IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 July 2024, accepted 29 July 2024, date of publication 12 August 2024, date of current version 11 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3442431

RESEARCH ARTICLE

Analysis of Voice Biomarkers for the Detection of Cognitive Impairment

MOISÉS R. PACHECO-LORENZO[1], HEIDI CHRISTENSEN[2], (Member, IEEE), LUIS E. ANIDO-RIFÓN[1], (Senior Member, IEEE), MANUEL J. FERNÁNDEZ-IGLESIAS[1], AND SONIA M. VALLADARES-RODRÍGUEZ[3]

[1]atlanTTic, Universidade de Vigo, 36310 Vigo, Spain
[2]Department of Computer Science, The University of Sheffield, S1 4DP Sheffield, U.K.
[3]Department of Electronics and Computing, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

Corresponding author: Moisés R. Pacheco-Lorenzo (mlorenzo@det.uvigo.es)

This work was supported in part by the Ministry of Science, Innovation and Universities (Formación de Profesorado Universitario) under Grant FPU19/01981 and Grant EST21/00739; and in part by the Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI)/10.13039/501100011033 under Grant PID2020-115137RB-I00.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of Pontevedra-Vigo-Ourense from the Galician Healthcare Service (SERGAS), Spain, under registration code 2021/213.

ABSTRACT The objective of this work is to determine whether speech obtained from interactions with a smart speaker can be used to predict the level of cognitive impairment (CI). We use a voice assistant to administer a cognitive test in Spanish, and we record the conversations in order to extract features that could potentially be used as voice biomarkers. A total of 21 participants (14 patients and 7 healthy controls) between the ages of 68 and 86 are included in the study (15 were women). Using just speech we are able to perform a regression with machine learning models, in order to predict the Global Deterioration Scale (GDS) of cognitive functions. Then, we measure the performance of the estimations with standard metrics - an R^2 of 0.74 was obtained in the best case using Support Vector Machine (SVM) algorithms. Despite needing a bigger sample of participants in future studies, this is a positive and promising result for such a non-intrusive procedure, which could potentially be used as a screening tool for automatic cognitive impairment assessment.

INDEX TERMS Biomarkers, cognitive impairment, dementia, regression, voice.

I. INTRODUCTION

Dementia currently affects more than 50 million people worldwide and it is estimated that this figure will triple by 2050 [1]. These numbers highlight the high prevalence of this medical condition and the major burden on the healthcare system that it represents. Specifically, Alzheimer's Disease (AD) is the most common type of dementia and may contribute to 60-80% of cases [2], [3]. Throughout, the World Health Organization (WHO) has recognized dementia as a public health priority and made a global appeal to fight it [4]. Additionally, there are three main risk factors (i.e., age, genetics and family history [5]) of which age is the greatest one. In this sense, the aging of the baby boom generation will significantly increase the number of people with Alzheimer's in developed countries. The early stages of dementias, particularly AD, involve an initial deterioration of cognitive functions (e.g., memory, reasoning, attention, language) known as Mild Cognitive Impairment (MCI). MCI is the prodromal and transitional phase between healthy aging and dementia; it is characterized by subtle cognitive deficits that do not meet the criteria for the diagnosis of a major neurocognitive disorder [6]. The early detection and diagnosis of MCI is vital for patient's well-being and facilitates a better intervention and treatment.

The current diagnostic methods for detecting dementia are the neuropsychological tests [7], [8]. These tools perform a cognitive evaluation of an individual in certain cognitive

The associate editor coordinating the review of this manuscript and approving it for publication was Kaustubh Raosaheb Patil.

domains (e.g., memory, language, attention, visuospatial abilities, etc.). The main limitations of these tests are that they are generally perceived by users as intrusive and extraneous tools [9]; they provide a delayed diagnosis [10]; they have a notorious lack of ecological validity [11], [12]; they depend on confounding variables (e.g., age, educational level [13], practice effect -improvement that results from practice or repetition of the test- [7], [14]), and they require manual data processing (e.g., acquisition and analysis).

For all these reasons, alternative cognitive assessment mechanisms were explored in recent years, including the straightforward digitization of classic tests [15], gamification [16], virtual reality [17], [18], [19], [20], [21], [22], [23], [24], as well as the identification of biomarkers for the early detection of cognitive impairment [25], [26], [27], [28]. This research is focused on the study of speech alterations, as they are one of earliest signs of cognitive decline [29], and they are part of the diagnosis in current clinical practice [30]. These alterations include variations in the percentage and number of voice breaks, shimmer (an increase in shimmer indicates greater variability in voice amplitude, which can result in a harsh or shaky voice) and noise-to-harmonics ratio [31]. Increasing evidence suggests that spoken language could be used as a powerful resource to automatically detect dementia at an early stage [30]. Some studies report the correct discrimination of people affected by MCI from people without cognitive impairment with accuracies as high as 90% [32], [33], which has raised the general interested in these automated methods.

According to a systematic review from 2021 [34], the use of smart conversational agents for the detection of neuropsychiatric disorders is ''an emerging and promising field of research, with a broad coverage of mental disorders and extended use of artificial intelligence techniques''. Here, it was concluded that additional research was needed in this field based on studies with more robust experimental designs. Furthermore, smart virtual assistants were considered a technology trend in 2019, following the trail of chatbots [15], and even replacing them as the preferred solution for human-machine interaction. Finally, some conceptual frameworks have been proposed for the use of smart speakers and conversational agents in order to carry out early detection of cognitive impairment [35], [36], [37], [38].

For all these reasons, smart assistants are of great interest for the administration of standardised classical tests, such as the Mini-Mental State Examination (MMSE) for dementia [16] or the Global Deterioration Scale (GDS) [39]. These tests provide, as a result, an evaluation of the cognitive or mood-related state of the patient, and require an interpretation by a trained medical practitioner (e.g., neurologists, psychiatrists).

Recently, work in the speech community has been addressing CI in detail. The virtual conference INTERSPEECH 2020 [40] proposed the Alzheimer's Dementia Recognition through Spontaneous Speech (ADRESS) challenge. This challenge proposes a classification task for the comparison of different approaches to the automated recognition of Alzheimer's dementia based on spontaneous speech. Machine learning methods provided a baseline accuracy of up to 62.5% with automatically extracted voice features (76.85% with linguistic features extracted from manually produced transcripts) [40]. In 2021 INTERSPEECH's challenge the shared task (i.e., ADRESSo) was focused on speech only [41], [42], meaning raw, non-annotated and non-transcribed speech. This was the first shared task to target cognitive status prediction using data in this format. The baseline accuracy of the task was 78.87% for the AD classification task, and a Root-Mean-Square-Error (RMSE) of 5.28 for prediction of cognitive scores.

Additionally, there has been recent work involving voice assistant data collections. A study [43] from 2022 extracted and analysed features from voice commands given to a voice assistant system (Amazon Alexa). Their machine learning classification models yielded an accuracy of 68% in distinguishing between healthy controls and MCI participants, and a RMSE of 3.53 when compared to standard cognitive assessment scores. Features of overall performance, music-related commands, call-related commands, and features from Automatic Speech Recognition (ASR) were the top-four feature sets most impactful on inference accuracy. Furthermore, in the last years, cross-lingual approaches have grown in popularity in the machine learning field [44], reaching an F-score of 0.80 for AD discrimination using acoustic and word embeddings in Spanish. Their feature extraction part, using Wav2Vec 2.0 [45], was composed of several temporal convolutions which converted the speech input into a latent space representation.

Thus, we can conclude from existing literature that voice might be a relevant biomarker to non-invasively identify early cognitive impairment. Based on a scoping review [46], there is already a variety in the existing literature of studies in terms of population of interest and methodology for automatic assessment. However, the language used is still almost exclusively English, and the speech is usually obtained from prompted conversations with other humans. This brings attention to the possibility of exploring different languages (in this case, Spanish) and different types of conversations, particularly between a human and a virtual assistant. Furthermore, it is worth researching to what extend we can predict the level of cognitive deterioration, particularly the GDS score, using just speech. Consequently, in this study we aim to answer the following research question:

Can speech obtained from interactions with a smart speaker be used to predict the level of cognitive impairment?

