DOI: 10.61782/fa.2025.0380

FORUM ACUSTICUM EURONOISE 2025

EXPLAINABLE FEATURE SELECTION FOR DEMENTIA RECOGNITION FROM ACOUSTIC AND LINGUISTIC CUES

Ascensión Gallardo-Antolín[1*] **Irati Oiza-Zapata**[1] **Javier Ferreiros-López**[2]
Fernando Fernández-Martínez[2] **Rubén San-Segundo**[2]
[1] Universidad Carlos III de Madrid,
Departmento de Teoría de la Señal y Comunicaciones, Spain
[2] Information Processing and Telecommunications Center,
Speech Technology and Machine Learning Group,
E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, Spain

ABSTRACT

Dementia represents a substantial health challenge, with early detection being critical to enable timely interventions and mitigate its progression. Speech analysis, integrated with Machine Learning (ML) techniques, has gained prominence as a promising approach for the automatic detection of dementia, as vocal and linguistic biomarkers serve as valuable indicators of cognitive impairment that can be effectively exploited by ML algorithms. Explainability is a critical requisite for the practical application of ML-based systems in clinical settings. In this study, we address this challenge by feeding the system with a reduced set of acoustic and linguistic features easily understood by humans. These features are determined utilizing SHapley Additive exPlanations (SHAP) values, an Explainable Artificial Intelligence (XAI) method. This way, SHAP helps to identify the most impactful features on the predictions, which allows not only to explain the model decisions, but also to select the characteristics according to their global relevance, thus optimizing the model and enhancing its explainability. The proposed framework is firstly applied independently to the acoustic and text (transcriptions) modalities and secondly, to the multimodal system. Experiments on the ADReSS dataset demonstrate its feasibility and highlight the potential of explainable feature selection to bridge ML techniques with clinically meaningful insights.

Keywords: *dementia detection, speech biomarkers, explainable machine learning, feature selection, Shapley values.*

1. INTRODUCTION

Dementia is a broad term referring to the loss of memory, language, problem-solving, and other cognitive abilities whose leading cause is the neurodegenerative disorder called Alzheimer's Disease (AD) [1]. As AD is an age-related disorder, it is expected to become more prevalent as the elderly population grows, what it is very likely to occur in the near future. This trend highlights the need for effective strategies for AD early detection and monitoring.

However, existing diagnostic methods are expensive and time-consuming [2], leading to a large number of people with AD not receiving timely interventions. Digital biomarkers can alleviate this problem as they offer remote and non-intrusive evaluations, saving time and costs [3]. In this context, speech biomarkers have become a promising approach, as it is well-known that one of the first signs of AD is degradation in voice and language production [4]. In fact, AD can alter the physical characteristics of the voice, leading to reduced articulatory precision and speech fluency, which in turn affects speech intelligibility. It also decreases prosodic variation, yielding a more monotonous speech. Additionally, AD impacts language production, often resulting in a diminished vocab-

**Corresponding author*: *gallardo@ing.uc3m.es.*

ulary, loss of word meaning, difficulty in finding the right words, and reduced coherence in discourse.

As a result, there is currently an active research line on AD detection from speech using acoustic, linguistic modalities or their combination [5]. Many of these studies have been conducted within the framework of the ADReSS Challenge, introduced at the INTERSPEECH 2020 conference [6], which has significantly contributed to the growing research interest in this area.

These works follow mainly two different approaches. The first one is based on the use of hand-crafted features, such as the eGeMAPSv2 repertory [7, 8] for the audio modality or lexical diversity descriptors for the textual one, in combination with traditional Machine Learning (ML) models [9]. The second one adopts Deep Learning (DL) methods often through the use of pre-trained models, such as VGGish [10] or BERT [11], for the extraction of, respectively, acoustic and linguistic embeddings. Although DL models typically achieve high accuracy, their opacity hinder their use in sensitive areas like healthcare, where interpretability is crucial. In fact, for the AD detection system to be clinically useful, it would need to provide the medical staff not only with a prediction of whether a subject is suffering or not AD, but also an explanation of the factors it relied on to make that decision.

