Alzheimer's Disease Detection Through Automated Speech Analysis in Spanish-Speaking Populations through Feature Engineering and Machine Learning

Marcos Saade Romano
Tecnológico de Monterrey

***Abstract*—Early detection of Alzheimer's Disease (AD) is crucial for implementing effective intervention strategies. Speech analysis has emerged as a non-invasive and cost-effective approach for identifying cognitive impairments associated with AD. This study investigates the detection of AD in Spanish-speaking individuals using machine learning models, specifically Support Vector Machines (SVM). We utilize a Spanish-language dataset comprising audio recordings from individuals with Alzheimer's dementia, mild cognitive impairment, and a control group. The research presents a pipeline for feature engineering, feature selection, and binary classification. Our analysis highlights specific acoustic features that effectively distinguish AD speech patterns from non-AD patterns. The SVM model demonstrates promising performance in accurately classifying the presence of AD based on speech data. These findings underscore the potential of SVM models for AD detection through speech analysis in Spanish-speaking populations and emphasize the importance of particular acoustic characteristics.**

***Index Terms*—Alzheimer's Disease diagnosis, automatic speech analysis, reading task, class imbalance.**

I. Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline, memory loss, and language impairments [1], [2]. The global prevalence of AD is rising as populations age, creating significant challenges for healthcare systems worldwide [3], [4]. Early detection of AD is crucial for implementing interventions that can slow disease progression and improve patient outcomes.

Traditional diagnostic methods, such as neuroimaging and cerebrospinal fluid analysis, are often invasive, expensive, and not easily accessible in many settings [1], [5], [7]. As a result, there is a growing need for non-invasive, cost-effective, and accessible tools for early AD detection.

Speech analysis has emerged as one such tool, providing a promising alternative that leverages changes in speech patterns linked to cognitive decline [1], [8], [29]. Previous studies have highlighted that acoustic, complexity, and temporal features in speech can serve as potential markers for cognitive impairments associated with AD [8]–[10]. Despite advancements in this area, research focused on Spanish-speaking populations remains limited [1], [14].

This study aims to address this gap by employing Support Vector Machine (SVM) models for the automatic detection of AD using Spanish speech data. We utilized a dataset comprising 361 audio recordings from native Spanish speakers, with 287 labeled as non-AD (Healthy Controls and Mild Cognitive Impairment) and 74 as AD. To address class imbalance, the Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors (SMOTE-ENN) was employed. Feature selection was performed using Recursive Feature Elimination (RFE), resulting in 15 relevant acoustic, complexity, and temporal features. Permutation feature importance analysis was conducted to identify key contributors to model predictions. The model was evaluated through cross-validation, yielding promising results that underscore the potential of speech analysis as an effective diagnostic tool for AD in Spanish-speaking populations.

The rest of the paper is organized as follows: Section II discusses related work, Section III describes the data and preprocessing steps, Section IV outlines the feature extraction process, Section V presents the methodology, Section VI details the experiments and results, Section VII provides the discussion, Section VIII explores future work, and Section IX concludes the study.

II. Related Work

Speech analysis has gained significant attention as a promising tool for early detection of AD, providing a non-invasive alternative to traditional diagnostic techniques [1], [2]. Numerous studies have demonstrated the potential of acoustic features, such as Mel-Frequency Cepstral Coefficients (MFCCs), formant frequency statistics, and spectral centroid in identifying cognitive decline [11], [15], [16].

A. Acoustic Features as Biomarkers

MFCCs are widely used in speech processing to represent the short-term power spectrum of sound [16]–[18]. Studies have investigated their utility in distinguishing between healthy individuals and those with dementia [1], [15]. MFCCs can capture subtle changes in speech patterns associated with cognitive decline, making them valuable features for automated dementia detection systems.

Formant frequencies, the resonant frequencies of the vocal tract, are key in speech production, especially for vowel sounds. The second formant (F2) relates to tongue position, while the first formant (F1) corresponds to tongue height. Variations in these frequencies can reveal speech motor control issues, often observed in cognitive or neurological impairments like dementia. Such formant variations have been studied as non-invasive diagnostic markers for detecting speech and language deficits linked to cognitive conditions [6], [33].

Spectral centroid indicates the "center of mass" of the spectrum and is perceived as the brightness of a sound. Alterations in spectral centroid values can signify changes in speech production mechanisms. Research has demonstrated that individuals with dementia may exhibit shifts in spectral centroid, reflecting modifications in speech characteristics due to cognitive decline [19], [32].

Additionally, complexity measures such as the Higuchi Fractal Dimension (HFD) have been used to quantify the fractal properties and complexity of speech signals [11], [19].

B. Temporal Features and Speech Rate

Temporal features, including pause rate, speech rate, and segment durations, play a crucial role in differentiating AD patients from healthy individuals [1], [20], [21]. AD patients often exhibit longer pauses and slower speech rates, reflecting difficulties in cognitive processing and language production.

C. Machine Learning Models in AD Detection

Various machine learning models, such as SVMs, K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Decision Trees (DT), have been employed to detect AD using speech features [11], [22], [23]. SVMs have shown effectiveness due to their ability to handle high-dimensional data and find optimal hyperplanes for classification [39].

Luz *et al.* [13] introduced the ADReSS Challenge, a benchmark dataset aiming to compare various machine learning models, focusing on speech features for automated Alzheimer's recognition, thereby highlighting the effectiveness of speech-based machine learning models in AD detection.

D. Use of Spanish Speech Databases

Several studies have utilized Spanish speech databases for dementia research, addressing the gap in research focusing on Spanish-speaking populations. Below, we summarize key studies that have leveraged Spanish speech data to classify dementia stages using various machine learning methodologies.