Section II delineates the methodology employed in this study, encompassing various key areas: sample selection, data collection procedures, analytical instruments utilized, and the subsequent data analysis conducted. Section III

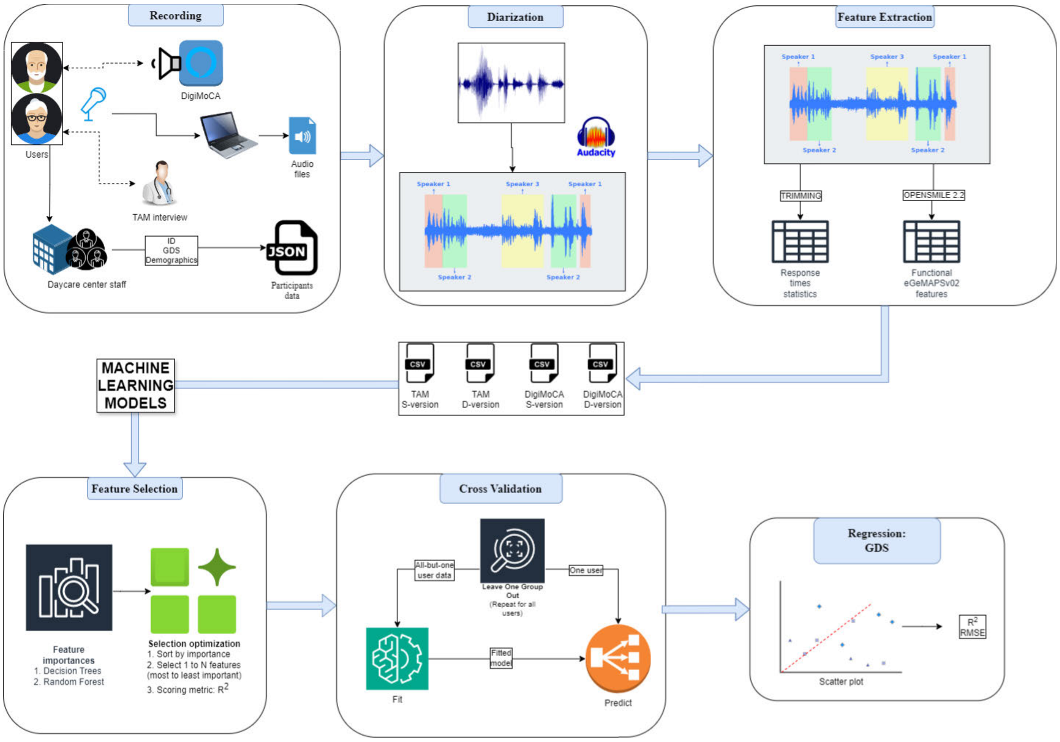

FIGURE 1. Data processing flowchart. This diagram depicts how the collected data enters all phases detailed in Sec. II.

presents and discusses the results of a pilot study involving a cohort study comprised of 21 seniors, concerning the potential of speech alterations as an appropriate cognitive biomarker. Finally, Section IV enumerates some concluding remarks.

II. METHODS

In the following paragraphs we describe the methodology undertook in order to collect, process and analyze the data used in this study. An overview flowchart on the entire process is depicted in Fig. 1. Data analyses were performed in Python 3.10, and the code generated will be made available upon reasonable request.

A. SAMPLE DESCRIPTION

In this pilot study a total of 21 participants were involved, with a mean age of 77.19 ± 4.88 years and a mean GDS value of 2.29 ± 1.24. The number of Healthy Controls (HC), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) participants is 7 for all groups and, as expected, participants with certain levels of cognitive impairment tend to be older. The clinical characteristics of them are shown in Table 1. A total of 84 recordings were captured during the sessions, 4 for each participant. For each recorded audio of the DigiMoCA administration, there are on average 74.10 ± 9.96 segments of participant speech after diarization. For each recording of the TAM questionnaire administration, there are 26.67 ± 8.58.

All the participants in this pilot study were recruited from the Spain-based Association of Relatives of Alzheimer's Patients (AFAGA). The inclusion criterion was to be a literate senior adult over the age of 65. Exclusion criteria, on the other hand, were (i) being in an advanced dementia state, as the main goal of the study is the early detection of cognitive impairment; (ii) having hearing or speech disability, since effective communication with a smart speaker is the main subject of analysis; and (iii) having aversion to technology, as this might cause disturbances in usability data.

The recruitment was carried out by AFAGA's social workers during the last week of August 2021 and the first week of September 2021, and all sessions took place during the remaining weeks of September 2021. This pilot study was officially approved by the Research Ethics Committee of Pontevedra-Vigo-Ourense from the Galician Healthcare Service (SERGAS), Spain. This is stated in the corresponding dictum with registration code 2021/213, signed on 30/06/2021.

Additionally, all the participants were given a patient information sheet and signed, written consent was requested to participate. Participants were classified according to existing Global Deterioration Scale (GDS) information, as GDS stage labels have been used as the golden standard in this study. GDS classifies individuals into seven stages. Stages 1-3 can be defined as pre-dementia stages: stage 1 meaning no cognitive decline; stage 2 means Subjective Cognitive Decline (SCD), consituting an age-associated memory impairment; and stage 3 represents MCI [47].

TABLE 1. Clinical characteristics of participants. Values are in (mean ± SD) or *n*.

Variable*	**HC** (n = 7)	**MCI** (n = 7)	**AD** (n = 7)	**Total**
Age (years)	73.86 ± 4.58	78.86 ± 4.22	78.86 ± 3.98	77.19 ± 4.88
Women	7	7	1	15
GDS	1.00 ± 0.00	2.00 ± 0.00	3.86 ± 0.64	2.29 ± 1.24
Segments per rec. (DigiMoCA)	68.14 ± 6.14	76.21 ± 4.81	77.93 ± 13.90	74.10 ± 9.96
Segments per rec. (TAM)	25.57 ± 5.09	24.79 ± 7.63	29.64 ± 11.55	26.67 ± 8.58

Stages 4-7 are considered dementia stages, and from stage 5 and onwards, an individual can no longer survive without assistance. Caregivers can get a rough idea of where an individual is at in the disease process by observing that individual's behavioral characteristics and comparing them to the GDS.

B. RECORDING

For each participant, 2 sessions were recorded, generating 2 audio files for each session: one capturing the administration of the DigiMoCA – a digital implementation of the Montreal Cognitive Assessment test [48] – by means of an Amazon's Echo Dot smart speaker, meaning the speaker's built-in conversational agent administered the cognitive test; and another session consisting on the administration of a questionnaire based on the Technology Acceptance Model (TAM) [49]. This makes a total of 4 WAV audio files per participant.

The hardware used for recording was a Huawei MateBook 14 laptop computer and an Audio-Technica ATR2500x-USB microphone. As a recording software, we used the open-source Audacity editor, with a bit depth (i.e., number of bits per sample) of 24 and a sampling rate of 192 kHz - the maximum parameters supported by the hardware used.

Ambient noise was practically non-existent, as the sessions took place in a silent room, and the microphone was situated as close to the participant as possible. Due to privacy protection, raw audio files will not be openly available.

C. DIARIZATION

Once the audio recordings are stored, the first step for pre-processing them is diarization. Speaker diarization is the task of establishing *who spoke when* in an audio recording that contains an unspecified quantity of speech and, sometimes, participants [50]. Since both the test administrator, that is, either the smart speaker or the researcher, and the participant were recorded, we need to separate the parts of the audio where each of them speaks, so that we only analyze the speech produced by the participant under test. There are several algorithms and software tools available for this purpose, but unfortunately they lack the required precision for this scenario. This is due to two reasons: firstly, the nature of the tests administered cause a fast-paced conversation where it is difficult for a computer to determine exactly when a participant starts and stops talking. Secondly, as discussed in Section III, response times (i.e., time between the ending of a test question and the start of the corresponding response) are fairly important metrics for the analysis, and therefore it is essential that they are measured with precision. As a consequence, we chose to perform the diarization process manually, again making use of the Audacity software.

