéjac et al., 2021).

The model was trained to predict participants' clinical status (CDR0, MCI, CDR1) from subject-level reading features. The `xgboost` classifier was implemented with stratified 5-fold cross-validation, using mean reading speed (silent and aloud) and comprehension accuracy as predictors. All features were z-normalised prior to training. Model performance was stable across validation schemes, yielding comparable results when re-trained with a 10-fold cross-validation.

As shown in Table 1, the XGBoost classifier achieved moderate performance in differentiating CDR0, MCI, and CDR1 participants based on subject-level reading features. Performance was highest for the CDR1 group $(F1 = 0.80)$, reflecting the model's ability to identify participants with mild dementia, while MCI participants showed lower F1 (0.50), likely due to the subtler behavioural deviations and the limited dataset from the first semester. The macro and weighted averages indicate that overall classification was above chance (NIR $= 0.56$), highlighting the discriminative potential of reading speed and comprehension measures.

Group	Precision	Recall	F1-score	NIR
CDR0 (HC, SCI)	0.83	0.56	0.67	0.56
MCI	0.43	0.60	0.50	0.56
CDR1	0.67	1.00	0.80	0.56
Macro Avg	0.64	0.72	0.66	-
Weighted Avg	0.69	0.63	0.63	-

Table 1: Performance of the XGBoost classifier (5-fold CV) across CDR groups. Precision = Positive Predictive Value; Recall = Sensitivity; F1-score = harmonic mean of precision and recall; NIR = No Information Rate.

To simplify the classification task and focus on distinguishing healthy participants ($N = 47$) from those with any cognitive impairment, we combined MCI and CDR1 participants into a single group (CDR>0, $N = 42$). Table 2 reports the performance of the XGBoost classifier for this binary task. The model achieved balanced Precision and Recall across the two groups, with F1-scores of 0.74 for CDR0 and 0.67 for CDR>0. The overall macro and weighted averages indicate that the model substantially outperformed the No Information Rate (NIR $= 0.53$), suggesting that subject-level reading features can effectively discriminate healthy participants from those showing mild cognitive impairments or mild dementia.

Group	Precision	Recall	F1-score	NIR
CDR0 (HC, SCI)	0.70	0.78	0.74	0.53
CDR>0 (MCI, CDR1)	0.71	0.63	0.67	0.53
Macro Avg	0.71	0.70	0.70	-
Weighted Avg	0.71	0.71	0.70	-

Table 2: Performance of the XGBoost classifier (5-fold CV) for binary classification: CDR0 (HC and SCI) vs CDR>0 (MCI and CDR1). Precision = Positive Predictive Value; Recall = Sensitivity for CDR0 / Specificity for CDR>0; F1-score = harmonic mean of Precision and Recall; NIR = No Information Rate.

These results provide preliminary evidence that fine-grained reading-derived features can support early cognitive profiling, particularly for detecting prodromal stages such as MCI, and suggest that inclusion of additional data in subsequent semesters is expected to improve model stability and accuracy. Table 3 presents the relative importance of these features for both the three-class and binary classification models.

Feature	% model 1	% model 2
Token Speed (silent)	31.4	26.7
Comprehension Accuracy (aloud)	27.9	9.8
Token Speed (aloud)	27.6	46.1
Comprehension Accuracy (silent)	13.1	17.1

Table 3: Features importance in the XGBoost classifier.

The first model achieved moderate performance for the three-class task, with higher F1 for CDR1 participants and lower F1 for MCI participants, likely reflecting subtler deviations in reading behaviour and the limited dataset size.