He et al. (2023) [34] conducted a study involving 119 subjects, comprising 76 individuals diagnosed with Alzheimer's Disease (AD) and 43 healthy controls (HC). The researchers employed a scene construction task to elicit speech and utilized a Random Forest classifier for analysis. The model achieved a high F1 score exceeding 0.9, with voice quality features proving to be the most effective indicators for classification.

García-Gutiérrez et al. (2024) [35] analyzed data from 1,373 participants, including 817 AD patients, 463 individuals with Mild Cognitive Impairment (MCI), and 93 healthy controls. Using a spontaneous speech protocol, various machine learning models were applied to the data. The study reported an F1 score of 0.92 for AD detection and found a significant correlation between the model's predictions and cognitive scores, underscoring the robustness of speech-based features in identifying cognitive decline.

Kaser et al. (2024) [36] explored speech elicitation through animal fluency, alternating fluency, and phonemic "F" fluency tasks with 174 subjects (78 AD patients and 96 healthy controls). Machine learning models were utilized to distinguish between normal and impaired cognitive states. The models achieved an Area Under the Curve (AUC) of 0.93 and an overall accuracy of 88.4%, demonstrating the effectiveness of fluency-based tasks in dementia classification.

E. Use of Reading Tasks in Dementia Detection

Reading tasks have been effective for eliciting speech patterns that differentiate individuals with dementia from healthy controls. Ivanova *et al.* [27] investigated linguistic features in speech during reading tasks and identified significant differences in syntactic complexity and semantic content between AD patients and controls.

Martínez-Nicolás *et al.* [8] examined speech characteristics during reading tasks and noted that individuals with AD exhibited more frequent pauses, slower speech rates, and reduced articulation accuracy compared to healthy controls.

F. Importance of Feature Selection Methods

Feature selection methods like Recursive Feature Elimination (RFE) are useful for improving model performance by eliminating irrelevant or redundant features [28]. Darst *et al.* [28] demonstrated the effectiveness of RFE in handling correlated variables in high-dimensional data, which is pertinent to speech analysis for AD detection.

III. Data and Preprocessing

A. Dataset Description

The dataset consists of 361 audio recordings from native European Spanish speakers aged between 50 and 96 years. The participants were categorized into two classes: 287 as non-AD (197 Healthy Controls and 90 with Mild Cognitive Impairment) and 74 as AD. All diagnoses were confirmed by clinical assessments in accordance with the criteria established by the Spanish National Health System [26], [37].

B. Audio Preprocessing

To improve the quality of the audio recordings and ensure the reliability of extracted features, the following preprocessing steps were applied:

- **Amplitude Normalization**: The audio signals were normalized to achieve a consistent amplitude level across all recordings.
- **Noise Reduction**: Spectral subtraction and Wiener filtering techniques were employed to reduce background noise.
- **Peak Reduction**: Sudden peaks in the audio signals were smoothed to prevent distortion in feature extraction.
- **Voice Activity Detection (VAD)**: A VAD algorithm was used to detect voiced and unvoiced (silence) segments, considering valid silences of at least 0.5 seconds.

IV. Feature Extraction

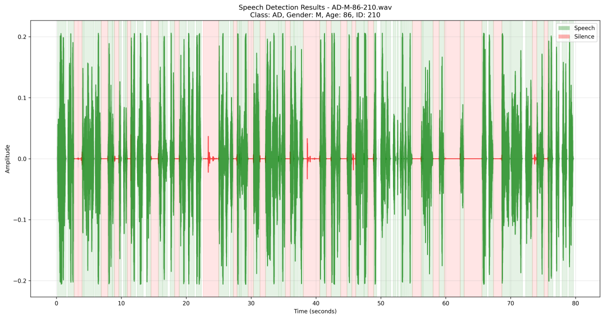
A comprehensive set of acoustic, prosodic, and complexity features was extracted from the preprocessed audio signals. These features are essential for capturing various aspects of speech that may indicate cognitive decline associated with conditions such as Alzheimer's disease (AD). The extracted features include:

- **Voice Quality Features**:
  - **Jitter and Shimmer**: Frequency and amplitude perturbations were measured, including jitter (local, ppq5) and shimmer (local, apq5). These metrics reflect irregularities in vocal fold vibrations and are associated with voice quality [38].
  - **Cepstral Peak Prominence Smoothed (CPPS)**: CPPS was calculated to assess the harmonic structure of the voice signal, providing insights into voice quality [39].
- **Formant Frequencies**:
  - Mean, standard deviation, and range of the first four formants (*F1–F4*) were extracted. These features reflect vowel articulation characteristics and resonant frequencies of the vocal tract [6], [33].
  - The bandwidth of the third formant (*F3_B3*) was calculated, providing additional information on vocal tract characteristics [26].
- **Spectral Features**:
  - **Mel-Frequency Cepstral Coefficients (MFCCs)**: The mean and standard deviation of the first 13 MFCCs were computed, capturing detailed spectral properties of speech. MFCCs effectively capture subtle changes in speech patterns associated with cognitive decline. [1], [15].
  - **Spectral Slope and Centroid**: The spectral slope was calculated to represent the tilt of the speech spectrum, while the spectral centroid indicated the center of mass of the spectrum. Alterations in these features can signal changes in speech production mechanisms [19], [32].
- **Harmonics-to-Noise Ratio (HNR)**: The mean HNR was extracted using the autocorrelation method, providing a measure of voice quality by quantifying the ratio of harmonic components to noise [26], [33].
- **Pitch Features**: The mean and standard deviation of the fundamental frequency (*pitch_mean*, *pitch_std*) were computed to assess variations in pitch, which can indicate prosodic abnormalities in speech [26], [40].
- **Amplitude Features**:
  - **Average Amplitude**, **Peak Amplitude**, and **Amplitude Variance**: These features capture variations in loudness and signal dynamics.
  - **Amplitude Minimum** and **Amplitude Maximum Difference Mean**: These metrics provide insights into the range and variability of the speech signal amplitude [26].
- **Temporal Features**:
  - **Timing Features**: Utilizing Voice Activity Detection (VAD), features such as total duration, total speech duration, silence count, speech segment count, and various pause statistics like maximum duration and standard deviation of pause lengths were extracted. These features capture speech timing irregularities and are significant in assessing speech fluency.
  - **Speech Rate and Articulation Rate**: These rates were calculated to assess the speed of speech production, which may slow in individuals with cognitive decline.
  - **Speech-to-Pause Ratios**: Ratios between speech and pause durations were computed to provide insights into speech continuity and fluency [41].
- **Rhythm Features**:
  - **Raw Pairwise Variability Index (rPVI)** and **Normalized Pairwise Variability Index (nPVI)**: These metrics quantify the variability in speech timing, reflecting rhythmic patterns that may change due to cognitive impairment [42].
- **Complexity Measures**:
  - **Higuchi's Fractal Dimension (HFD)**: Statistical measures including mean, maximum, minimum, standard deviation, and variance of HFD were computed to analyze the complexity of the speech signal. HFD quantifies the fractal properties of time-series data and has shown potential in distinguishing healthy individuals from those with AD. [11], [19]
- **Additional Features**:
  - **Asymmetry**: The skewness of the amplitude distribution was calculated to assess asymmetry in the speech signal.
  - **Trajectory Intra (TrajIntra)**: Mean absolute differences of the signal were computed to capture signal variability.
  - **Acoustic Voice Quality Index HNR_sd (AVQI_HNR_sd)**: The standard deviation of the Root Mean Square (RMS) energy was calculated, providing an estimate of voice quality. [26]
- **Silence/Voice Segment Visualization**:
  Fig. 1 depicts the differences in silence and voice segments between AD and HC participants, highlighting the importance of temporal features.

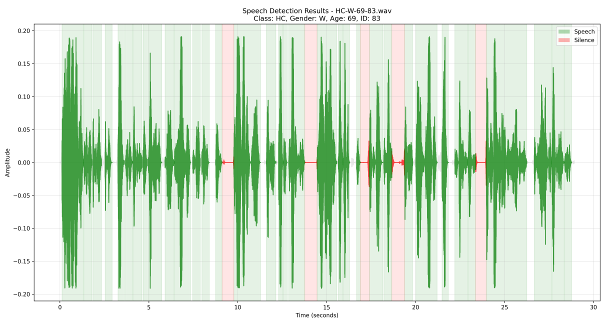
V. Methodology

A. Pipeline Overview

Our methodology consists of several key steps, as illustrated in Fig. 2. The pipeline begins with data collection and preprocessing, followed by feature extraction. To address class imbalance, SMOTE-ENN is applied before feature selection and model training using SVM.



(a) AD Subject



(b) HC Subject

Fig. 1: Visualization of Silence and Voice Segments for AD and HC Participants. The figures illustrate the amplitude over time, where green sections represent speech segments and red-highlighted areas correspond to silent intervals. Panel (a) shows an audio excerpt from an AD subject, revealing more frequent and prolonged silent intervals. Panel (b) presents a sample from an HC subject, which displays shorter and less frequent silences, indicative of more continuous speech. These visual differences emphasize the potential significance of temporal features such as pause rate, total silence duration, and speech-to-silence ratio in distinguishing between AD and HC cases.

B. Handling Class Imbalance with SMOTE-ENN

Class imbalance is a common issue in medical datasets, where the number of cases in one class significantly outnumbers the other. In our dataset, the non-AD class has 287 samples, while the AD class has only 74 samples. This imbalance can lead to biased model performance favoring the majority class.

To mitigate this issue, we employed the Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors (SMOTE-ENN) [30], [31]. SMOTE-ENN is a hybrid method that combines oversampling the minority class using SMOTE and cleaning the data using ENN. Specifically:

- **SMOTE**: Generates synthetic samples of the minority class (AD cases) by interpolating between existing minority instances and their nearest neighbors.
- **ENN**: Removes samples (from both classes) that are misclassified by their nearest neighbors, thus cleaning overlapping regions and reducing noise.

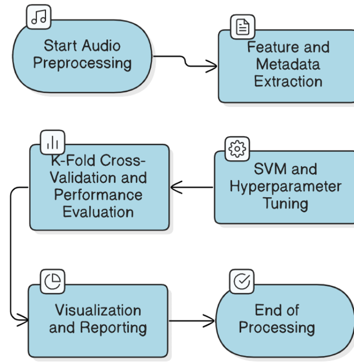


Fig. 2: Pipeline Diagram Illustrating the Preprocessing Steps, SMOTE-ENN Application, Feature Extraction, and SVM Classification

Applying SMOTE-ENN resulted in a more balanced and cleaner dataset, enabling the SVM model to learn decision boundaries more effectively and improve generalization. This approach has been shown to enhance classification performance in imbalanced medical datasets [30], [31].

To visually illustrate the impact of applying SMOTE-ENN on our dataset, we include a bar graph comparing the class distribution before and after resampling, as shown in Fig. 3.