D. FEATURE EXTRACTION

Feature extraction is the most important part of a speech analysis system. It is common practice to first process the speech into window frames of a few ms (typically 20 ms). Then, a Discrete Fourier Transform (DFT) is applied, in order to convert the audio samples into the frequency domain. This is usually very informative, and a set of mathematical and statistic properties can be extracted from it, known as "Low-Level Descriptors" (LDDs). Furthermore, we can then combine these and extract general –and more insightful– metrics for all the frames, instead of individually (these are called "Functionals"). There are additional non-DFT based techniques in speech processing, such as Mel Frequency Cepstral Coefficients (MFCC) or Discrete Wavelet Transforms (DWT) [51], which were also used for this study, together with the Functional parameters. The following paragraphs describe the process applied in order to obtain the desired features.

Using the timestamp labels produced by the diarization process, the first features extracted are response time statistics. Response times indicate the amount of time that passes between the moment when the participant/interviewer stops talking and the participant starts answering. Making use of the `stats.describe()` function from Python's *Scipy* library,[1] we obtain the following statistics of the response times: (i) number of observations (i.e., number of times the participant speaks); (ii) shortest and longest response time of the whole recording; (iii) average value and variance; (iv) skewness and kurtosis of the observed probability distribution of the response times. This forms a dataframe with these 7 features, the participant's ID and their GDS level.

Next, we need to trim the audio files. This can be done in two ways: we can either "select" (from now on, denoted the "S" version) only the segments where the participant speaks, or we can "discard" (from now on, the "D" version) only the segments where the participant/interviewer talks. The difference is that in the second case, we are not discarding the parts where nobody speaks, i.e., the silences. In order to

[1] https://scipy.org/

trim the WAV files, we first read them by making use of the *Scipy* library; then, we convert it into a NumPy[2] array; next, we iterate over each timestamp, discarding or selecting it depending on the mode we are trimming; and finally, we store the resulting array of samples using the same library.

The remaining features are extracted directly from the audio recordings, once trimmed. For this, the open source toolkit openSMILE[3] version 2.2 was used. openSMILE is a popular toolkit for feature extraction and speech classification [52]. For each of the 8 WAV files generated per participant (4 with S version and 4 with D), we obtain a dataframe by using the `process_file()` function from the `opensmile.Smile` class. When instantiating this class, we can establish a few parameters, including the feature set and the feature level, as well as logging and multithreading parameters. The feature set is the collection of features to extract, and we selected the *extended Geneva Minimalistic Acoustic Parameter Set version 02 (eGeMAPSv02)* for being more minimalistic and providing a common baseline for the related research [53]. The feature level, on the other hand, indicates the level of abstraction at which the features are extracted; essentially, we can choose between Low-Level Descriptors (LLDs) calculated over a sliding window, or Functionals, which are statistical values computed from variable series of LLDs. The *eGeMAPSv02* set contains 88 functional parameters or features, grouped into 3 categories: frequency related, energy/amplitude related, and spectral parameters.

This leaves us with 2 dataframes that we merge based on ID; the resulting dataframe's dimension is 84 × 97, and we have one for each of the S and D versions.

E. FEATURE SELECTION

Once the main dataframes are created, the next step is to perform a selection of the most important features. This is common practice in the ML field, where typically a subset of the most important features yield a better and faster result than the whole feature set does.

In order to make the selection, we ordered the features from most to least important, and then selected the subset of the most important ones that gives the best possible result. For measuring the performance of the result (and hence how to order the features), we used two metrics: R^2 and $RMSE$. R^2 is a statistical value called the coefficient of determination, and quantifies how much the dependent variable is determined by the independent variables [54]. In its general form, it is calculated using the following formula, where *SSres* is the residual sum of squares and *SStot* is the total sum of squares.

$$R^2 = 1 - \frac{SSres}{SStot} = 1 - \frac{\sum(y_i - f_i)^2}{\sum(y_i - \bar{y})^2}$$

The Root-Mean-Square-Error (RMSE) is a statistical measure of the difference between an original list of values and a predicted one, and it is calculated as the square root of the average of the residual squares (M is the sample size):

$$RMSE = \sqrt{\frac{SSres}{M}} = \sqrt{\frac{\sum(y_i - f_i)^2}{M}}$$

We computed the R^2 value from ML regression with cross-validation, for the top N most important features, varying N from 1 to 97 (number of total features). This means that, in order to obtain the importance of each feature, we need to already perform a prediction of the target variable (i.e., GDS in this case). This was carried out following the process described in Sec. II-F.

The criterion followed for calculating the feature importances was directly obtained from the *feature_importances_* property of the ensemble methods, which is a measure of the usefulness of a feature in terms of classification. The sum of all importances is 1, and therefore the importance of each feature can be interpreted as the percentage of contribution of the feature to the prediction (the higher the better).

The ensemble algorithms used for obtaining the feature importances were Random Forest Regressor (RFR) and Decision Tree Regressor (DTR) from the `sklearn.ensemble` ML Python library.

Additionally, p-values for each feature were obtained by performing a T-test between them and the GDS (target variable), in order to determine wether there is a statistically significant difference. This was performed using the `ttest_ind()` function from the `scipy.stats` module.

F. REGRESSION AND CROSS-VALIDATION

In order to predict the GDS level from the speech features, we utilize the Sklearn[4] library. From the resulting dataframe of the feature extraction process, we extract the GDS column as the target feature, and discard the ID. Then, we pre-process the data by applying a scaling function that transforms it and normalizes it to a range between 0 and 1.

In order to measure the performance of each algorithm, we employ cross-validation. This is a commonly used resampling technique, where, instead of training a model with the whole dataset we have available, we leave out a portion of it, which we use as new unknown data later to test the performance of the model [55]. In this case, since we have a small dataset, we perform a Leave-One-Group-Out (LOGO) type of cross-validation. Instead of training the model just once, we split the dataframe into several training and testing subsets. Specifically, we train once for each participant - excluding their own data - and then run the regression with that particular participant's data; we repeat the process for all individuals, until we have a prediction of GDS for each one of them separately. (This technique is equivalent to having as many *folds* as participants, i.e., 21 in this case).

It should be noted that this method differentiates from a regular Leave-One-Out (LOO) type of cross-validation since

[2] https://numpy.org/

[3] https://www.audeering.com/research/opensmile/

[4] https://scikit-learn.org/stable/

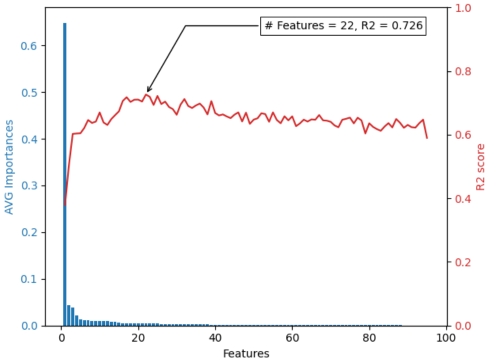

FIGURE 2. RFR feature importances for the DigiMoCA dataset S-version.

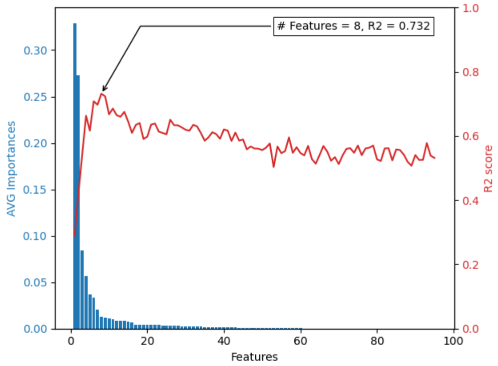

FIGURE 3. DTR feature importances of TAM dataset S-version.