Simplifying to a binary classification task (CDR0 vs CDR>0) yielded more balanced performance and slightly higher F1 scores, suggesting that temporal and comprehension-related features can effectively discriminate healthy participants from those with cognitive impairments and supporting the feasibility of using reading dynamics as non-invasive indicators of early cognitive decline. Only a few participants were misclassified, and errors were roughly balanced across the two binary groups (CDR0 and CDR>0). These misclassified participants tend to exhibit slower reading speeds (-0.2 on silent reading and -0.06 on aloud reading) while maintaining relatively high accuracy (0.5 and 0.4). This pattern suggest that the model relies more heavily on temporal reading measures – both silent and aloud – than on comprehension accuracy. Thus, subtle slowing in reading speed appears to be a stronger signal of early cognitive change than errors in comprehension.

To ensure stability of the classification results, we also ran a repeated 5-fold cross-validation (50 repeats). Performance metrics for each group, as well as macro- and weighted averages, are reported in Table 8 in Appendix A. These results confirm that temporal reading measures – both silent and aloud speed – contribute more strongly than comprehension accuracy to the discrimination between CDR0 and CDR>0 participants.

We expect that including additional training data from the second and third trials will further improve model performance and stability, enhancing the discriminative power of reading-derived features across cognitive profiles. Additionally, other model families and parameter configurations should be tested and evaluated – such as non-tree based approaches (e.g. multi-layer neural networks), general ensemble methods (e.g. AdaBoost), and other tree-based algorithms (e.g. Random Forests, basic Gradient Boosting) – to further validate or potentially enhance predictive performance.

6. Applications and Perspectives

The proposed dataset is designed for twofold use: (i) as a benchmark resource for linguistic and cognitive modelling of reading behaviour in typical and atypical aging, and (ii) as a foundation for assistive or screening technologies based on linguistic-behavioural data. Preliminary analyses on reading-derived features suggest that they can capture meaningful differences between participant groups, highlighting their potential for automatic classification and predictive modelling. While the obtained accuracy is lower than that typically reported in studies using neuroimaging-based models, it highlights the promise of linguistic and behavioural features as accessible, non-invasive indicators of cognitive status. Developing such scalable approaches is crucial for supporting large-scale screening and early detection of cognitive decline.

Building on this foundation, the resource is being expanded to include detailed annotations of acoustic features collected from spontaneous and semi-structured speech tasks (Albertin et al., 2024). This will enable multimodal analyses, integrating reading and speech data to investigate the interplay between linguistic, temporal, and acoustic markers of cognitive status (Belmonte et al., 2024; Badal et al., 2024; Huang et al., 2024; Hajjar et al., 2023; Gagliardi and Tamburini, 2021). Specifically, a range of voice-based digital markers – such as acoustic/prosodic, lexical, syntactic and dictionary based features (Calzà et al., 2021) – can be extracted to complement reading-derived measures and support predictive modelling of early cognitive decline, thus fostering further research on early detection of cognitive decline and the development of non-invasive assessment tools.

7. Conclusion

Our results show how fine-grained behavioural reading data can serve as a valuable language resource for investigating cognitive aging – both healthy and pathological/degenerative. The combination of linguistic annotation, temporal resolution, and comprehension data makes this dataset a unique tool for both theoretical and applied research. Additionally, the proposed comprehensive pipeline from feature selection to multivariate analysis, highlights both exploratory patterns in reading behaviour and their predictive utility for identifying cognitive profiles in aging populations.

Taken together, our analyses highlight the potential of language resources to inform accessibility initiatives, assistive technologies, and cognitive health monitoring. The methodological framework – from feature selection and preprocessing, through exploratory clustering, to supervised classification – offers a structured approach to characterising the behavioural signatures of reading performance across cognitive profiles. While the current focus is on temporal and comprehension-based measures, the framework is readily extendable to include additional modalities, such as acoustic and lexical features from speech data, paving the way for future multimodal investigations and applications in early cognitive screening.

Importantly, the ability to detect subtle, prodromal signs of cognitive decline at an early stage is critical for timely intervention. Early identification of individuals at risk of progression to mild cognitive impairment or dementia allows the implementation of cognitive stimulation programs, lifestyle modifications, and other preventive strategies, in line with recent evidence highlighting the benefits of early intervention (e.g., Fowler et al., 2025; Woods et al., 2023; Kim et al., 2017; Wenisch et al., 2007; Spector et al., 2003). Our results suggest that detailed reading behaviour, captured through temporal and comprehension measures, may serve as sensitive markers of early cognitive changes, offering a non-invasive, scalable approach for monitoring at-risk populations.