From this perspective, an appealing approach is the use of more interpretable methods, such as ML models powered by understandable features. Besides, a reduced number of characteristics would make easier to explain the system output. The first issue can be addressed through the application of eXplainable Artificial Intelligence (XAI) methods. Specifically, the Shapley Additive Explanations (SHAP) approach [12] is widely employed in a diversity of tasks. However, for the best of our knowledge, it has been scarcely explored for dementia recognition from speech [8]. Regarding the second issue, the literature offers various feature selection techniques [13]. However, these methods typically do not incorporate interpretability considerations into the selection process.

Considering both aspects, this paper focuses on designing a reduced set of interpretable multimodal speech biomarkers for dementia detection. To achieve this, an explainable feature selection method is proposed, which relies on SHAP values as a metric for choosing the most relevant features. Results show that this strategy not only enhances the system interpretability and clinical utility but also improves its accuracy, specificity and sensitivity.

The rest of the paper is organized as follows: Section 2 outlines AD-related acoustic and linguistic changes. Section 3 details our methodology. Section 4 covers the dataset, experiments and results. Finally, Section 5 presents the main conclusions and future work.

2. DEMENTIA SPEECH PATTERNS

As speech can reveal cognitive disorders, tasks eliciting spontaneous speech are extensively used for assessing the vocal and language abilities of patients. In this context, the task of describing the Cookie Theft picture from the Boston Diagnostic Aphasia Examination [14] , that is represented in Fig. 1, is one of most commonly employed.

From an acoustic perspective, the analysis of the spoken descriptions facilitates the identification of characteristic speech patterns commonly associated with dementia [15, 16]: (1) changes in prosodic features, such as reduced pitch variability, resulting in a more monotonous voice, and decreased speech intensity, leading to a weaker voice; (2) distortions in voice quality, including increased breathiness and hoarseness; (3) reduced articulatory precision, which reduces speech clarity and overall intelligibility; and (4) a slower speech rate with more pauses and hesitations, disrupting natural speech fluency.

Regarding the linguistic domain, the Cookie Theft task is useful to reveal diminishing skills in several aspects [17]: (1) salience of information: AD subjects may not describe the low salience or background details (e.g. curtains or window in the picture); (2) semantic categories: AD subjects may employ more general than concrete terms (e.g., "lady" instead of "mother"); (3) referential cohesion: AD subjects may misuse the pronouns to refer to a particular person or object; (4) causal and temporal relations: AD subjects may fail in describing this kind of relationships; (5) mental state language: descriptions by AD subjects may lack terms like "want", "attention" or "see"; (6) structural language and speech: descriptions by AD persons may contain many unfilled or filled ("eh", "well") pauses and non-specific vocabulary ("it", "something"); and (7) general cognition and perception: AD people may produce fragmented or disorganized descriptions.

3. METHODOLOGY

Fig. 2 represents the proposed multimodal system for AD detection. As shown in Fig. 2(a), the process starts by training a ML model fed with the whole set of acoustic and linguistic features. Then, the contribution of each feature to the model output is computed with SHAP (see Subsection 3.4). As depicted in Fig. 2(b), these SHAP values are

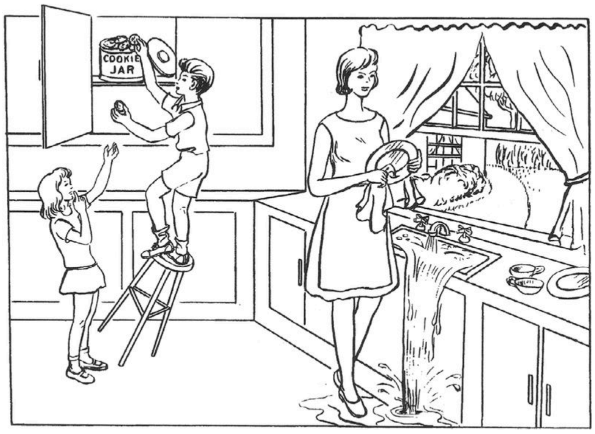

Figure 1. Cookie Theft picture from the Boston Diagnostic Aphasia Examination [14].

used as metric for selecting the more relevant characteristics. This subset is the input to the final ML model, that is subsequently trained. For a particular example, this model outputs both the prediction and the corresponding explanation in terms of its SHAP values. In next subsections, the main components of the system are described.