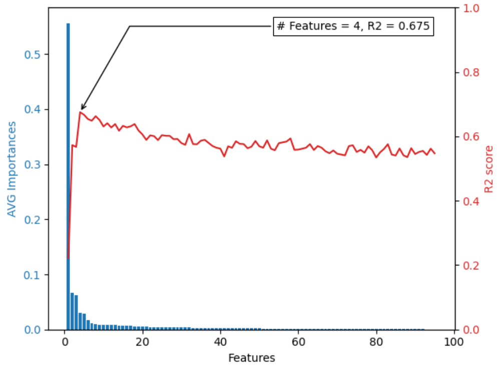

FIGURE 4. RFR feature importances of combined dataset S-version.

each participant has more than one recording, and therefore a group -hence the name- of more than one row in the main dataframe.

III. RESULTS AND DISCUSSION

In this section we present and discuss the results of the pilot study, concerning the potential of speech features as an appropriate cognitive biomarker.

When it comes to the characterization of the participants, from Table 1 we can generally notice the slight increase of segments per recordings (of both DigiMoCA and TAM types) with the GDS. This means that the higher the cognitive impairment suffered, the more speech fragments a person needs in order to complete the same questionnaire. This is a reasonable result as the average speaking time increases with cognitive impairment, which in turn is known to decrease speech fluency and quality, and therefore provokes the need for more speaking turns (as well as breaks) to transmit the same information, as it was shown in [56]. We can also notice that the difference between HC and MCI is much lower than the one between MCI and AD, which reinforces the difficulty of an early detection of MCI.

We can see the results of the feature selection process in Figs. 2, 3 and 4 for the DigiMoCA, TAM and combined datasets respectively. The first contains the features of the recordings from the 2 administrations of DigiMoCA per participant (42 × 97 dataframe); equivalently, the second contains the features of the recordings from the 2 TAM questionnaires administered (also 42 × 97 dataframe). The combined dataframe contains the features from all the recordings, resulting in a size of 84 rows x 97 columns. In all three figures we can see in sorted blue bars the average feature importances given by the used algorithm: Random Forest Regressor (RFR) or Decision Tree Regressor (DTR). We used these two ensemble algorithms since neither of the other families of algorithms offer a feature importance metric.

The average of the feature importances is computed over a leave-one-group-out (LOGO) 21-fold cross-validation. We can also see in red the best R^2 score achieved for the regression of GDS level, again with a LOGO 21-fold CV. This is computed for each possible subset of the most important features, as described in Section II-E, and the subset that yields the highest R^2 score is indicated with an arrow. For the DigiMoCA recordings (Fig. 2), the best possible outcome is reached by using the 22 most important features based on RFR, which yields an R^2 of 0.726. For the TAM recordings (Fig. 3), the best R^2 obtained is 0.732 by selecting the top 8 most important features given by DTR. For the combined dataset of all recordings (Fig. 4), the best possible R^2 actually decreases to 0.675, using the 4 most important features based on RFR.

Regarding Figs. 2, 3 and 4, we can notice that generally the R^2 value rapidly increases and then makes a slight decrease; this is potentially related to the huge difference in feature importances between the first and last ones, meaning that using just a small proportion of them provides the most information for the regression of the GDS, and adding more is just noise that simply worsens the result.

The main difference among the 3 datasets in terms of feature selection is that in the TAM data (Fig. 3) there are two highly important features (F3_BW_SD and F0_SEMI_P80). As opposed to this, in the DigiMoCA and combined datasets the "F3_BW_SD" feature is noticeably standing out from the rest in terms of importance. The "F3_BW_SD" feature is the standard deviation of the F3 bandwidth, which means it is related to the third formant. The "F0_SEMI_P80" is the percentile 80-th of logarithmic F0 on a semitone frequency

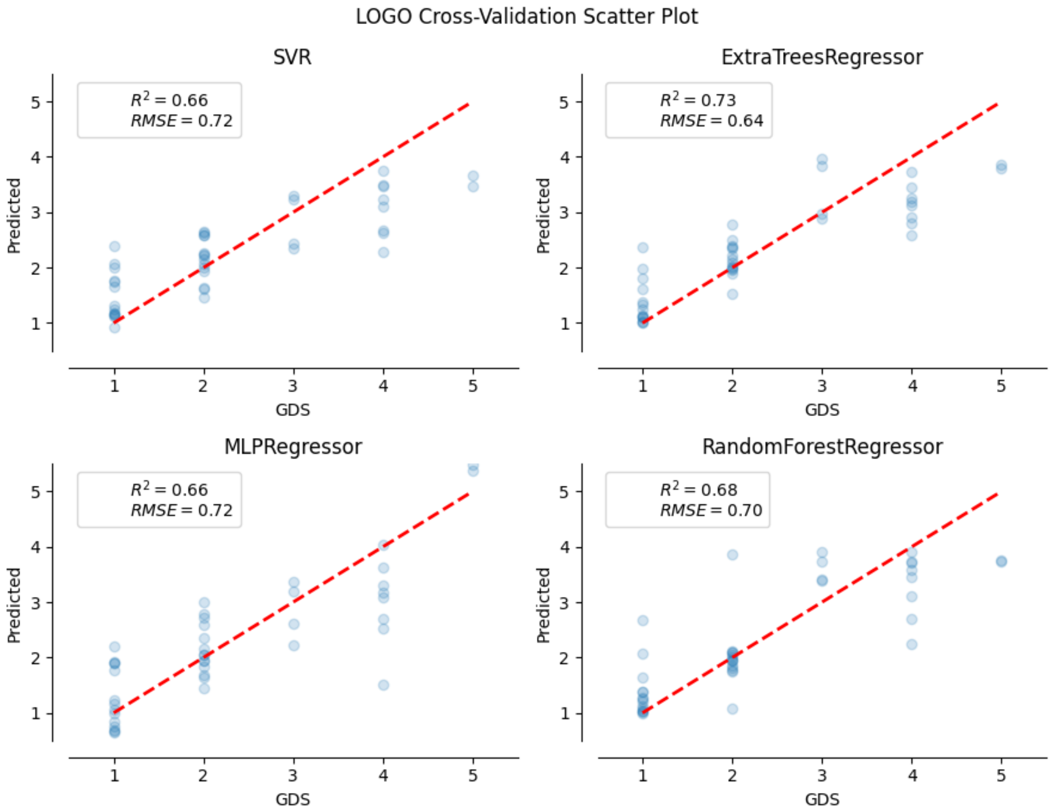

FIGURE 5. LOGO CV scatter plot with best 22 features in DigiMoCA S-version.

TABLE 2. Top 8 features (of 22 selected) from the DigiMoCA S-version.

Feature	Importance	p-value
F3_BW_SD	0.6484	4.2174e-7
F0_SEMI_P50	0.0438	7.5803e-8
F0_SEMI_P20	0.0382	0.0003
JTT_SD	0.0208	0.0015
F0_SEMI_MEAN	0.0127	1.5650e-8
F1_MEAN	0.0116	8.9335e-5
RT_MEAN	0.0112	1.6584e-8
F0_SEMI_SD	0.0104	0.5817

TABLE 3. Top 8 selected features from the TAM S-version.

Feature name	Importance	p-value
F3_BW_SD	0.3292	3.4243e-7
F0_SEMI_P80	0.2728	4.0269e-8
F0_SEMI_P20	0.0843	0.0012
SHM_SD	0.0564	0.0136
MFCC4V_MEAN	0.0372	3.7731e-8
RT_KUR	0.0330	0.2362
RT_SKEW	0.0205	0.0589
MFCC4_SD	0.0125	0.7657

TABLE 4. Top 4 selected features from combined S-version.