7.1. Limitations

While these results are promising, several limitations should be acknowledged. The analyses presented here are based exclusively on data from the first test session, limiting longitudinal interpretation. The sample size for some groups (particularly SCI and CDR1) is relatively small, which may affect model stability and generalisability. In particular, potential misclassifications – especially in intermediate profiles such as MCI – may partly reflect the graded and transitional nature of cognitive decline rather than purely algorithmic limitations. Additionally, only a subset of potential behavioural features was included; the integration of additional modalities such as acoustic features from speech tasks is expected to enhance predictive performance.

8. Acknowledgements

The study was financially supported by the PRIN project *ReMind: an ecological, cost-effective AI platform for early detection of prodromal stages of cognitive impairment*, 2022YKJ8FP, conducted in accordance with the Declaration of Helsinki and approved by the Italian National Research Council Committee for Research Ethics (Ethical Clearance Statement n. 0369498, 08/10/2024). All participants provided their informed consent.

9. Bibliographical References

Giorgia Albertin, Paolo Bongioanni, Cristina Dolciotti, Marcello Ferro, Gloria Gagliardi, Alessandro Lento, Claudia Marzi, Andrea Nadalini, Vito Pirrelli, and Fabio Tamburini. 2024. Early detection of prodromal stages of cognitive impairment in the Italian elderly population using NLP and AI technologies: the ReMind project. In *VII International Conference on Clinical Linguistics*.

Eider M Arenaza-Urquijo and Prashanthi Vemuri. 2018. Resistance vs resilience to Alzheimer disease: clarifying terminology for preclinical studies. *Neurology*, 90(15):695–703.

Varsha D Badal, Jenna M Reinen, Elizabeth W Twamley, Ellen E Lee, Robert P Fellows, Erhan Bilal, and Colin A Depp. 2024. Investigating acoustic and psycholinguistic predictors of cognitive impairment in older adults: modeling study. *JMIR aging*, 7(1):e54655.

Marica Belmonte, Gloria Gagliardi, Dimitrios Kokkinakis, and Fabio Tamburini. 2024. Automatic detection of rhythmic features in pathological speech of mci and dementia patients. In *Proceedings of the 5th Workshop on Resources and Processing of linguistic, para-linguistic and extra-linguistic aata from people with various forms of cognitive/psychiatric/developmental impairments@ LREC-COLING*, pages 34–44.

Candice Bentéjac, Anna Csörgő, and Gonzalo Martínez-Muñoz. 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3):1937–1967.

Deborah M Burke and Meredith A Shafto. 2004. Aging and language production. *Current directions in psychological science*, 13(1):21–24.

Roberto Cabeza, Marilyn Albert, Sylvie Belleville, Fergus IM Craik, Audrey Duarte, Cheryl L Grady, Ulman Lindenberger, Lars Nyberg, Denise C Park, Patricia A Reuter-Lorenz, et al. 2018. Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience*, 19(11):701–710.

Laura Calzà, Gloria Gagliardi, Rema Rossini Favretti, and Fabio Tamburini. 2021. Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia. *Computer Speech & Language*, 65:101113.

Davide Crepaldi, Marcello Ferro, Claudia Marzi, Andrea Nadalini, Vito Pirrelli, and Loukia Taxitari. 2022. Finger movements and eye movements during adults' silent and oral reading. In Ronit Levie, Amalia Bar-On, Orit Ashkenazi, Elitzur Dattner, and Gilad Brandes, editors, *Developing Language and Literacy: Studies in Honor of Dorit Diskin Ravid*, pages 443–471. Springer International Publishing.

