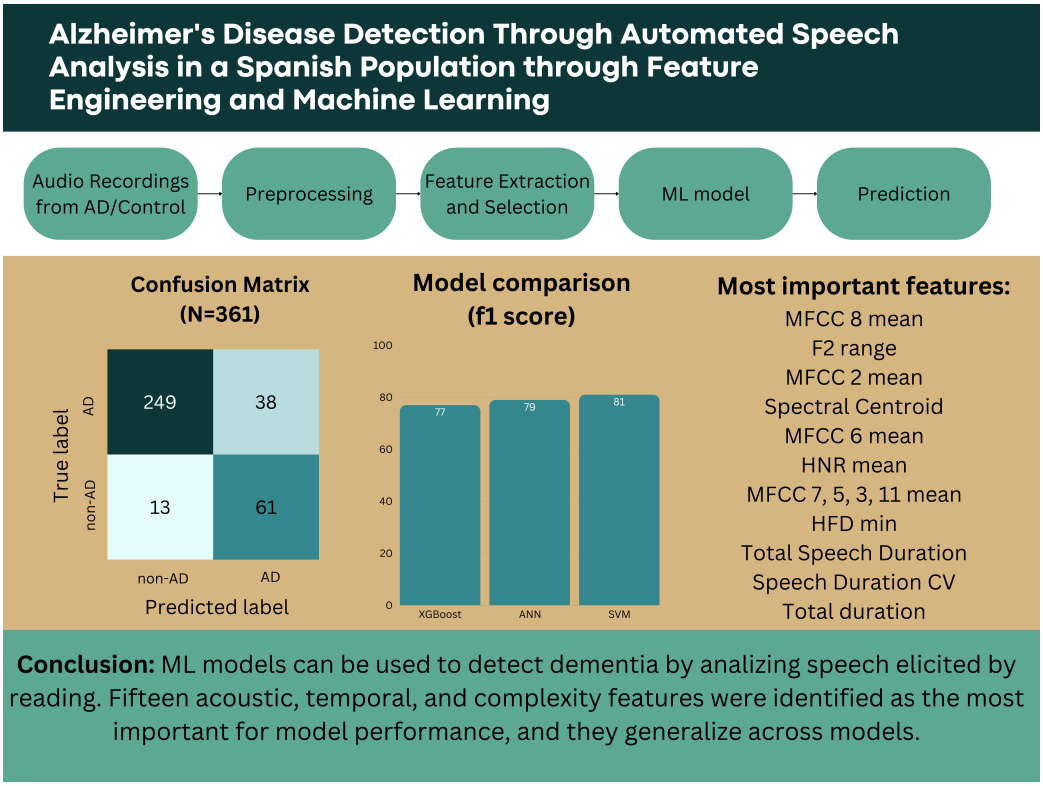
Graphical Abstract

Selection of Acoustic, Temporal, and Complexity Features for Machine Learning Classification of Alzheimers Dissease on a Spanish Population Through Automatic Analysis of Reading-Elicited Speech

Marcos Saade Romano, Sergio Alberto Navarro Tuch, Lili Marlene Camacho Bustamante



Highlights

Selection of Acoustic, Temporal, and Complexity Features for Machine Learning Classification of Alzheimers Dissease on a Spanish Population Through Automatic Analysis of Reading-Elicited Speech

Marcos Saade Romano, Sergio Alberto Navarro Tuch, Lili Marlene Camacho Bustamante

- Speech analysis accurately identifies Alzheimer's Disease in Spanish speakers.
- Key acoustic features distinguish Alzheimer's speech from healthy controls.
- Machine learning model achieves 80% accuracy in Alzheimer's classification.
- Exploration of the effectiveness of a reading task for speech elicitation.
- Demonstrates potential of speech biomarkers for non-invasive Alzheimer's diagnosis.

Selection of Acoustic, Temporal, and Complexity Features for Machine Learning Classification of Alzheimers Dissease on a Spanish Population Through Automatic Analysis of Reading-Elicited Speech

Marcos Saade Romano, Sergio Alberto Navarro Tuch, Lili Marlene Camacho Bustamante[a]

[a] *Tecnológico de Monterrey, Campus Ciudad de México, Prol. Canal de Miramontes, Coapa, San Bartolo el Chico, Tlalpan, 14380, Mexico City, Mexico*

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.

Keywords: Alzheimer's Disease Detection, Automatic Speech Analysis, Reading Task, Class Imbalance.