Feature name	Importance	p-value
F3_BW_SD	0.5563	3.8849e-13
F0_SEMI_P20	0.0673	1.0645e-6
F0_SEMI_P80	0.0618	1.8538e-13
SHM_SD	0.0307	0.0005

scale, starting at 27.5 Hz [57]. The importance of frequency and formant related features -especially F3- in the context of dementia and MCI detection has been outlined in previous works [31], [58], [59], [60]. The meaning of each prosodic feature is covered in App.

In Tables 2, 3 and 4 we can see the selected most important features, their importance given by the ensemble methods, and their p-value for the DigiMoCA, TAM and combined S-versions datasets respectively.

We can notice the exact difference in importances in Tables 2, 3 and 4, and how their corresponding p-value reinforces it. If we apply the Bonferroni correction (i.e., divide the threshold of significance by the number of features $P \leq 0.05/97 = 5.155e - 4$) we find that, if we group together all the datasets, 13 out of the 20 shown best features are statistically significant and associated with CI incidence, and are highlighted in green. This is a sign of statistical congruence between importances and p-value, which indicates the former to be a relevant metric.

In Figs. 5, 6 and 7 we can see the scatter plot of different regression algorithms applied with LOGO cross-validation for DigiMoCA, TAM and combined datasets respectively.

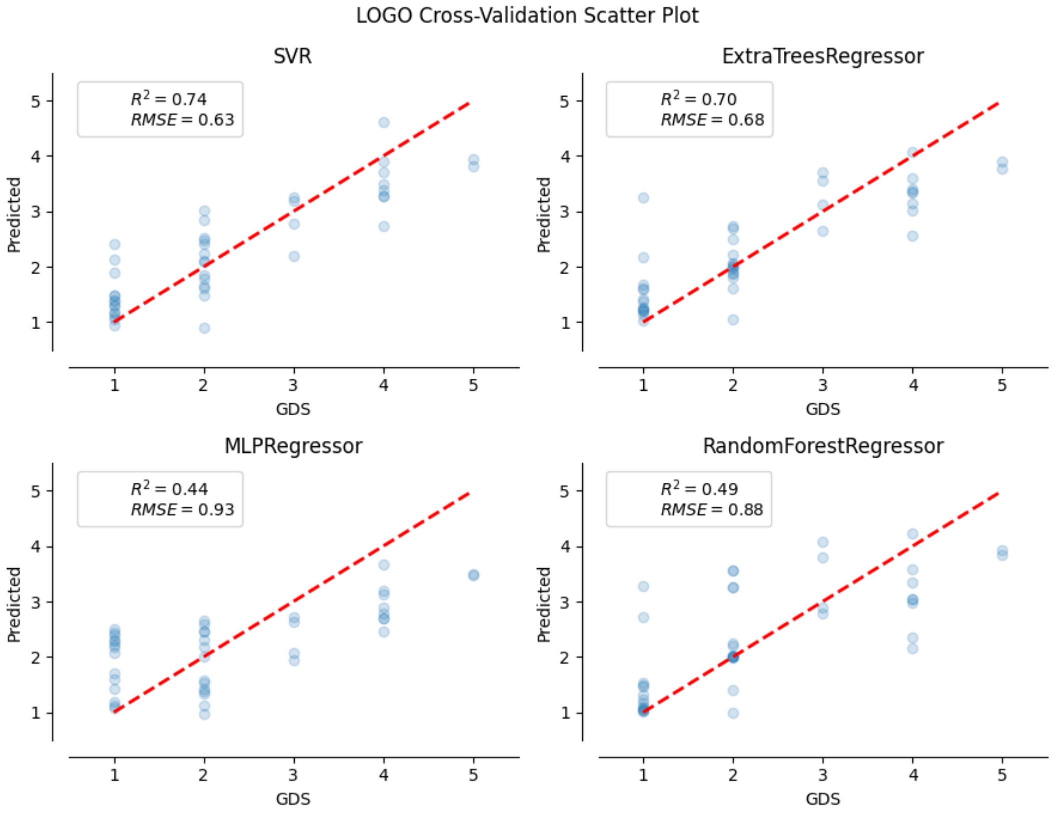

FIGURE 6. **LOGO CV scatter plot with best 8 features in TAM S-version.**

The features used for the regression of the GDS are the ones selected in the previous step for each dataset. In the scatter plots we can see the representation in blue dots of the predicted GDS value for each recording versus the real GDS of the participant. Therefore, the closer to the diagonal, the better the regression. We can also see the R^2 value and the *RMSE* in the top-left corner of each plot. We show the results for 4 popular regressors: Support Vector Regressor (SVR),[5] Random Forest Regressor, Extremely Randomized Trees Regressor and Multi-Layer Perceptron Regressor.

When it comes to the actual performance of the ML regression of the GDS, we can highlight that the best R^2 possible is 0.74, obtained with the S-version TAM dataset, using the top 8 most important features and with Support Vector Regression (SVR). The optimal R^2 obtained for the DigiMoCA dataset is similar (0.73), in this case with the Extra Trees Regressor, but it is noticeable how the performance decreases if we combine both datasets, to an optimal R^2 of 0.68. This decrease in performance is potentially related to the difference in nature of the speech conversations of the recordings: during the administration of the DigiMoca test with the smart speaker, the flow of the conversation is much more similar to that of a quick test, where the speaker asks a question and the participant answers; whereas during the TAM administration, the tone of the conversation is much more relaxed and casual, where the participant has more time to think. We also think this is one of the reasons why the response time mean (RT_MEAN) variable is one of the most significant ones in the DigiMoCA dataset, but not in the others. Nevertheless, it seems that mixing up different types of conversations does decrease the overall performance, and it is therefore a better idea to stick to a singular type of test when extracting audio features from a conversation. This looks especially important when having a small number of participants and recordings, as it happens to be the case in this pilot study.

From the scatter plots shown in Figs. 5, 6 and 7 we can also notice another subtle detail. If we look at the distribution of the points (real vs. predicted GDS) we can see that in the cases of 1 and 2 real GDS, the prediction is usually grouped around the real value and quite accurate. However, for higher GDS values (3, 4 and 6) the predictions are more scattered and not as accurate; in fact, in most of the plots and for the majority of the algorithms, the prediction for the different real GDS values do not differ noticeably on average. Therefore, the regression performs better with lower GDS values than higher, and when we obtain a prediction close to 1 or 2, we can be fairly confident that it is accurate; if it is higher than 3, not as much. This is probably due to the sample of participants,

[5] SVR is a regression variant of the Support Vector Machine (SVM) classification algorithm.