Marcello Ferro, Claudia Cappa, Sara Giulivi, Claudia Marzi, Ouafae Nahli, Franco Alberto Cardillo, and Vito Pirrelli. 2018. Readlet: Reading for understanding. In *Proceedings of 5th IEEE Congress on Information Science & Technology*.

Marcello Ferro, Claudia Marzi, Andrea Nadalini, Loukia Taxitari, Alessandro Lento, and Vito Pirrelli. 2024. Readlet: a dataset for oral, visual and tactile text reading data of early and mature readers. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*.

Nicole R Fowler, Katherine A Partrick, James Taylor, Michael Hornbecker, Kevin Kelleher, Malaz Boustani, Jeffrey L Cummings, Tim MacLeod, Michelle M Mielke, Jared R Brosch, et al. 2025. Implementing early detection of cognitive impairment in primary care to improve care for older adults. *Journal of internal medicine*.

Gloria Gagliardi and Fabio Tamburini. 2021. Linguistic biomarkers for the detection of mild cognitive impairment. *Lingue e linguaggio*, 20(1):3–31.

Peter C Gordon, Matthew W Lowder, and Renske S Hoedemaker. 2016. Reading in normally aging adults. *Cognition, language and aging*, pages 165–191.

Ihab Hajjar, Maureen Okafor, Jinho D Choi, Elliot Moore, Anees Abrol, Vince D Calhoun, and Felicia C Goldstein. 2023. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(1):e12393.

Lihe Huang, Hao Yang, Yiran Che, and Jingjing Yang. 2024. Automatic speech analysis for detecting cognitive decline of older adults. *Frontiers in Public Health*, 12:1417966.

Kayoung Kim, Ji Won Han, Yoonseop So, Jiyeong Seo, You Joung Kim, Joon Hyuk Park, Seok Bum Lee, Jung Jae Lee, Hyun-Ghang Jeong, Tae Hui Kim, et al. 2017. Cognitive stimulation as a therapeutic modality for dementia: A meta-analysis. *Psychiatry investigation*, 14(5):626.

Young-Suk Grace Kim. 2015. Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading research quarterly*, 50(4):459–481.

Eesha Kokje, Simge Celik, Hans-Werner Wahl, and Christiane von Stutterheim. 2022. Can discourse processing performance serve as an early marker of Alzheimer's disease and mild cognitive impairment? A systematic review of text comprehension. *European Journal of Ageing*, 19(1):3–18.

William S Kremen, Malin Ericsson, Margaret Gatz, Ida K Karlsson, Marianne Nygaard, Nancy L Pedersen, and Matthew S Panizzon. 2025. A lifespan perspective on cognitive reserve and risk for dementia. *Alzheimer's & Dementia*, 21(6):e70176.

Changye Li, Weizhe Xu, Trevor Cohen, and Serguei Pakhomov. 2024. Useful blunders: Can automated speech recognition errors improve downstream dementia classification? *Journal of biomedical informatics*, 150:104598.

Guillaume Lio, Roberta Fadda, Giuseppe Doneddu, Jean-René Duhamel, and Angela Sirigu. 2019. Digit-tracking as a new tactile interface for visual perception analysis. *Nature Communications*, 10(1):5392.

Claudia Marzi, Andrea Nadalini, Alessandro Lento, Manu Srivastava, Alice Todesco, Vito Pirrelli, and Marcello Ferro. 2025. Oral text reading as a multisensory task. *Lingue e Linguaggio*, 24(1):141–156.

Claudia Marzi, Anna Rodella, Andrea Nadalini, Loukia Taxitari, and Vito Pirrelli. 2020. Does finger-tracking point to child reading strategies? In *Proceedings of 7th Italian Conference on Computational Linguistics*, volume 2769, Bologna.

Andrea Nadalini, Claudia Marzi, Marcello Ferro, Loukia Taxitari, Alessandro Lento, Davide Crepaldi, and Vito Pirrelli. 2023. Eye-voice and finger-voice spans in adults' oral reading of connected texts: Implications for reading research and assessment. *The Mental Lexicon*, 18(3):366–400.