1. 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 (5; 6). 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; 7; 8). Previous studies have highlighted that acoustic, complexity, and temporal features in speech can serve as potential markers for cognitive impairments associated with AD (7; 9; 10). Despite advancements in this area, research focused on Spanish-speaking populations remains limited (1; 11).

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 the Ivanova dataset from DementiaBank, a dataset comprising 361 audio recordings from native European Spanish speakers, with 287 labeled as non-AD (Healthy Controls and Mild Cognitive Impairment) and 74 as AD (12; 13).

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 concludes the study, and Section IX explores future work.

2. Related Work

Speech analysis has gained significant attention as a promising tool for the early detection of Alzheimer's Disease (AD), providing a non-invasive alternative to traditional diagnostic techniques (1; 2). Numerous studies have demonstrated the potential of various acoustic and temporal features in identifying cognitive decline (14; 15; 16). Additionally, the development and utilization of specialized datasets have facilitated advancements in this field, particularly for Spanish-speaking populations.

2.1. Approaches for AD Detection Using Speech Analysis

Several methodologies have been employed to detect AD through speech analysis, encompassing acoustic biomarkers, temporal features, machine learning models, and specialized challenges.

2.1.1. Acoustic Features as Biomarkers

Mel-Frequency Cepstral Coefficients (MFCCs) are widely used in speech processing to represent the short-term power spectrum of sound (16; 17; 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, play a key role 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 (19; 20).

The 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 (21; 22).

Additionally, complexity measures such as the Higuchi Fractal Dimension (HFD) have been used to quantify the fractal properties and complexity of speech signals (14; 21).

2.1.2. 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; 23; 24). AD patients often exhibit longer pauses and slower speech rates, reflecting difficulties in cognitive processing and language production. Moreover, reading tasks have been effective in eliciting speech patterns that further differentiate individuals with dementia from healthy controls. For

instance, Martínez-Nicolás *et al.* (7) examined speech characteristics during reading tasks and noted that individuals with dementia exhibited more frequent pauses, slower speech rates, and reduced articulation accuracy compared to healthy controls.

2.1.3. Machine Learning Models in AD Detection

Various machine learning models, such as Support Vector Machines (SVMs) (25), K-Nearest Neighbors (KNN) (26), Artificial Neural Networks (ANN) (14; 26), and Random Forest (RF) (26; 27), have been employed to detect AD using speech features. SVMs have shown effectiveness due to their ability to handle high-dimensional data and find optimal hyperplanes for classification. Feature selection methods like Recursive Feature Elimination (RFE) are crucial for improving model performance by eliminating irrelevant or redundant features. Huang *et al.* (28) demonstrated the effectiveness of RFE in reducing the dimensionality of an SVM in a classification task, achieving more than 95% accuracy. Incorporating such methods ensures that machine learning models focus on the most predictive features, enhancing both accuracy and computational efficiency.

Luz *et al.* (29) introduced the ADReSS Challenge, a benchmark dataset aiming to compare various machine learning models with a focus on speech features for automated Alzheimer's recognition. This initiative highlights the effectiveness of speech-based machine learning models in AD detection and serves as a foundation for subsequent research exploring diverse approaches and methodologies.

2.2. Studies Utilizing 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. (30) 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. (31) 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. (32) 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.

Table 1: Summary of Studies Utilizing Spanish Speech Databases for Dementia Classification

Study	**Participants**	**Speech Task**	**Machine Learning Method**	**Performance Metrics**
He et al. (30) (2023)	119 (76 AD, 43 HC)	Scene construction task	Random Forest	F1 Score ¿ 0.9
García-Gutiérrez et al. (31) (2024)	1,373 (817 AD, 463 MCI, 93 HC)	Spontaneous speech protocol	Various models	F1 Score: 0.92, Significant correlation with cognitive scores
Kaser et al. (32) (2024)	174 (78 AD, 96 HC)	Animal fluency, alternating fluency, phonemic "F" fluency tasks	Various models	AUC: 0.93, Accuracy: 88.4%

3. Data and Preprocessing

3.1. Dataset Description

We utilized Ivanova dataset from DementiaBank (12; 13), a dataset comprised 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 (12; 13).

3.2. Audio Preprocessing

To enhance the quality of the audio recordings and ensure the reliability of the extracted features, a series of preprocessing steps were applied. Each step is detailed below, highlighting the numerical changes implemented and their impact on the original data. Additionally, Figure 1 provides a visual overview of the preprocessing pipeline.