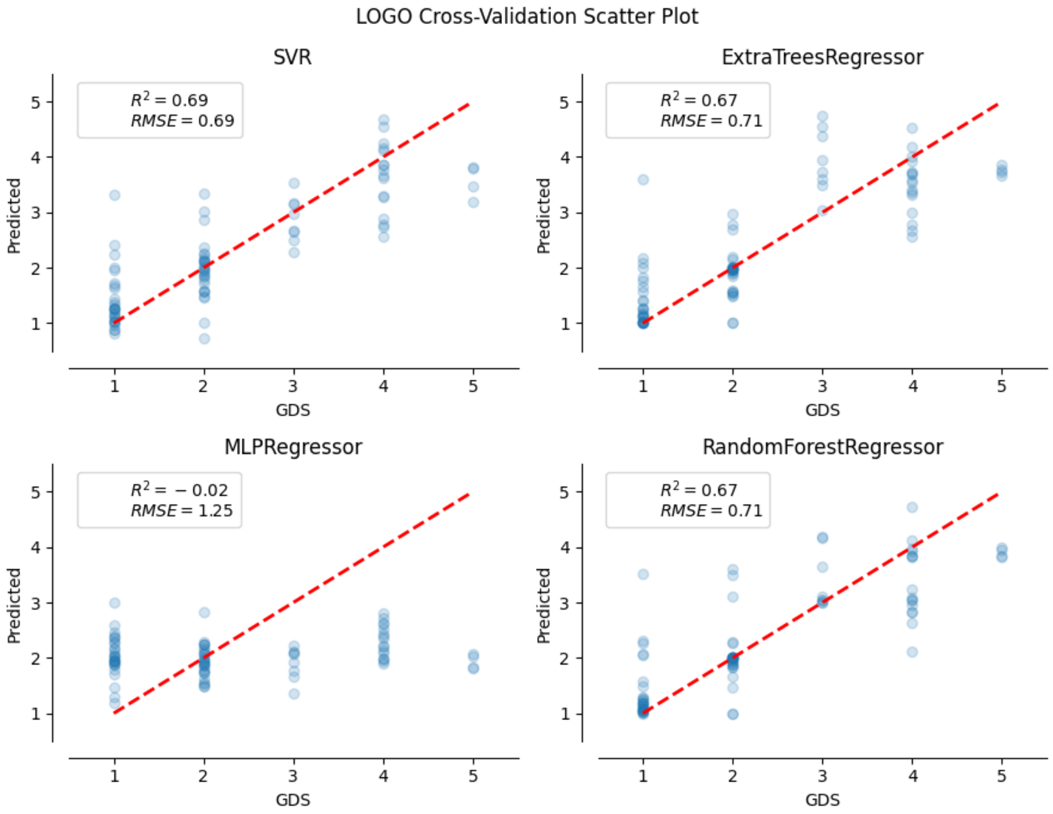

FIGURE 7. LOGO CV scatter plot with best 4 features in combined S-version.

since it is equally split between HC, MCI and AD, which corresponds to 1, 2 and 3 onwards. Consequently, if we group the regression in these three categories, we notice that the prediction of GDS makes for a good classification in them. This is obviously a very positive outcome since it is more important to be able to classify between the different stages of CI incidence than knowing the exact value of GDS.

IV. CONCLUSION

In this paper we present and discuss the analysis of voice biomarkers (i.e., features) for the detection of cognitive impairment. More specifically, we measured their capacity for predicting the GDS of the recorded person's cognitive state, by means of Machine Learning regression. We automatically extracted and selected the most relevant features, and measured the performance with standard metrics and cross-validation techniques. The best result obtained was an R^2 of 0.74, using the 8 most important features of the TAM S-version dataset (answers recorded during the administration of a TAM questionnaire), with an SVR model.

The main strength of this study is the fact that we obtain the results from audio and speech features only. The GDS rates the stage of the cognitive decline, and thus takes into account several cognitive characteristics, such as memory deficit, time and spatial orientation, concentration deficit, etc. Therefore, being able to predict this scale with a coefficient of determination of 0.74 is significant and a positive indicator of the informativeness of voice and speech characteristics.

Additionally, we should highlight the ecological nature of this technique, as it pertains to cognitive skills utilized in everyday life. Ecological validity is defined as the degree to which results obtained in controlled experimental conditions (i.e., interaction with a smart speaker in this research) are related to those obtained in naturalistic environments [61]. By using speech features only, we are able to extract the necessary information just by recording the speaker while talking in a conversation. In this case, we were able to obtain interesting results from a DigiMoCA and TAM questionnaire administration, and we suspect the analysis could be generalized for a general relaxed conversation of similar length. The advantage of this is that the end user can be subject to a screening process without even being aware of it, which is the least intrusiveness possible in a screening tool for this purpose.

In terms of the limitations of the study, the main one is the small sample size of the participants. Despite being ideally distributed among the three main layers of cognitive impairment (i.e., HC, MCI and AD), 21 is not the desired sample size in order to determine the true relevance of voice biomarkers in terms of CI incidence. Therefore, we can establish that the results found in this study are neither definitive nor really representative of the true potential

TABLE 5. List of used prosodic features and brief description.

Feature	OpenSMILE name	Description [57]
F0_SEMI_MEAN	F0semitoneFrom27.5Hz_sma3nz_amean	Mean of logarithmic F0 on a semitone frequency scale, starting at 27.5 Hz
F0_SEMI_P20	F0semitoneFrom27.5Hz_sma3nz_percentile20.0	Percentile 20-th of logarithmic F0 on a semitone frequency scale, starting at 27.5 Hz
F0_SEMI_P50	F0semitoneFrom27.5Hz_sma3nz_percentile50.0	Percentile 50-th of logarithmic F0 on a semitone frequency scale, starting at 27.5 Hz
F0_SEMI_P80	F0semitoneFrom27.5Hz_sma3nz_percentile80.0	Percentile 80-th of logarithmic F0 on a semitone frequency scale, starting at 27.5 Hz
F0_SEMI_SD	F0semitoneFrom27.5Hz_sma3nz_stddevNorm	Standard deviation of logarithmic F0 on a semitone frequency scale, starting at 27.5 Hz
F1_MEAN	F1frequency_sma3nz_amean	Mean of the centre frequency of first formant in voiced regions
F3_BW_SD	F3bandwidth_sma3nz_stddevNorm	Standard deviation of the F3 bandwidth
JTT_SD	jitterLocal_sma3nz_stddevNorm	Standard deviation of the variations in individual consecutive F0 period lengths
MFCC4V_MEAN	mfcc4V_sma3nz_amean	Mean of Mel-Frequency Cepstral Coefficient 4 in voiced regions
MFCC4_SD	mfcc4_sma3_stddevNorm	Standard deviation of Mel-Frequency Cepstral Coefficient 4
SHM_SD	shimmerLocaldB_sma3nz_stddevNorm	Standard deviation of difference of the peak amplitudes of consecutive F0 periods
RT_KUR	-	Kurtosis of the distribution of the response times
RT_SKEW	-	Skewness of the distribution of the response times
RT_MEAN	-	Mean response time

for acoustic features in this regard. Additionally, another limitation is the fact that all the participants had biological roots in Spain and were spanish-speaking, which prevents us from generalizing the results to different ethnicity and/or language classes.

In order to improve and make our evaluation more robust, a bigger number of participants is needed, so that we ensure a sufficient variety in terms of gender, GDS level and age. Furthermore, a similar study with participants of a different language-speaking country would be extremely insightful.

APPENDIX

PROSODIC FEATURES

See Table 5.

REFERENCES

[1] S. Cahill, "WHO's global action plan on the public health response to dementia: Some challenges and opportunities," *Aging Mental Health*, vol. 24, no. 2, pp. 197–199, Feb. 2020.

[2] M. V. F. Silva, C. D. M. G. Loures, L. C. V. Alves, L. C. de Souza, K. B. G. Borges, and M. D. G. Carvalho, "Alzheimer's disease: Risk factors and potentially protective measures," *J. Biomed. Sci.*, vol. 26, no. 1, p. 33, May 2019.

[3] Z. Breijyeh and R. Karaman, "Comprehensive review on Alzheimer's disease: Causes and treatment," *Molecules*, vol. 25, no. 24, p. 5789, Dec. 2020.

[4] *First WHO Ministerial Conference on Global Action Against Dementia: Meeting Report*, World Health Org., Geneva, Switzerland, 2015.

[5] M. P. Ates, Y. Karaman, S. Guntekin, and M. A. Ergun, "Analysis of genetics and risk factors of Alzheimer's disease," *Neuroscience*, vol. 325, pp. 124–131, Jun. 2016.

[6] C. Gillis, F. Mirzaei, M. Potashman, M. A. Ikram, and N. Maserejian, "The incidence of mild cognitive impairment: A systematic review and data synthesis," *Alzheimer's Dementia, Diagnosis, Assessment Disease Monit.*, vol. 11, no. 1, pp. 248–256, Dec. 2019.

[7] M. D. Lezak, *Neuropsychological Assessment*, 4th ed., London, U.K.: Oxford Univ. Press, 2004.

[8] D. B. Howieson and M. D. Lezak, *The Neuropsychological Evaluation*. Washington, DC, USA: American Psychiatric Publishing, 2010, pp. 29–54.