Viet Chau Linh Nguyen, Thomas Perret, Valentine Fabre, Alice Gomez, and Angela Sirigu. 2024. Cost and benefit of parafoveal information during reading acquisition as revealed by finger movement patterns. *Scientific Reports*, 14(1):25127.

Peng Peng, Marcia Barnes, CuiCui Wang, Wei Wang, Shan Li, H Lee Swanson, William Dardick, and Sha Tao. 2018. A meta-analysis on the relation between reading and working memory. *Psychological bulletin*, 144(1):48.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.

Patricia A Reuter-Lorenz, Sara B Festini, and Tiffany K Jantz. 2021. Executive functions and neurocognitive aging. In *Handbook of the psychology of aging*, pages 67–81. Elsevier.

Pauline EJ Spaan. 2016. Cognitive decline in normal aging and early Alzheimer's disease: A continuous or discontinuous transition? A historical review and future research proposal. *Cogent Psychology*, 3(1):1185226.

Aimee Spector, Lene Thorgrimsen, BOB Woods, Lindsay Royan, Steve Davies, Margaret Butterworth, and Martin Orrell. 2003. Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia: randomised controlled trial. *The British Journal of Psychiatry*, 183(3):248–254.

Yaakov Stern. 2002. What is cognitive reserve? theory and research application of the reserve concept. *Journal of the international neuropsychological society*, 8(3):448–460.

Loukia Taxitari, Claudia Cappa, Marcello Ferro, Claudia Marzi, Andrea Nadalini, and Vito Pirrelli. 2021. Using mobile technology for reading assessment. In *Proceedings of 6th IEEE Congress on Information Science & Technology*.

Emilie Wenisch, Inge Cantegreil-Kallen, Jocelyne De Rotrou, Pia Garrigue, Florence Moulin, Fériel Batouche, Aurore Richard, Martha De Sant'Anna, and Anne Sophie Rigaud. 2007. Cognitive stimulation intervention for elders with mild cognitive impairment compared with normal aged subjects: preliminary results. *Aging clinical and experimental research*, 19(4):316–322.

Robert S Wilson, Tianhao Wang, Lei Yu, David A Bennett, and Patricia A Boyle. 2020. Normative cognitive decline in old age. *Annals of neurology*, 87(6):816–829.

Bob Woods, Harleen Kaur Rai, Emma Elliott, Elisa Aguirre, Martin Orrell, and Aimee Spector. 2023. Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane database of systematic reviews*, (1).

Jiaqi Zhang, Kayleigh L Warrington, Lin Li, Ascensión Pagán, Kevin B Paterson, Sarah J White, and Victoria A McGowan. 2022. Are older adults more risky readers? evidence from meta-analysis. *Psychology and Aging*, 37(2):239.

Appendix A: Model Details

Here we report detailed information on model specifications used in this study.

Comparison	Test statistic	Adjusted p-value	Significant
HC vs CDR1	-80.14	0.0000	*
MCI vs HC	4.63	0.0000	*
SCI vs MCI	-47.08	0.0000	*
SCI vs HC	-21.54	0.0000	*
SCI vs CDR1	-78.25	0.0000	*
MCI vs CDR1	-48.06	0.0000	*

Table 4: Post-hoc pairwise comparisons at the token level of the accuracy on comprehension questions across participant groups using *Dunn's test* with *Bonferroni* correction. Significant differences (p $\leq$ 0.05) are marked with an asterisk (*).

Comparison	Test statistic	Adjusted p-value	Significant
HC vs CDR1	-4.529	0.0000	*
MCI vs HC	7.073	0.0000	*
SCI vs MCI	-3.304	0.0057	*
SCI vs HC	1.876	0.3638	
MCI vs CDR1	0.429	1.0000	

Table 5: Post-hoc pairwise comparisons of the silent-aloud speed ratio across participant groups using *Dunn's test* with *Bonferroni* correction. Significant differences (p $\leq$ 0.05) are marked with an asterisk (*).