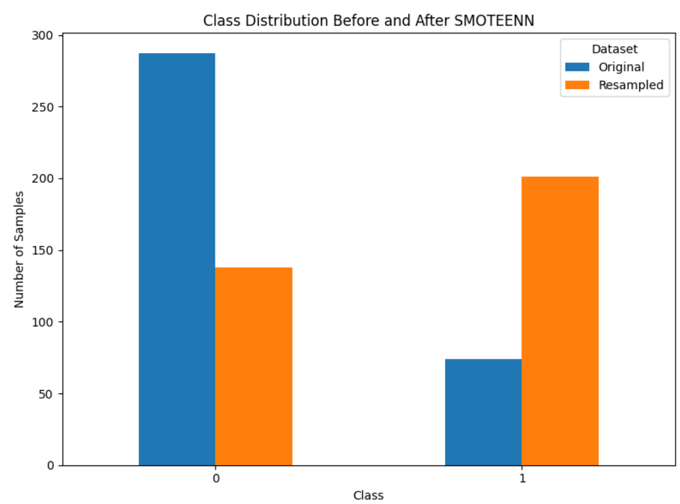


Fig. 3: Class distribution before and after applying SMOTE-ENN. The original distribution shows a significant imbalance between the non-AD class and the AD class. After applying SMOTE-ENN, the distribution is more balanced, with fewer overlapping and noisy samples.

The graph highlights how SMOTE-ENN effectively oversamples the minority class while also reducing noise and cleaning up overlapping samples in both classes. This step is crucial in ensuring that the model is not biased towards the majority class and can better generalize to new data.

C. Feature Selection

Recursive Feature Elimination (RFE) with a Random Forest estimator was used for feature selection. RFE iteratively removes the least important features based on feature importance scores until the optimal subset of features is obtained [28]. The selected features were:

- F2_range
- spectral_centroid
- mfcc_2_mean
- mfcc_3_mean
- mfcc_5_mean
- mfcc_6_mean
- mfcc_7_mean
- mfcc_7_std
- mfcc_8_mean
- mfcc_11_mean
- hnr_mean
- HFD_min
- total_duration
- total_speech_duration
- speech_duration_coefficient_of_variation

After feature selection, permutation feature importance was used to rank features based on their contribution to the SVM model's performance (see Figure 5). This method measures how much the model's prediction error increases when a feature's values are randomly shuffled, breaking its relationship with the target variable.

This approach helped identify which speech features were most critical for distinguishing between dementia and non-dementia cases. Features that caused a significant drop in model performance when permuted were considered highly important.

D. Support Vector Machine Model

A Support Vector Machine (SVM) with a radial basis function (RBF) kernel was employed due to its effectiveness in handling nonlinear data [24]. The hyperparameters of the SVM, including the regularization parameter C and the kernel coefficient γ, were optimized using grid search. For the hyperparameter optimization process, the F1 score was selected as the evaluation metric.

Maximizing the F1 score was chosen because it provides a balanced measure that considers both precision and recall. This is particularly important in scenarios where the class distribution is imbalanced or when both false positives and false negatives carry significant consequences.

The grid search procedure systematically explored the parameter space, and the configuration that yielded the highest F1 score on the validation set was selected. The resulting optimal parameters are as follows:

- **C (Regularization Parameter)**: 10
- **Kernel**: RBF
- **Gamma**: 'scale'

E. Evaluation Metrics

The model's performance was evaluated using the following metrics, chosen to provide a comprehensive analysis of its effectiveness across multiple dimensions:

- **Accuracy**:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives. Accuracy provides an overall indication of the proportion of correct predictions made by the model. While it is a useful initial metric, it can be misleading in the presence of class imbalance, which is relevant in the context of dementia diagnosis where the distribution between classes may not be even.

- **Precision**:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision was selected to evaluate the model's ability to correctly identify true positive cases without including false positives. This metric is particularly important for the dementia classification task, as false positive diagnoses can lead to unnecessary stress and follow-up procedures for patients who are actually healthy.

- **Recall** (Sensitivity):

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall is essential in the context of dementia detection, where it is crucial to identify as many true positive cases as possible. High recall ensures that the model effectively captures most cases of Alzheimer's Disease (AD), minimizing the risk of false negatives, which could result in missed diagnoses and delayed treatment.

- **F1-Score**:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-Score was included to provide a balance between precision and recall, offering a single metric that accounts for both false positives and false negatives. This is particularly useful when there is an uneven distribution of classes, as it provides a better sense of the model's performance in capturing both key aspects.

This set of metrics allows for a nuanced evaluation of the model, ensuring that it is not only accurate but also reliable in correctly diagnosing Alzheimer's Disease while minimizing errors that could impact patient outcomes.

Cross-validation with 5 folds was used to assess the model's performance.

VI. Experiments and Results

A. Cross-Validation Performance

The SVM model achieved the following cross-validation metrics:

- **Accuracy**: $80.06\% \pm 2.20\%$
- **F1-Score (Weighted)**: $81.19\% \pm 2.00\%$
- **Recall (Weighted)**: $80.06\% \pm 2.20\%$
- **Precision (Weighted)**: $83.48\% \pm 1.79\%$

B. Classification Report on the Entire Dataset

After training on the entire dataset, the SVM model produced the following classification report:

TABLE I: Classification Report

Class	Precision	Recall	F1-Score	Support
Non-AD	0.95	0.87	0.91	287
AD	0.62	0.82	0.71	74
Accuracy			0.86	
Macro Avg	0.78	0.85	0.81	361
Weighted Avg	0.88	0.86	0.87	361

C. Confusion Matrix

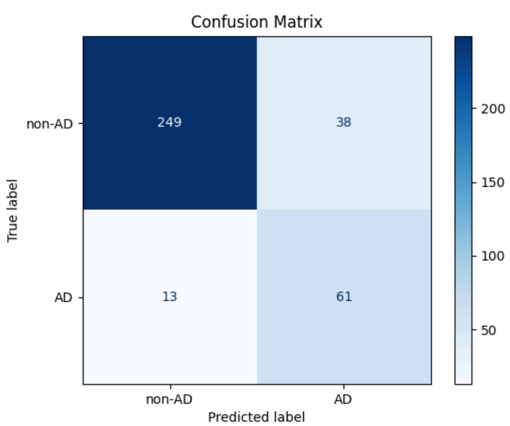


Fig. 4: Confusion Matrix of the SVM Model

The confusion matrix in Fig. 4 shows that the model correctly classified 249 out of 287 non-AD cases and 61 out of 74 AD cases. There were 38 false positives (non-AD cases incorrectly classified as AD) and 13 false negatives (AD cases incorrectly classified as non-AD). This indicates strong performance in identifying non-AD individuals but highlights some difficulty in detecting all AD cases.

