ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10890866

Cognitive Decline Detection using DLB Extraction Pipelines

Shibingfeng Zhang
FICLIT
University of Bologna
Bologna, Italy
shibingfeng.zhang@unibo.it

Nadia Khlif
ILC
CNR
Pisa, Italy
nadia.khlif@ilc.cnr.it

Marcello Ferro
ILC
CNR
Pisa, Italy
marcello.ferro@cnr.it

Gloria Gagliardi
FICLIT
University of Bologna
Bologna, Italy
gloria.gagliardi@unibo.it

Fabio Tamburini
FICLIT
University of Bologna
Bologna, Italy
fabio.tamburini@unibo.it

***Abstract*—The Prediction and Recognition of Cognitive Decline through Spontaneous Speech (PROCESS) Signal Processing Grand Challenge focuses on detecting dementia by analyzing spontaneous speech production. The challenge proposes a classification task to distinguish between subjects categorized as healthy controls, mild cognitive impairment, and dementia. Our team tackled this task by leveraging Digital Linguistic Biomarkers (DLBs) extracted from speech. Our system outperformed over 100 competing systems, earning us first place in the classification task.**

***Index Terms*—digital linguistic biomarker, cognitive decline, speech signal processing.**

I. Introduction

The Prediction and Recognition of Cognitive Decline through Spontaneous Speech (PROCESS) Signal Processing Grand Challenge [1] focuses on detecting dementia through spontaneous speech processing. The task classifies subjects into three categories: Healthy Control (HC), Mild Cognitive Impairment (MCI), and Dementia (DEM). Each subject provides three spontaneous speech samples: describing a picture (Cookie Theft Description, CTD), listing words starting with the sound 'P' (Phonemic Fluency Task, PFT), and listing animal names (Semantic Fluency Task, SFT).

We tackled this classification task using Digital Linguistic Biomarkers (DLBs) extracted through a pipeline designed for cognitive decline detection [2], [3]. DLBs refer to linguistic features extracted from individuals' verbal productions that act as indicators of their medical state. The DLBs were used in a cascade classifier combining a Random Forest classifier and a Multi-Layer Perceptron classifier. Our system achieved first place in the competition, achieving an F1-score of 69.6%, outperforming the second-place competitor by about 5 points.

II. DLB Pipelines for Feature Extraction

A newly built DLB pipeline (v2.0), based on our previous work [2], 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 undergoes text transcriptions [4], voice activity detection [5], voiced segment identification, vowel-consonant distinction [6], dependency [7] and constituency [8] parsing. In the feature extraction phase, the DLBs listed in Table I are derived [9] using the information obtained during preprocessing. These DLBs are categorized into five groups: Acoustic, Rhythmic, Lexical, LIWC based counts, and Syntactic DLBs, which offer a fine-grained representation of the linguistic patterns related to cognitive impairment [2]. It is worth noticing that, for PFT and SFT, lexical and syntactic DLBs are not extracted since these tasks involve word listing with no meaningful syntactic structure or lexical information. The number of correct words listed in tasks PFT and SFT is also included as an additional feature (CHA) for the PROCESS challenge. Figure 1 shows the structure of the DLBs extraction pipeline.

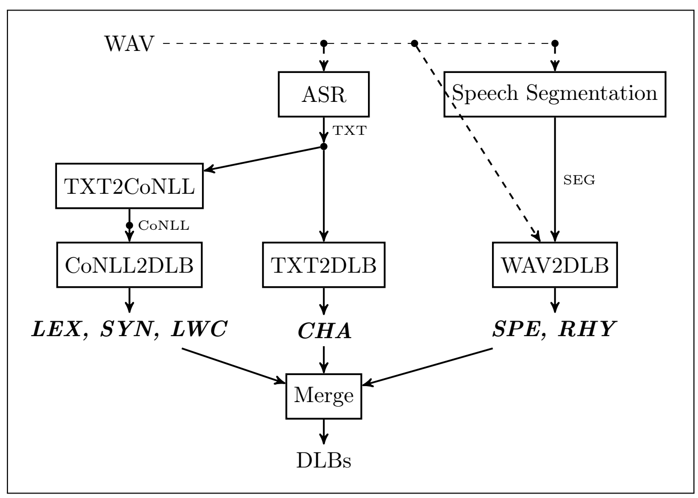

Fig. 1. The Structure of DLBs Extraction Pipeline [2], [3].

III. Tools for Classification

The model is designed to classify subjects into three categories: HC, MCI, or DEM, based on task-specific features and a two-stage classification process.

The first classification stage employs a Random Forest Classifier to perform binary classification, distinguishing between

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

Nadia Khlif is a PhD student in the *Computer Science Research Laboratory*, Faculty of Sciences, at the University Mohammed First of Oujda, Morocco.