3.1 Feature extraction

3.1.1 Acoustic characteristics

In this study, we have focused on acoustic features with a clear relationship with the perceptual alterations that dementia can produce on speech as described in Section 2, with the aim of obtaining a repertory of interpretable and clinically useful acoustic biomarkers of AD. The set of acoustic characteristics is based on those proposed in [18, 19], and is composed of the following 34 features:

- Prosodic features: Intensity (minimum, maximum, mean, standard deviation, q1, median and q3), and pitch (minimum, maximum, mean, standard deviation, q1, median, q3, fraction of voiced frames and the mean of the absolute pitch slope). Both are linked to prosody perception.
- Voice quality features: Glottal to Noise Excitation (GNE) ratio (maximum, mean, standard deviation), that is related to breathiness.
- Articulatory features: Formants (standard deviation of the first, second, third and fourth formants, formant dispersion and average formant), that are associated to the control of the articulatory muscles. Also, articulation rate (number of syllables divided by the duration of the audio excluding pauses) and average duration of the syllable (the duration of the audio excluding pauses divided by the number of voiced peaks) are considered as they are related to the articulation precision.
- Fluency and pause features: Speech rate (number of syllables divided by the duration of the audio including pauses) that is a metric for the perceived speaking speed. Additionally, pause features (total speech time, total pause time, percentage of pauses, pause-to-speech ratio, and mean and standard deviation of pause length) are included for measuring the speech fluency.

The acoustic features were extracted following [18], that in turn used the *Parselmouth* [20, 21] and the *DigiPsychProsody* [22] packages.

3.1.2 Linguistic characteristics

The linguistic characteristics consists of the "Term Frequency-Inverse Document Frequency" (TF-IDF) features, previously proposed for this task in [9]. TF-IDF represents the importance of a term in a document (transcription) balancing its frequency in that document with its rarity across the corpus, and can reflect the linguistic patterns of dementia speech mentioned in Section 2.

The TF-IDF representation was extracted from the preprocessed transcriptions of the audio recordings. The preprocessing stage consisted of the removal of special characters and lemmatization, that was performed with the spaCy package [23] and the *en_core_web_sm* model. Then, the TF-IDF model was generated with the *scikit-learn* package [24] setting the dictionary size to 150 terms and eliminating the stopwords. The list of stopwords was created *ad hoc* for this task, before the examination of the tokens contained in the training corpus.

3.2 Classifier

In this work, we have chosen to employ traditional ML models over more complex DL approaches, as the former provide enhanced interpretability, especially when using hand-crafted features. In particular, the developed classifier is based on Support Vector Machines (SVM) that has demonstrated good performance over small datasets, as is our case. For the implementation of the SVM classifier,