D. Misclassification Analysis

To gain deeper insights into the model's performance, we analyzed the misclassifications for each true class. Specifically, among the Non-AD cases, which include Healthy Control (HC) and Mild Cognitive Impairment (MCI) individuals, there were a total of 38 misclassifications:

- **Healthy Control (HC)**: 17 instances were incorrectly classified as Alzheimer's Disease (AD).
- **Mild Cognitive Impairment (MCI)**: 21 instances were incorrectly classified as AD.

Additionally, for the AD class:

- **Alzheimer's Disease (AD)**: 13 instances were incorrectly classified as Non-AD.

This breakdown indicates that while the model demonstrates strong performance in identifying AD cases, there is a significant number of Non-AD instances (both HC and MCI) being misclassified as AD. The higher misclassification rate from Non-AD to AD suggests overlapping features between these groups, particularly between MCI and AD, which may pose challenges for the model's discriminative ability. Addressing this overlap through feature engineering or employing more sophisticated classification techniques could potentially enhance the model's accuracy in distinguishing between these classes.

E. Feature Importance Analysis

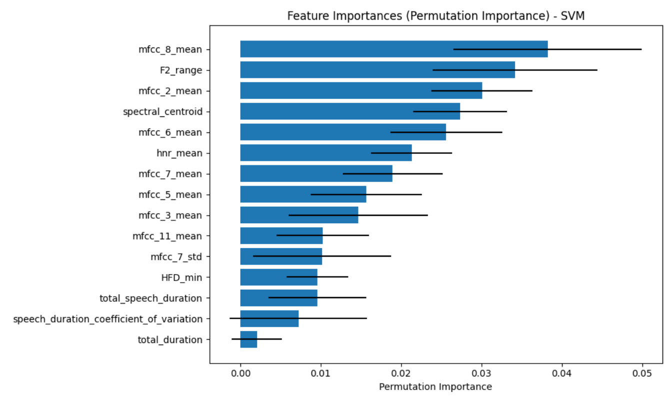


Fig. 5: Permutation Feature Importance of the SVM Model

Permutation feature importance was computed to identify the most influential features contributing to the SVM model's predictions. The features were ranked based on the decrease in model performance when each feature's values were randomly shuffled. Fig. 5 illustrates the importance of each feature, with error bars representing the standard deviation over multiple shuffles.

The top features identified were:

- **mfcc_8_mean**: Mean of the 8th MFCC, capturing higher-order spectral characteristics.
- **F2_range**: Range of the second formant frequency, associated with vowel articulation.
- **mfcc_2_mean**: Mean of the 2nd MFCC, reflecting spectral envelope properties.
- **spectral_centroid**: Represents the center of mass of the spectrum, indicating brightness of the sound.
- **mfcc_6_mean**: Mean of the 6th MFCC.

Features such as **HFD_min**, **total_duration**, and **speech_duration_coefficient_of_variation** had lower importance, suggesting that temporal and complexity features were less influential in the model's predictions compared to specific acoustic features.

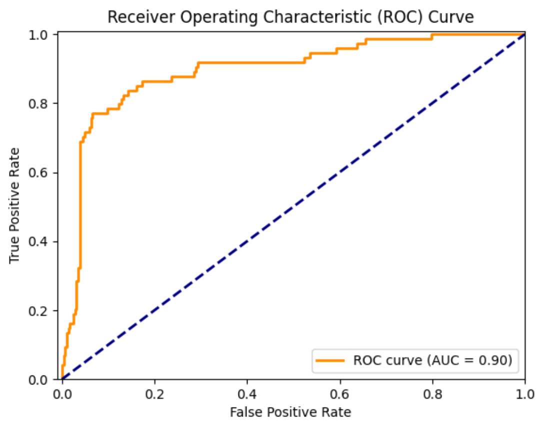


Fig. 6: ROC Curve of the SVM Model

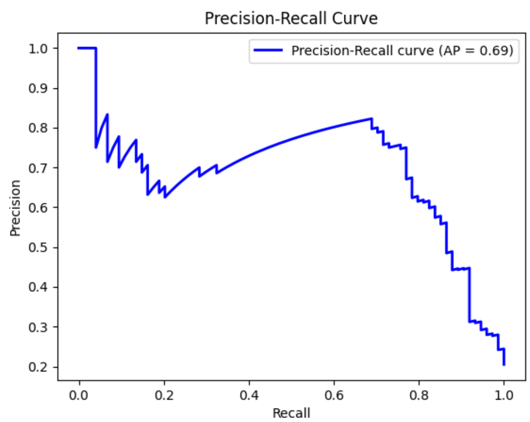


Fig. 7: Precision-Recall Curve of the SVM Model

F. ROC and Precision-Recall Curves

The ROC curve in Fig. 6 and the Precision-Recall curve in Fig. 7 illustrate the model's discriminative ability and its performance in handling class imbalance. The area under the ROC curve (AUC) was 0.90, indicating strong ability to distinguish between classes. The area under the Precision-Recall curve (AP) was 0.69, reflecting decent performance in detecting true positives while balancing false positives.