[9] N. Chaytor and M. Schmitter-Edgecombe, "The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills," *Neuropsychol. Rev.*, vol. 13, no. 4, pp. 181–197, Dec. 2003.

[10] D. Holtzman, J. Morris, and A. Goate, "Alzheimer's disease: The challenge of the second century," *Sci. Transl. Med.*, vol. 3, no. 77, 2011, Art. no. 77sr1.

[11] R. G. Knight and N. Titov, "Use of virtual reality tasks to assess prospective memory: Applicability and evidence," *Brain Impairment*, vol. 10, no. 1, pp. 3–13, May 2009.

[12] S. Farias, "The relationship between neuropsychological performance and daily functioning in individuals with Alzheimer's disease: Ecological validity of neuropsychological tests," *Arch. Clin. Neuropsychol.*, vol. 18, no. 6, pp. 655–672, Aug. 2003.

[13] C. B. Cordell, S. Borson, M. Boustani, J. Chodosh, D. Reuben, J. Verghese, W. Thies, and L. B. Fried, "Alzheimer's association recommendations for operationalizing the detection of cognitive impairment during the medicare annual wellness visit in a primary care setting," *Alzheimer's Dementia*, vol. 9, no. 2, pp. 141–150, Mar. 2013.

[14] K. A. Hawkins, D. Dean, and G. D. Pearlson, "Alternative forms of the rey auditory verbal learning test: A review," *Behavioural Neurol.*, vol. 15, nos. 3–4, pp. 99–107, Jan. 2004.

[15] S. J. Robinson and G. Brewer, "Performance on the traditional and the touch screen, tablet versions of the corsi block and the tower of Hanoi tasks," *Comput. Hum. Behav.*, vol. 60, pp. 29–34, Jul. 2016.

[16] T. Tong and M. Chignell, "Developing a serious game for cognitive assessment: Choosing settings and measuring performance," in *Proc. 2nd Int. Symp. Chin. CHI*, Apr. 2014, pp. 70–79.

[17] L. Beck, M. Wolter, N. F. Mungard, R. Vohn, M. Staedtgen, T. Kuhlen, and W. Sturm, "Evaluation of spatial processing in virtual reality using functional magnetic resonance imaging (fMRI)," *Cyberpsychol., Behav., Social Netw.*, vol. 13, pp. 211–215, Apr. 2010.

[18] T. Parsons, P. Larson, K. Kratz, M. Thiebaux, B. Bluestein, J. Buckwalter, and A. Rizzo, "Sex differences in mental rotation and spatial rotation in a virtual environment," *Neuropsychologia*, vol. 42, no. 4, pp. 555–562, Feb. 2004.

[19] G. Plancher, A. Tirard, V. Gyselinck, S. Nicolas, and P. Piolino, "Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and Alzheimer's disease: Influence of active and passive encoding," *Neuropsychologia*, vol. 50, no. 5, pp. 592–602, Apr. 2012.

[20] P. Nolin, F. Banville, J. Cloutier, and P. Allain, "Virtual reality as a new approach to assess cognitive decline in the elderly," *Academic J. Interdiscipl. Stud.*, vol. 2, pp. 612–616, Oct. 2013.

[21] F. Banville, P. Nolin, S. Lalonde, M. Henry, M.-P. Dery, and R. Villemure, "Multitasking and prospective memory: Can virtual reality be useful for diagnosis?" *Behav. Neurol.*, vol. 23, no. 4, pp. 209–211, 2010.

[22] R. Nori, L. Piccardi, M. Migliori, A. Guidazzoli, F. Frasca, D. De Luca, and F. Giusberti, "The virtual reality walking corsi test," *Comput. Hum. Behav.*, vol. 48, pp. 72–77, Jul. 2015.

[23] Y. Iriarte, U. Diaz-Orueta, E. Cueto, P. Irazustabarrena, F. Banterla, and G. Climent, "AULA—Advanced virtual reality tool for the assessment of attention: Normative study in Spain," *J. Attention Disorders*, vol. 20, no. 6, pp. 542–568, Jun. 2016.

[24] P. Nolin, A. Stipanicic, M. Henry, Y. Lachapelle, D. Lussier-Desrochers, A. Rizzo, and P. Allain, "*ClinicaVR: Classroom-CPT*: A virtual reality tool for assessing attention and inhibition in children and adolescents," *Comput. Hum. Behav.*, vol. 59, pp. 327–333, Jun. 2016.

[25] S. E. Counts, M. D. Ikonomovic, N. Mercado, I. E. Vega, and E. J. Mufson, "Biomarkers for the early detection and progression of Alzheimer's disease," *Neurotherapeutics*, vol. 14, no. 1, pp. 35–53, Jan. 2017.

[26] S. Cavedoni, A. Chirico, E. Pedroli, P. Cipresso, and G. Riva, "Digital biomarkers for the early detection of mild cognitive impairment: Artificial intelligence meets virtual reality," *Frontiers Hum. Neurosci.*, vol. 14, p. 245, Jul. 2020.

[27] P. Lewczuk, B. Mroczko, A. Fagan, and J. Kornhuber, "Biomarkers of Alzheimer's disease and mild cognitive impairment: A current perspective," *Adv. Med. Sci.*, vol. 60, no. 1, pp. 76–82, Mar. 2015.

[28] M. N. Sabbagh, M. Boada, S. Borson, P. M. Doraiswamy, B. Dubois, J. Ingram, A. Iwata, A. P. Porsteinsson, K. L. Possin, G. D. Rabinovici, B. Vellas, S. Chao, A. Vergallo, and H. Hampel, "Early detection of mild cognitive impairment (MCI) in an at-home setting," *J. Prevention Alzheimer's Disease*, vol. 7, no. 3, pp. 171–178, 2020.

[29] I. Hajjar, M. Okafor, J. D. Choi, E. Moore, A. Abrol, V. D. Calhoun, and F. C. Goldstein, "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 Monit.*, vol. 15, no. 1, Jan. 2023, Art. no. e12393.

[30] M. Verma and R. J. Howard, "Semantic memory and language dysfunction in early Alzheimer's disease: A review," *Int. J. Geriatric Psychiatry*, vol. 27, no. 12, pp. 1209–1217, Dec. 2012.

[31] J. J. G. Meilán, F. Martínez-Sánchez, J. Carro, D. E. López, L. Millian-Morell, and J. M. Arana, "Speech in Alzheimer's disease: Can temporal and acoustic parameters discriminate dementia?" *Dementia Geriatric Cognit. Disorders*, vol. 37, nos. 5–6, pp. 327–334, 2014.

[32] M. Asgari, J. Kaye, and H. Dodge, "Predicting mild cognitive impairment from spontaneous spoken utterances," *Alzheimer's Dementia, Transl. Res. Clin. Intervent.*, vol. 3, no. 2, pp. 219–228, Jun. 2017.

[33] Z. Noorian, C. Pou-Prom, and F. Rudzicz, "On the importance of normative data in speech-based assessment," 2017, *arXiv:1712.00069*.

[34] M. R. Pacheco-Lorenzo, S. M. Valladares-Rodríguez, L. E. Anido-Rifón, and M. J. Fernández-Iglesias, "Smart conversational agents for the detection of neuropsychiatric disorders: A systematic review," *J. Biomed. Informat.*, vol. 113, Jan. 2021, Art. no. 103632.

[35] M. R. Pacheco-Lorenzo, S. Valladares-Rodríguez, L. Anido-Rifón, and M. J. Fernández-Iglesias, "A conceptual framework based on conversational agents for the early detection of cognitive impairment," in *Proc. 2nd Int. Conf. Artif. Intell., Adv. Appl.*, G. Mathur, M. Bundele, M. Lalwani, and M. Paprzycki, Eds. Singapore: Springer, 2022, pp. 801–813.