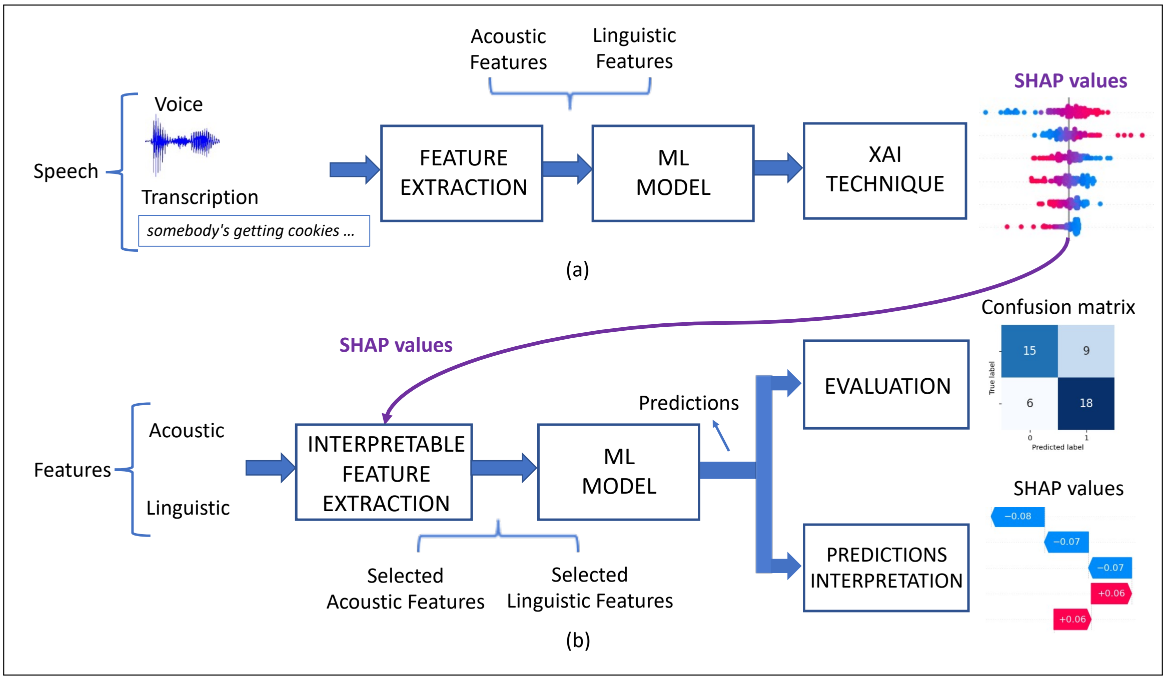

Figure 2. Interpretable feature selection for the dementia recognition system based on acoustic and linguistic characteristics: (a) Block diagram of the computation of the Shapley values; (b) Block diagram of the dementia detection system where the input is the set of selected interpretable features and the output are the predictions made by the ML model and their explanation in terms of the Shapley values.

the *sckit-learn* package [24] was employed. In the training process, the hyperparameters were tuned using grid search and a 5 fold cross-validation strategy.

3.3 Interpretability method

To enhance the transparency of the ML model, we have employed the SHAP technique due to its capability to offer both local explanations for individual predictions and global interpretability of the model's overall behavior. This last property allows the use of SHAP as a feature selection mechanism, as shown in Section 3.4.

SHAP is a model-agnostic XAI technique grounded on game theory, that quantifies the contribution of each feature to a prediction [12]. In binary classification tasks, as is our case, SHAP values indicate the extent to which each feature influences the model output (score), pushing it toward either the positive class (AD) or the negative one (non-AD). The prediction for a given instance can be reconstructed as the sum of its SHAP values, plus a baseline term representing the mean prediction in the dataset.

3.4 Interpretable Feature Selection

As previously stated, SHAP quantifies the contribution of each feature to a particular prediction. This contribution can be seen as a measure of the feature relevance, and therefore, it can be used as a criterion for feature selection.

In order to employ SHAP as an interpretable feature selection technique, we leverage its capability to provide a global interpretation of the model's behavior as proposed in [25]. For doing that, the SHAP values across the entire training dataset are computed and averaged afterwards. Then, the ranking of features is obtained by sorting the resulting aggregated SHAP values in decreased order, and the top-n features are chosen.

SHAP exhibits two notable properties as feature selection method: firstly, it identifies the most influential features through an interpretability-driven approach, thereby improving model transparency, and secondly, in contrast to other methods, such as those based on Mutual Information [26], it takes into account the interactions between features.

4. EXPERIMENTS AND RESULTS

4.1 Database and experimental protocol

The database used in this study is the "Alzheimer's Dementia Recognition through Spontaneous Speech" (ADReSS) dataset curated for the ADReSS Challenge [6]. It contains speech recordings and the corresponding transcripts of descriptions of the Cookie Theft picture by 156 speakers (78 AD and 78 non-AD). The number of subjects in the training and test sets is, respectively, 108 and 48.