G. Other Models

To validate the generalizability of the 15 selected features, additional models were tested and compared. These models included Artificial Neural Network (ANN), and XGBoost. The purpose of this comparison was to understand how well these features performed across different algorithms beyond the primary SVM model used in earlier analyses.

As shown in Fig. 8, the SVM model displayed the best balance between accuracy and F1 Score, reinforcing its robustness with the selected 15 features. The ANN model followed closely, demonstrating that these features are also effective in a neural network context. Although XGBoost had slightly lower metrics, its performance was still competitive, confirming that the chosen feature set carries valuable information for different model types.

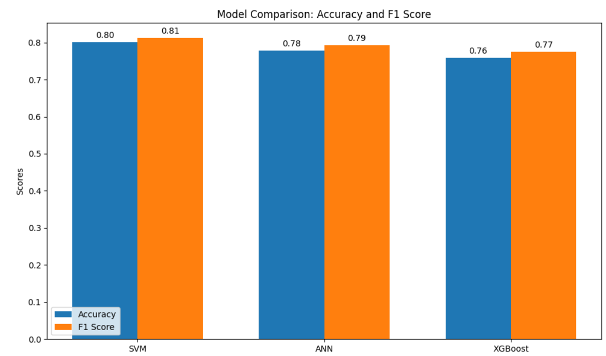


Fig. 8: Comparison of model performance in terms of Accuracy and F1 Score for SVM, ANN, and XGBoost. The SVM model achieved an accuracy of 0.80 and an F1 Score of 0.81, showing strong classification performance. The ANN model reached an accuracy of 0.78 with an F1 Score of 0.79, closely following the SVM. XGBoost, while performing reasonably well, achieved an accuracy of 0.76 and an F1 Score of 0.77, slightly lower than the other models.

VII. Discussion

The SVM model demonstrated strong performance in detecting AD using Spanish speech data. The high precision for non-AD cases indicates that the model is effective at identifying healthy individuals, while the recall for AD cases shows it can detect a significant proportion of true AD cases.

SMOTE-ENN Application: The use of SMOTE-ENN effectively addressed the class imbalance in the dataset. By generating synthetic samples of the minority class and cleaning overlapping regions, the model's ability to detect AD cases improved. This approach aligns with findings from previous studies that highlight the benefits of SMOTE-ENN in medical data classification [30], [31].

Feature Importance: The feature importance analysis revealed that acoustic features, particularly the mean values of specific MFCCs, **F2_range**, and **spectral_centroid**, played a significant role in the model's ability to distinguish between AD and non-AD cases. The prominence of **mfcc_8_mean** suggests that higher-order spectral features are critical in capturing the subtle changes in speech associated with AD. MFCCs have been shown to capture changes in the short-term power spectrum of sound, reflecting alterations in vocal tract configurations due to cognitive decline [1], [11].

The importance of **F2_range** aligns with the understanding that vowel articulation is affected in individuals with AD due to motor speech impairments [1]. Variations in **F2_range** may reflect articulatory deficits, as changes in tongue position during vowel production are common in dementia.

Similarly, the **spectral_centroid** feature indicates shifts in the "brightness" of the sound, which can result from modifications in speech production mechanisms caused by cognitive decline [32]. These acoustic features collectively contribute to capturing the speech characteristics associated with AD.

Limitations: While the SVM model showed promising results, there are limitations to consider. The dataset size, particularly the number of AD cases (74), is relatively small, which may affect the model's ability to generalize. Additionally, the model exhibited some difficulty in distinguishing AD cases, as indicated by the lower precision for the AD class and the confusion matrix results.

VIII. Future Work

Future research should aim to enhance model robustness by utilizing larger and more diverse datasets. The integration of linguistic features may further improve detection accuracy. Investigating temporal dependencies in voice features, particularly in MFCCs, could provide deeper insights. Additionally, exploring the use of more advanced models, such as transformers, known for their effectiveness in sequence analysis, may yield better results. Expanding the availability of Spanish-language data, including different regional variations such as Latin American and Caribbean Spanish, is also recommended to improve model generalizability and performance across varied Spanish-speaking populations.

IX. Conclusion

This study highlights the efficacy of support vector machines for detecting Alzheimer's Disease (AD) using Spanish speech data. It outlines the comprehensive pipeline employed, including audio preprocessing, feature extraction and selection, and model training. By integrating SMOTE-ENN to address class imbalance, we enhanced the model's capability to identify AD cases effectively. The model achieved a cross-validation accuracy of 80%. Analysis of feature importance revealed that specific acoustic features, such as **mfcc_8_mean**, **F2_range**, and **spectral_centroid**, were key differentiators between AD and non-AD cases.

The generalizability of the 15 selected acoustic, temporal, and complexity features was validated using additional models, including artificial neural networks (ANN) and XGBoost.

These findings support the potential of speech analysis as a non-invasive and practical diagnostic tool for AD, particularly within Spanish-speaking communities where research remains limited.