Acoustic Features (SPE)
- Silence segments duration (M, MD, SD)
- Speech segments duration (M, MD, SD)
- Temporal regularity of voiced segments
- Verbal rate
- Transformed phonation rate
- Standardized phonation time
- Standardized pause rate
- Root mean square energy (M, SD)
- Pitch (M, SD)
- Spectral centroid (M, SD)
- Higuchi fractal dimension (M, SD)
Rhythmic Features (RHY)
- Percentage of vocalic intervals
- Vocalic, ΔV, and consonantal, ΔC, interval durations (SD)
- Pairwise variability index, raw, rPVI, and normalized, nPVI
- Variation coefficient for ΔV and ΔC
Lexical Features (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
Correct Word counts for Challenge (CHA)
- Number of correct words listed in Semantic Fluency audio
- Number of correct words listed in Phonemic Fluency audio
Linguistic Inquiry and Word Count Features (LWC)
- Language metrics (e.g., words per sentence, words $>$ 6 letters)
- Function words (e.g., pronouns, articles, auxiliary verbs)
- Affect words (e.g., positive/negative emotion)
- Cognitive processes (e.g., insight, certainty, tentativeness)
- Perceptual processes (e.g., seeing, hearing, feeling)
- Biological processes (e.g., body, health/illness, ingesting)
- Personal concerns (e.g., work, leisure, money, religion, death)
- Social words (e.g., family, friends)
- Punctuation (e.g., periods, commas, colons, question marks)
Syntactic Features (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 I
THE LIST OF DLBS EXTRACTED BY THE PIPELINE (MEANS (M), MEDIANS (MD), AND STD. DEVS (SD)). PLEASE REFER TO [9] FOR DETAILS.

HC and Non-HC subjects. If the prediction is HC, it is directly finalized. For Non-HC predictions, the model proceeds to the second stage, where a Multi-layer Perceptron classifier performs a binary classification to differentiate between MCI and DEM. This hierarchical organization was introduced to address the class imbalance in the challenge development dataset, which has limited representation of the DEM class.

For each subject, the DLBs extracted from the CTD, SFT, and SPT audio files are concatenated and used as input features for the classifiers. The stage 1 classifier utilizes all DLBs listed in Table I, while the stage 2 classifier utilizes all DLBs except lexical features. To address the limited number of samples in the non-HC classes, the features for stage 2 are reshaped into a 4-dimensional space using PCA.

Leave-One-Subject-Out Cross-Validation is used to evaluate the proposed classification system on the development set, as the labels of test set are not available to participants of the PROCESS challenge. Figure 2 illustrates the confusion matrix obtained from this evaluation. As shown, the proposed system performs well in distinguishing HC from Non-HC subjects, achieving an F1 of 0.727 for class HC. However, it struggles to categorize non-HC subjects into MCI and DEM, with an F1, respectively, of 0.586 and 0.242.

		Predicted		
		HC	MCI	DEM
True	HC	60	15	7
	MCI	19	34	6
	DEM	4	8	4

Fig. 2. Confusion Matrix of proposed two-stage classification using DLBs

The overall performance of our system, when evaluated on the test set, reached an F1 of 69.6%. Despite being the top-performing system among over 100 competitor systems, there remains significant room for improvement and further exploration.

IV. CONCLUSIONS

Our study highlights the potential of digital linguistic biomarkers (DLBs) derived from spontaneous speech production as powerful tools for detecting cognitive impairment.

Future work will focus on enhancing the DLB extraction pipeline, integrating a broader range of linguistic and acoustic features, and conducting comprehensive ablation studies to assess the contribution and impact of each feature group on model performance.

REFERENCES

[1] F. Tao, B. Mirheidari, M. Pahar, S. Young, Y. Xiao, H. Elghazaly, F. Peters, C. Illingworth, D. Braun, R. O'Malley *et al.*, "Early dementia detection using multiple spontaneous speech prompts: The process challenge," *arXiv preprint arXiv:2412.15230*, 2024.

[2] G. Gagliardi and F. Tamburini, "The automatic extraction of linguistic biomarkers as a viable solution for the early diagnosis of mental disorders," in *Proc. LREC2020*, Marseille, France, Jun. 2022, pp. 5234–5242.

[3] S. Zhang, G. Gagliardi, and F. Tamburini, "Voice activity detection on Italian language," in *Proc. CLiC-it*, Pisa, Italy, 2024.

[4] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *PMLR*, 2023, pp. 28 492–28 518.

[5] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin *et al.*, "Pyannote. audio: neural building blocks for speaker diarization," in *Proc. ICASSP*, 2020, pp. 7124–7128.

[6] X. Li, S. Dalmia, J. Li, M. Lee, P. Littell, J. Yao, A. Anastasopoulos *et al.*, "Universal phone recognition with a multilingual allophone system," in *Proc. ICASSP*, 2020, pp. 8249–8253.

[7] D. Kondratyuk and M. Straka, "75 languages, 1 model: Parsing universal dependencies universally," *arXiv preprint arXiv:1904.02099*, 2019.

[8] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages," in *Proc. ACL2020: System Demo*, 2020.

[9] L. Calzà, G. Gagliardi, R. Rossini Favretti, and F. Tamburini, "Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and dementia," *Comp. Speech & Lang.*, vol. 65, pp. 101–113, 2021.