The developed systems have been evaluated in terms of accuracy, F1-Score, specificity, and sensitivity. Also, we have measured the Clinical Utility Index (CUI) [27], a clinically oriented assessment that is divided into two parts: (1) Positive Utility (CUI+), computed as the product of the positive predictive value and sensitivity; and (2) Negative Utility (CUI-), computed as negative predictive value and specificity. The CUI is the average of CUI+ and CUI- weighted by the prevalence. There are four levels of diagnostic utility: "poor utility" (CUI $<$ 49%), "fair utility" (49% $\leq$ CUI $<$ 64%), "good utility" (64% $\leq$ CUI $<$ 81%), and "excellent utility" (CUI $\geq$ 81%).

4.2 Results of the single-modality systems

Following the pipeline depicted in Fig. 2(a), the SHAP values over the training set were computed using only one of the modalities (either audio or text). Fig. 3(a) and (b) are the summary plots of the 10 highest SHAP values for the acoustic and linguistic features, respectively. In these graphs, features are sorted by their global importance and each dot represents a single prediction for a feature. The x-axis shows the SHAP contribution, where positive values indicate that the feature pushes the prediction towards the positive class (AD), while negative values push it towards the negative class (non-AD). The color reflects the feature value (red for high and blue for low). For example, low values of the feature *speech rate* (i.e., the person speaks slowly) makes the system tends to predict AD.

As can be observed, for the acoustic modality, the more influential features are mainly related to speech fluency and pauses (*speech_rate*, *pause_time* and *pause_length_mean*), followed by prosodic characteristics (*intensity_median*, *intensity_q3*, *intensity_stddev*, *pitch_mean*, *absolute_pitch_slope_mean* and *voiced_fraction*). For the linguistic case, the frequency of appearance of concrete terms (*cookie*, *sink*, *mother*, *stool*, *run*), mental state terms (*see*) and filled pauses (*well*), the use of non-specific vocabulary (*it*) and the description of background details (*curtain*) are the most impactful features on the predictions.

Fig. 4 depicts the accuracy and F1-score achieved by the single-modality systems as a function of the percentage of features selected by SHAP. As can be seen, keeping around the 40% of features for the audio modality (13 out 34) and the 30% for the textual one (45 out 150) improves the corresponding baselines (when all features are used). Tab. 1 contains the results achieved by the baselines and the optima subsets of features for both modalities. As observed, all metrics either improve or remain unchanged when the interpretable feature selection is applied, showing that SHAP is able to choose the combination of features that are more informative for the AD task. As for the comparison between both modalities, linguistic features perform significantly better than the acoustic ones in all cases independently of the number of features selected.

4.3 Results of the multimodal system

In the multimodal system, early fusion was implemented by concatenating features from both modalities at the input level. Results for both scenarios, without and with feature selection, are reported in Tab. 1. As can be observed, while the direct use of all features does not surpass the performance of the linguistic-only modality, SHAP-based feature selection effectively identifies an optimal subset of acoustic and linguistic features, yielding superior results across all evaluated systems. In this configuration, only $\approx$ 20% of the original features (33 out of 184) are retained. Moreover, the multimodal system achieves the highest CUI, exhibiting an "excellent utlilty" level and thereby underscoring its potential clinical applicability.

Tab. 2 contains the best configuration of characteristics selected by SHAP in the multimodal system. As can be observed, these features have a clearly interpretable meaning and their observed behavior matches the typical dementia speech patterns trends described in Section 2.

Fig. 5 shows an example of the SHAP explanation for a test instance that has been classified as AD with an score of $f(x) = 0.631$. For each feature, its value (on the left side) and its positive (number inside the red bar) or negative (blue bar) contribution to AD prediction are shown. In this case, the large amount of filled pauses (high value of *well*), small presence of concrete terms (low values of *cookie* and *stool*), and monotonous voice (low value of *absolute_slope_pitch_mean*) pushes the model's prediction to AD, whereas the description of background details (high values of *outside*, *open* and *window*) pushes it to non-AD.