[36] B. Mirheidari, D. Blackburn, T. Walker, M. Reuber, and H. Christensen, "Dementia detection using automatic analysis of conversations," *Comput. Speech Lang.*, vol. 53, pp. 65–79, Jan. 2019.

[37] B. Mirheidari, D. Blackburn, R. O'Malley, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Computational cognitive assessment: Investigating the use of an intelligent virtual agent for the detection of early signs of dementia," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2732–2736.

[38] M. Fukuda, R. Nishimura, M. Umezawa, K. Yamamoto, Y. Iribe, and N. Kitaoka, "Elderly conversational speech corpus with cognitive impairment test and pilot dementia detection experiment using acoustic characteristics of speech in Japanese dialects," in *Proc. 30th Lang. Resour. Eval. Conf.*, 2022, pp. 1016–1022.

[39] B. Reisberg et al., "Two year outcomes, cognitive and behavioral markers of decline in healthy, cognitively normal older persons with global deterioration scale stage 2 (subjective cognitive decline with impairment)," *J. Alzheimer's Disease*, vol. 67, no. 2, pp. 685–705, Jan. 2019.

[40] S. Luz, F. Haider, S. de la F. Garcia, D. Fromm, and B. Macwhinney, "Alzheimer's dementia recognition through spontaneous speech: The ADReSS challenge," in *Proc. Interspeech*. The International Speech Communication Association (ISCA), 2020, pp. 2172–2176, doi: 10.21437/Interspeech.2020-2571.

[41] S. Luz, F. Haider, S. de la Fuente, D. Fromm, and B. MacWhinney, "Detecting cognitive decline using speech only: The ADReSSo challenge," 2021, *arXiv:2104.09356*.

[42] Z. S. Syed, M. S. S. Syed, M. Lech, and E. Pirogova, "Tackling the ADRESSO challenge 2021: The MUET-RMIT system for Alzheimer's dementia recognition from spontaneous speech," in *Proc. Interspeech*, 2021, pp. 3815–3819.

[43] X. Liang, J. A. Batsis, Y. Zhu, T. M. Driesse, R. M. Roth, D. Kotz, and B. MacWhinney, "Evaluating voice-assistant commands for dementia detection," *Comput. Speech Lang.*, vol. 72, Mar. 2022, Art. no. 101297.

[44] P. Perez-Toro, P. Klumpp, A. Hernandez, T. Arias-Vergara, P. Lillo, A. Slachevsky, A. García, M. Schuster, A. Maier, E. Noeth, and J. Orozco-Arroyave, "Alzheimer's detection from English to Spanish using acoustic and linguistic embeddings," in *Proc. Interspeech*, 2022, pp. 2483–2487.

[45] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "Wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. Red Hook, NY, USA: Curran Associates, 2020, pp. 12449–12460.

[46] R.-P. Filiou, N. Bier, A. Slegers, B. Houzé, P. Belchior, and S. M. Brambati, "Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: A scoping review," *Aphasiology*, vol. 34, no. 6, pp. 723–755, Jun. 2020.

[47] C. Hardcastle, B. Taylor, and C. Price, *Global Deterioration Scale*. Cham, Switzerland: Springer, 2021, pp. 2198–2201.

[48] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment," *J. Amer. Geriatrics Soc.*, vol. 53, no. 4, pp. 695–699, Apr. 2005.

[49] A. M. Aburbeian, A. Y. Owda, and M. Owda, "A technology acceptance model survey of the metaverse prospects," *AI*, vol. 3, no. 2, pp. 285–302, Apr. 2022.

[50] X. Anguera, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, and O. Vinyals, "Speaker diarization: A review of recent research," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 2, pp. 356–370, Feb. 2012.

[51] A. Vibhute, "Feature extraction techniques in speech processing a survey," *Int. J. Comput. Appl.*, vol. 107, pp. 1–8, Dec. 2014.

[52] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: The Munich versatile and fast open-source audio feature extractor," in *Proc. 18th ACM Int. Conf. Multimedia*, Oct. 2010, pp. 1459–1462.

[53] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 190–202, Apr. 2016.

[54] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021.

[55] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 36, no. 2, pp. 111–133, Jan. 1974.

[56] C. Themistocleous, M. Eckerström, and D. Kokkinakis, "Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls," *PLoS ONE*, vol. 15, no. 7, Jul. 2020, Art. no. e0236009.

[57] M. Corrales-Astorgano, D. Escudero, and C. González-Ferreras, "Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with down syndrome," *Speech Commun.*, vol. 99, pp. 90–100, Mar. 2018.

[58] I. Martínez-Nicolás, T. E. Llorente, F. Martínez-Sánchez, and J. J. G. Meilán, "Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: A systematic review article," *Frontiers Psychol.*, vol. 12, Mar. 2021, Art. no. 620251.

[59] K. Nishikawa, H. Kawano, R. Hirakawa, and Y. Nakatoh, "Analysis of prosodic features and formant of dementia speech for machine learning," in *Proc. 5th Int. Conf. Inf. Comput. Technol. (ICICT)*, Mar. 2022, pp. 173–176.

[60] A. Khodabakhsh and C. Demiroglu, "Analysis of speech-based measures for detecting and monitoring Alzheimer's disease," in *Data Mining in Clinical Medicine*. New York, NY, USA: Springer, 2015, pp. 159–173.

[61] D. E. Tupper and K. D. Cicerone, *The Neuropsychology of Everyday Life: Assessment and Basic Competencies* (Foundations of Neuropsychology). New York, NY, USA: Springer, 1990.

MOISÉS R. PACHECO-LORENZO received the B.S. degree in telecommunication technologies engineering, in 2017, and the M.S. degree in telecommunication engineering, in 2020. Currently, he is pursuing the Ph.D. degree with the University of Vigo. His fields of research interests include computer networks, cloud development, artificial intelligence, machine learning, and big data.

HEIDI CHRISTENSEN (Member, IEEE) received the M.Sc. and Ph.D. degrees from Aalborg University, Aalborg, Denmark, in 1996 and 2002, respectively. She is currently a Professor with the Spoken Language Technologies, Computer Science Department, The University of Sheffield, Sheffield, U.K. Before that, she held postdoctoral positions with The University of Sheffield, IDIAP, Switzerland, and Aalborg University. Her main research interests include the recognition of disordered speech, the automatic processing of conversations, and the automatic detection and tracking of paralinguistic information, such as emotions and general interactional behaviors.

LUIS E. ANIDO-RIFÓN (Senior Member, IEEE) received the Telecommunication Engineering degree (Hons.) from the University of Vigo, the Galician Government, and the Spanish Ministry of Education, in 1997, and the Ph.D. degree (Hons.) in telecommunication engineering from the University of Vigo, in 2001. He graduated both in the Telematics and Communication Branches. He joined the Telematics Engineering Department, in 1997. Since 2009, he has been a Full Professor. In November 2021, he was appointed as the Head of the Telematics Engineering Department. He has received awards from the World Wide Web Consortium (W3C), the Royal Academy of Sciences, and the Official Spanish Telecommunication Association. During the Ph.D. degree, he received many awards from the University of Vigo.

MANUEL J. FERNÁNDEZ-IGLESIAS received the degree in telecommunication engineering from Universidade de Santiago de Compostela, Spain, in 1990, and the Ph.D. degree in telecommunication engineering from Universidade de Vigo, Spain, in 1997. Since 1990, he has been involved in lecturing and research with the School of Telecommunication Engineering, Universidade de Vigo. His current research interests include how technology can support elder adults and blockchain applications in education. He is also looking into how blockchain can be used in education, as this technology has the potential to revolutionize the way educational institutions manage, validate, and share data.

SONIA M. VALLADARES-RODRÍGUEZ received the Ph.D. degree in telematics engineering from the University of Vigo, in April 2019. Currently, she performs teaching and research tasks with the University of Santiago de Compostela. With entrepreneurial concern, she is the Co-Founder of DataSalus. As a Data Scientist, she is convinced of the transfer of knowledge from the university to the business environment and, therefore, to society.