Annex I: Extracted Features

In this annex, we provide an overview of the extracted features used for the dementia classification research. The features span across various acoustic, timing, and complexity aspects:

- **Voice Quality Features**:
  - Jitter (local, ppq5)
  - Shimmer (local, apq5)
  - CPPS (Cepstral Peak Prominence Smoothed)
- **Formant Features**:
  - F1, F2, F3, F4 (mean, standard deviation, range, coefficient of variation)
  - F3_B3, F1_sd
- **Spectral Features**:
  - Spectral slope
  - Spectral centroid
- **MFCC (Mel-frequency Cepstral Coefficients)**:
  - MFCC 1–13 (mean and standard deviation)
- **Harmonics-to-Noise Ratio (HNR)**:
  - Mean HNR
- **Amplitude and Intensity Features**:
  - Amplitude variance, average amplitude, peak amplitude
  - Amplitude asymmetry, trajectory intra-analysis (TrajIntra)
  - Amplitude maximum difference (mean), amplitude minimum
- **Complexity Features**:
  - Higuchi Fractal Dimension (mean, max, min, standard deviation, variance)
- **Pitch Features**:
  - Mean pitch, standard deviation of pitch
- **Timing and Duration Features**:
  - Total duration, silence count, speech segment count
  - Total silence duration, total speech duration
  - Speech rate, articulation rate
  - Mean, standard deviation, maximum, and minimum pause durations
  - Pause rate, pause ratio
  - Mean, standard deviation, maximum, and range of speech duration
  - Speech duration coefficient of variation, speech-to-pause ratio
- **Rhythm Features**:
  - rPVI (raw Pairwise Variability Index)
  - nPVI (normalized Pairwise Variability Index)
- **Additional Metadata**:
  - Filename, age, gender, ID, class label

References

[1] 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 in Psychology*, vol. 12, p. 620251, 2021.

[2] I. Vigo, M. R. Lourenço, J. R. Rato, and A. C. Teixeira, "Speech- and language-based classification of Alzheimer's disease: A systematic review," *Bioengineering*, vol. 9, no. 3, p. 125, 2022.

[3] M. Prince, A. Wimo, M. Guerchet, G. C. Ali, Y. T. Wu, and M. Prina, *World Alzheimer report 2015: The global impact of dementia*. Alzheimer's Disease International, 2015.

[4] B. Winblad *et al.*, "Defeating Alzheimer's disease and other dementias: a priority for european science and society," *The Lancet Neurology*, vol. 15, no. 5, pp. 455–532, 2016.

[5] K. Blennow and H. Zetterberg, "Biomarkers for Alzheimer's disease: current status and prospects for the future," *Journal of Internal Medicine*, vol. 284, no. 6, pp. 643–663, 2018.

[6] M. Zolnoori, A. Zolnoori, and M. Topaz, “ADscreen: A speech processing-based screening system for automatic identification of patients with Alzheimer’s disease and related dementia,” *Artificial Intelligence in Medicine*, vol. 143, p. 102624, 2023.

[7] K. K. Leung *et al.*, “Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s disease,” *NeuroImage*, vol. 51, no. 4, pp. 1345–1359, 2010.

[8] I. Martínez-Nicolás, T. E. Llorente, F. Martínez-Sánchez, and J. J. G. Meilán, “Speech biomarkers of risk factors for vascular dementia in people with mild cognitive impairment,” *Frontiers in Human Neuroscience*, vol. 16, p. 1057578, 2022.

[9] L. Tóth *et al.*, “Automatic detection of mild cognitive impairment from spontaneous speech using ASR,” in *Proc. Interspeech*, 2015, pp. 2694–2698.

[10] S. Ahmed, A.-M. F. Haigh, C. A. de Jager, and P. Garrard, “Connected speech as a marker of disease progression in autopsy-proven Alzheimer’s disease,” *Brain*, vol. 136, no. 12, pp. 3727–3737, 2013.

[11] K. López-de Ipiña *et al.*, “On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis,” *Sensors*, vol. 13, no. 5, pp. 6730–6745, 2013.

[12] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, “Linguistic features identify Alzheimer’s disease in narrative speech,” *Journal of Alzheimer’s Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[13] S. Luz *et al.*, “Alzheimer’s dementia recognition through spontaneous speech: The ADReSS challenge,” in *Proc. Interspeech*, 2020, pp. 2172–2176.

[14] O. Ivanova *et al.*, “Discriminating speech traits of Alzheimer’s disease assessed through a corpus of reading task for Spanish language,” *Computer Speech & Language*, vol. 73, p. 101341, 2022.

[15] J. Laguarta and B. Subirana, “Longitudinal speech biomarkers for automated Alzheimer’s detection,” *Frontiers in Computer Science*, vol. 3, p. 624694, 2021.

[16] A. Satt, R. Hoory, A. König, P. Aalten, and P. H. Robert, “Speech-based automatic and robust detection of very early dementia,” in *Proc. Interspeech*, 2013, pp. 1692–1696.

[17] S. Liu *et al.*, “Early diagnosis of Alzheimer’s disease with deep learning,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 1015–1018.

[18] T. Alhanai, R. Au, and J. Glass, “Spoken language biomarkers for detecting cognitive impairment,” in *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017, pp. 409–416.

[19] K. Lopez-de Ipiña *et al.*, “Feature selection for automatic analysis of emotional response based on nonlinear speech modeling suitable for diagnosis of Alzheimer’s disease,” *Neurocomputing*, vol. 150, pp. 392–401, 2015.

[20] J. J. G. Meilán *et al.*, “Speech in Alzheimer’s disease: can temporal and acoustic parameters discriminate dementia?” *Dementia and Geriatric Cognitive Disorders*, vol. 37, no. 5–6, pp. 327–334, 2014.

[21] B. Roark *et al.*, “Spoken language derived measures for detecting mild cognitive impairment,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2081–2090, 2011.

[22] M. R. Kumar *et al.*, “Dementia detection from speech using machine learning and deep learning architectures,” *Sensors*, vol. 22, no. 23, p. 9311, 2022.