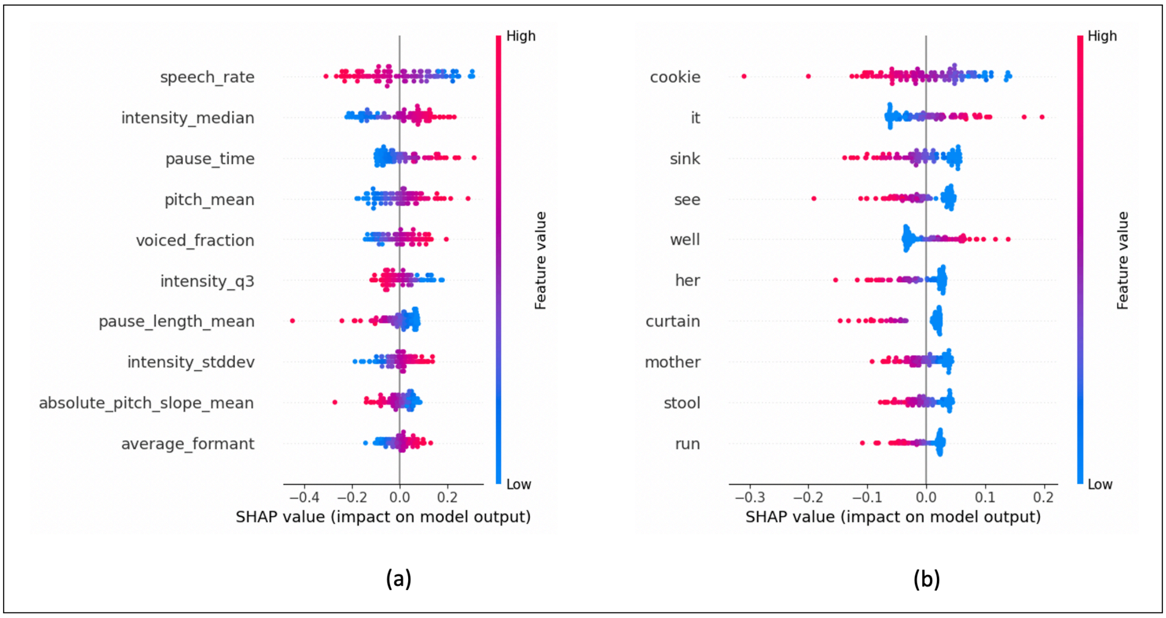

Figure 3. SHAP summary plots of single-modality SVM-based systems for AD detection: (a) Acoustic modality; (b) Linguistic modality.

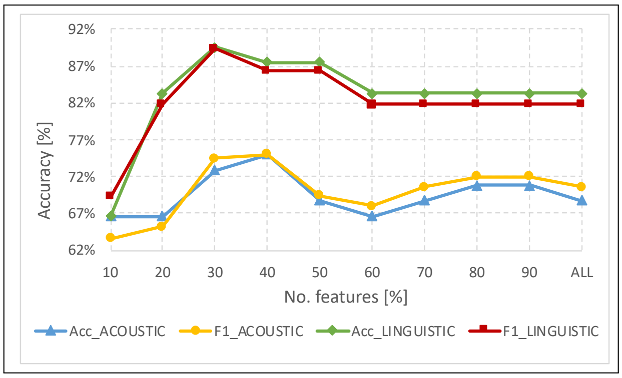

Figure 4. Accuracy and F1-score as a function of the percentage of selected features for the acoustic and linguistic modalities.