Effect	Estimate	Ref.df	t/F	p-value
Intercept (HC)	16.85	3.35	5.03	<.001
SCI	-1.62	3.12	-0.52	>.05
MCI	-4.56	2.16	-2.11	<.05
CDR1	-9.70	2.79	-3.48	<.001
cognitive reserve	0.95	0.21	4.45	<.001
Participants	82.69	84	64.9	<2e-16
R^2				0.32
explained deviance				32.4%

Table 6: Summary of a GAM fitting reading speed as a function of cognitive reserve (as grade level) across participant groups in silent reading. Parametric rows report coefficients, standard errors, and t-values; participant random effects report estimated degrees of freedom (edf), reference degrees of freedom, and F-values.

Effect	Estimate	Ref.df	t/F	p-value
Intercept (HC)	17.23	2.96	5.382	<.001
SCI	0.13	2.75	0.05	>.05
MCI	-1.96	1.91	-1.03	<.05
CDR1	-6.08	2.46	-2.47	<.05
cognitive reserve	0.41	0.19	2.17	<.05
Participants	82.47	84	23.55	<2e-16
R^2				0.14
explained deviance				14.8%

Table 7: Summary of a GAM fitting reading speed as a function of cognitive reserve (as grade level) across participant groups in aloud reading. Parametric rows report coefficients, standard errors, and t-values; participant random effects report estimated degrees of freedom (edf), reference degrees of freedom, and F-values.

Group	Precision	Recall	F1-score	NIR
CDR0 (HC, SCI)	0.92	0.79	0.85	0.40
CDR>0 (MCI, CDR1)	0.62	0.83	0.71	0.60
Macro Avg	0.77	0.81	0.78	-
Weighted Avg	0.74	0.81	0.77	-
variable importance				
	speed aloud	100.00		
	speed silent	93.79		
	accuracy	0.00		

Table 8: Performance of the XGBoost classifier using repeated 5-fold cross-validation (50 repeats) for binary classification: CDR0 (HC and SCI) vs CDR>0 (MCI and CDR1). Precision = Positive Predictive Value; Recall = Sensitivity for CDR0 / Specificity for CDR>0; F1-score = harmonic mean of Precision and Recall; NIR = No Information Rate.

Appendix B: Resource Description

The dataset described in this paper is publicly available on Zenodo: https://doi.org/10.5281/zenodo.18927776.

All data are anonymised. Access complies with institutional ethical approval guidelines (n. 0369498, 08/10/2024).

The dataset provides token-level reading time measures derived from finger-tracking (silent and aloud sessions) and audio recordings (aloud sessions). Only processed tabular data are shared; raw audio files are not included in order to comply with privacy and data protection regulations.

Post-processed finger-tracking data are organised at the token level.

Tracking duration (*dt*) indexes the total time a word token was tracked, including possible regressive movements. The tracking onset time (*t*) indicates the time point at which the tracking of a word begins (in seconds from the start of the reading session).

Table 9 summarises the structure of the post-processed finger-tracking dataset organised by word tokens and reading sessions. Each row corresponds to a single word token tracked by a participant during a reading session.

heading	type	gloss
idSession	*int*	reading session id
tid	*num*	token id
token	*factor*	word token string
dt	*num*	total tracking duration
t	*num*	tracking onset time
len	*int*	word token length
freq	*int*	word token frequency
idUser	*int*	participant id
years	*int*	participant age
group	*factor*	participant CDR group
readingType	*factor*	reading mode (aloud or silent)
questAccuracy	*num*	comprehension question accuracy

Table 9: Column headings for the word-token level.

To protect participant privacy, original session and participant identifiers were anonymised by replacing them with sequential numeric indexes. The mapping to the original identifiers cannot be recovered.