[23] Z. Jahan, S. B. Khan, and M. Saraee, “Early dementia detection with speech analysis and machine learning techniques,” *Discover Sustainability*, vol. 5, no. 1, pp. 1–18, 2024.

[24] L. Tóth *et al.*, “A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech,” *Current Alzheimer Research*, vol. 15, no. 2, pp. 130–138, 2018.

[25] R. Ossewaarde, R. Jonkers, F. Jalvingh, and R. Bastiaanse, “Classification of spontaneous speech of individuals with dementia based on automatic prosody analysis using support vector machines (SVM),” in *The Thirty-Second International Flairs Conference*, 2019.

[26] F. Martínez-Sánchez, J. J. G. Meilán, J. Carro, and O. Ivanova, “A prototype for the voice analysis diagnosis of Alzheimer’s disease,” *Journal of Alzheimer’s Disease*, vol. 64, no. 2, pp. 473–481, 2018.

[27] O. Ivanova *et al.*, “Reading and lexical–semantic retrieval tasks outperform single-task speech analysis in the screening of mild cognitive impairment and Alzheimer’s disease,” *Scientific Reports*, vol. 13, p. 9728, 2023.

[28] B. F. Darst, K. C. Malecki, and C. D. Engelman, “Using recursive feature elimination in random forest to account for correlated variables in high dimensional data,” *BMC Genetics*, vol. 19, no. 1, pp. 1–6, 2018.

[29] J. Liu, F. Fu, L. Li, J. Wang, and L. Zhang, “Efficient pause extraction and encode strategy for Alzheimer’s disease detection using only acoustic features from spontaneous speech,” *Brain Sciences*, vol. 13, no. 3, p. 477, 2023.

[30] M. Muntasir Nishat, F. Faisal, I. Jahan Ratul, A. Al-Monsur, A. M. Ar-Rafi, S. M. Nasrullah, *et al.*, “A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset,” *Scientific Programming*, vol. 2022, p. 3649406, 2022.

[31] M. Lamari, N. Azizi, N. E. Hammami, A. Boukhamla, S. Cheriguene, N. Dendani, and N. E. Benzebouchi, “SMOTE–ENN-based Data Sampling and Improved Dynamic Ensemble Selection for Imbalanced Medical Data Classification,” in *Advances on Smart and Soft Computing: Proceedings of ICACIn 2020*, Springer Singapore, pp. 37–49, 2021.

[32] M. L. B. Pulido, J. B. A. Hernández, M. Á. F. Ballester, C. M. T. González, J. Mekyska, and Z. Smékal, “Alzheimer’s disease and automatic speech analysis: a review,” *Expert Systems with Applications*, vol. 150, p. 113213, 2020.

[33] K. Nishikawa, H. Kawano, R. Hirakawa, and Y. Nakatoh, “Analysis of Prosodic Features and Formant of Dementia Speech for Machine Learning,” in *2022 5th International Conference on Information and Computer Technologies (ICICT)*, 2022, pp. 173–176, IEEE.

[34] He, R., Marquié, M., Sanabria, A., Alegret, M., Rosende-Roca, M., & Valero, S. (2023). Automated classification of cognitive decline and probable Alzheimer’s dementia from speech in Spanish/Catalan. *American Journal of Speech-Language Pathology, 32*(5), 2075-2086.

[35] García-Gutiérrez, F., Alegret, M., Marquié, M., Muñoz, N., Ortega, G., & Valero, S. (2024). Unveiling the sound of the cognitive status: Machine learning-based speech analysis in the Alzheimer’s disease spectrum. *Alzheimer’s Research & Therapy, 16*(1), 26.

[36] Kaser, A. N., Lacritz, L. H., Winiarski, H. R., Gabirondo, P., Schaffert, J., & Cullum, C. M. (2024). A novel speech analysis algorithm to detect cognitive impairment in a Spanish population. *Frontiers in Neurology, 15*, 1342907.

[37] A. M. Lanzi, A. K. Saylor, D. Fromm, H. Liu, B. MacWhinney, and M. Cohen, ”DementiaBank: Theoretical rationale, protocol, and illustrative analyses,” *American Journal of Speech-Language Pathology*, 2023.

[38] Y. Yamada, K. Shinkawa, M. Nemoto, M. Ota, K. Nemoto, and T. Arai, “Speech and language characteristics differentiate Alzheimer’s disease and dementia with Lewy bodies,” *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 14, no. 1, p. e12364, 2022.

[39] R. Ossewaarde, R. Jonkers, F. Jalvingh, and R. Bastiaanse, “Classification of spontaneous speech of individuals with dementia based on automatic prosody analysis using support vector machines (SVM),” in *The Thirty-Second International FLAIRS Conference*, 2019.

[40] J. J. G. Meilán, F. Martínez-Sánchez, I. Martínez-Nicolás, T. E. Llorente, and J. Carro, “Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia,” *Behavioural Neurology*, vol. 2020, no. 1, p. 4683573, 2020.

[41] P. Pastoriza-Dominguez, I. G. Torre, F. Dieguez-Vide, I. Gómez-Ruiz, S. Geladó, J. Bello-López, A. Ávila-Rivera, J. A. Matias-Guiu, V. Pytel, and A. Hernández-Fernández, “Speech pause distribution as an early marker for Alzheimer’s disease,” *Speech Communication*, vol. 136, pp. 107–117, 2022.

[42] F. Martínez-Sánchez, J. J. G. Meilán, J. A. Vera-Ferrandiz, J. Carro, I. M. Pujante-Valverde, O. Ivanova, and N. Carcavilla, “Speech rhythm alterations in Spanish-speaking individuals with Alzheimer’s disease,” *Aging, Neuropsychology, and Cognition*, vol. 24, no. 4, pp. 418–434, 2017.