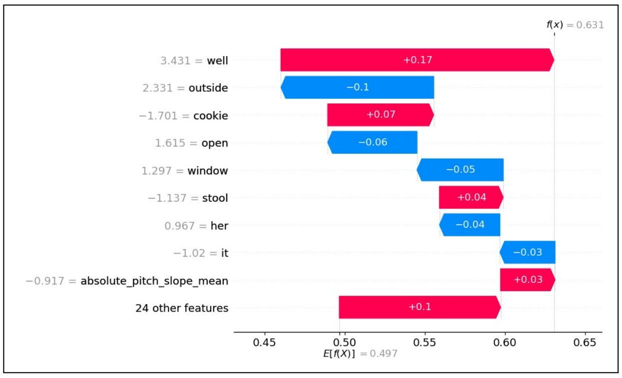

Figure 5. Local explanation based on SHAP values for an specific test instance.

5. CONCLUSIONS

In this paper, we have proposed an interpretable multi-modal dementia detection system from speech. It uses a reduced number of clinically and perceptual meaningful acoustic and linguistic features that have been chosen by means of the application of a SHAP-guided feature selection method. Results over the ADReSS dataset show that this approach not only enhances the system explainability and clinical utility but also improves its accuracy, F1-score, specificity and sensitivity, highlighting the potential of explainable feature selection to bridge ML techniques with clinically meaningful insights. For future work, we plan to extend our research to the other related tasks, such as AD progression prediction.

Table 1. Results of the SVM model for acoustic, linguistic and acoustic+linguistic modalities. ALL and FS stand, respectively, for the case where all the features and the selected features by SHAP are considered.

	No. features	No. features [%]	Accuracy	F1-score	Specificity	Sensitivity	CUI+	CUI-	CUI
Acoustic ALL	34	100	68.75	70.59	62.50	75.00	50.00	44.64	47.32
Acoustic FS	13	≈ 40	75.00	75.00	75.00	75.00	56.25	56.25	56.25
Linguistic ALL	150	100	83.33	81.82	91.67	75.00	67.50	72.02	69.76
Linguistic FS	45	30	89.58	89.36	91.67	87.50	79.89	80.67	80.28
Multimodal ALL	184	100	85.42	84.44	91.67	79.17	71.63	74.69	73.16
Multimodal FS	33	≈ 20	91.67	91.30	95.83	87.50	83.52	84.78	84.15

6. ACKNOWLEDGMENTS

The authors acknowledge the support of the Comunidad de Madrid through grant PEJ-2023-AI/SAL-GL-27629 and the Spanish State Research Agency (MICIU/AEI/10.13039/501100011033) through projects PID2021-126061OB-C43, PID2023-150584OB-C21 and PID2023-146684NB-I00.

7. REFERENCES

[1] World Health Organization: WHO, "Dementia." https://www.who.int/news-room/fact-sheets/detail/dementia, Mar. 2023.

[2] G. Prabhakaran and R. Bakshi, "Analysis of structure and cost in a longitudinal study of Alzheimer's disease," *J. Health Care Financ.*, May 2018.

[3] L. C. Kourtis, O. B. Regele, J. M. Wright, and G. B. Jones, "Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity," *npj Digit. Med.*, vol. 2, Feb 2019.

[4] I. Martínez-Nicolás, T. Llorente, F. Martínez-Sánchez, and J. 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," *Front. Psychol.*, vol. 12, Mar 2021.

[5] K. Ding, M. Chetty, A. N. Hoshyar, T. Bhattacharya, and B. Klein, "Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges," *Artif. Intell. Rev.*, vol. 57, 2024.

[6] S. Luz, F. Haider, S. D. L. F. Garcia, D. Fromm, and B. MacWhinney, "Editorial: Alzheimer's dementia recognition through spontaneous speech," *Front. Comput. Sci.-Switz*, vol. 3, Oct 2021.

[7] F. Eyben *et al.*, "The Geneva minimalistic acoustic parameter set (GEMAPS) for voice research and affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, pp. 190–202, Apr 2016.

[8] I. Oiza-Zapata and A. Gallardo-Antolín, "Alzheimer's disease detection from speech using Shapley additive explanations for feature selection and enhanced interpretability." Available at SSRN: http://dx.doi.org/10.2139/ssrn.5156692, 2025.

[9] M. Martinc and S. Pollak, "Tackling the adress challenge: a multimodal approach to the automated recognition of Alzheimer's dementia," in *Proc. of Interspeech 2020*, 2020.

[10] N. Wang, Y. Cao, S. Hao, Z. Shao, and K. Subbalakshmi, "Modular multi-modal attention network for Alzheimer's disease detection using patient audio and language data," in *Proc. of Interspeech 2021*, 2021.

[11] R. Pappagari, J. Cho, S. Joshi, L. Moro-Velázquez, P. Żelasko, J. Villalba, and N. Dehak, "Automatic detection and assessment of Alzheimer disease using speech and language technologies in low-resource scenarios," in *Proc. of Interspeech 2021*, 2021.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. of NIPS 2017*, p. 4768–4777, 2017.

[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, p. 1157–1182, Mar. 2003.

[14] H. Goodglass, E. Kaplan, and B. Barresi, *BDAE-3: Boston Diagnostic Aphasia Examination–Third Edition*. Philadelphia, PA: Lippincott Williams & Wilkins, 2001.

Table 2. Top features obtained by the SHAP-based feature selection method in the multimodal system.

Features	Definition	Observed Behavior	Interpretation Cue
Acoustic Features			
pause_time	Total time of pauses	↑	Hesitant speech and with difficult to find words
absolute_pitch_slope_mean	Average of the absolute pitch slope	↓	Monotonous voice
pitch_mean, pitch_median, pitch_q3	Mean, median and third quartile of the pitch	↑	High-pitched voice due to tension or stress
speech_rate	Speaking speed	↓	Slow, lethargic and disengaged speech
Linguistic Features			
curtain, window, outside, out_of, open	Terms	↓	No description of background details
sink, cookie, stool, mother, overflow, blow, dry, steal, hand, stand, reach, take	Terms	↓	Little reference to concrete concepts (persons, objects or actions)
her	Terms	↓	Incorrect use of referential pronouns
see, say, quiet	Terms	↓	Little use of mental language
here, it, this	Terms	↑	Use of non-specific vocabulary
well, oh	Terms	↑	Large number of filled and unfilled pauses
spill	Terms	↑	-

[15] J. J. 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?," *Dement. Geriatr. Cogn. Disord.*, vol. 37, no. (5-6), p. 327–334, 2014.

[16] C. Themistocleous, M. Eckerström, and D. Kokkinakis, "Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls," *PLoS One*, 2020.

[17] L. Cummings, "Describing the Cookie Theft picture," *Pragmat. Soc.*, vol. 10, no. 2, pp. 153–176, 2019.

[18] A. Favaro, N. Dehak, T. Thebaud, J. Villalba, E. Oh, and L. Moro-Velázquez, "Discovering invariant patterns of cognitive decline via an automated analysis of the Cookie Thief picture description task," in *Proc. of Odyssey 2024*, pp. 201–208, 2024.

[19] A. Favaro, A. Butala, T. Thebaud, J. Villalba, N. Dehak, and L. Moro-Velázquez, "Unveiling early signs of Parkinson's disease via a longitudinal analysis of celebrity speech recordings," *npj Parkinsons Dis.*, vol. 10, no. 1, p. 207, 2024.

[20] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A python interface to Praat," *J. Phon.*, vol. 71, no. 7, p. 15, 2018.

[21] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." `http://www.praat.org/`, 2021.

[22] L. Zhang, "DigiPsych Prosody." `https://github.com/NeuroLexDiagnostics/DigiPsych_Prosody`, 2019.

[23] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spaCy: Industrial-strength natural language processing in Python," 2020.

[24] F. Pedregosa, et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, 2011.

[25] W. E. Marcílio and D. M. Eler, "From explanations to feature selection: assessing shap values as feature selection mechanism," in *Proc. of SIBGRAPI*, 2020.

[26] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *J. Mach. Learn. Res.*, vol. 13, no. 2, 2012.

[27] M. A. J., "Sensitivity × PPV is a recognized test called the clinical utility index (CUI+)," *Eur. J. Epidemiol.*, vol. 26, no. 3, pp. 251–252, 2011.