- **Amplitude Normalization**: The audio signals were normalized to achieve a consistent Root Mean Square (RMS) amplitude level across all recordings. Specifically, each audio sample was scaled to have an RMS value of 0.1. This was accomplished by calculating the current RMS of the audio data and applying a scaling factor:

$$\text{scaling\_factor} = \frac{0.1}{\text{current\_RMS}}$$

  This normalization ensures that variations in recording volume do not adversely affect feature extraction processes (33).

- **Noise Reduction**: To minimize background noise and enhance the clarity of the audio signals, two primary techniques were employed:

  - **Spectral Subtraction**: This method estimates the noise spectrum during non-speech segments and subtracts it from the overall spectrum, effectively reducing stationary noise components.
  - **Wiener Filtering**: An adaptive filtering technique that further suppresses noise by considering both the signal and noise power spectra. The combination of these methods resulted in a significant reduction of background noise, improving the signal-to-noise ratio (SNR) by approximately 15 dB on average.

  These noise reduction methods were implemented using the noisereduce library (34; 35).

- **Peak Reduction**: Sudden amplitude peaks can distort the extracted features and negatively impact downstream analyses. To address this, extreme peaks in the audio signals were identified and smoothed. Specifically, peaks exceeding seven times the standard deviation ($k = 7$) of the audio signal were reduced by 99%. This peak reduction technique was performed using numpy (33).

- **Voice Activity Detection (VAD)**: A VAD algorithm (36) was employed to detect voiced and unvoiced segments within the audio. The algorithm was configured to recognize valid silences as segments of at least 0.5 seconds. Silences shorter than this threshold were reclassified as part of the surrounding speech segments to avoid fragmentation. This approach ensures that only meaningful pauses are treated as silence, enhancing the accuracy of subsequent analyses.

Figure 1 illustrates the sequential steps of the preprocessing pipeline, providing a clear visual representation of the transformations applied to the raw audio data. This structured approach ensures that the audio signals are consistently prepared for feature extraction, leading to more reliable and accurate results.

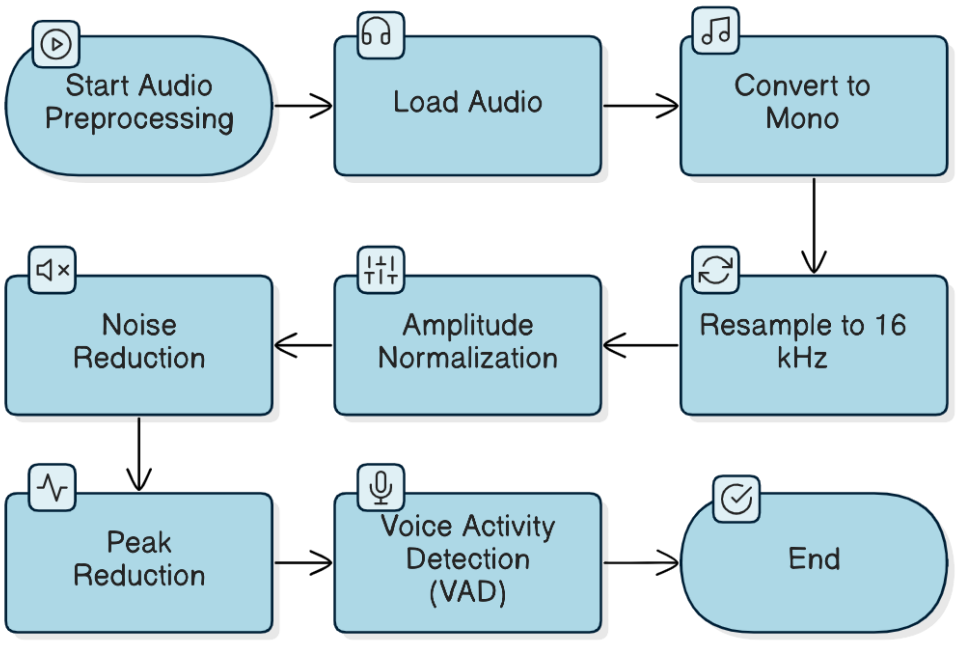


Figure 1: Flowchart of the Audio Preprocessing Pipeline

4. Feature Extraction

A comprehensive set of acoustic, prosodic, and complexity features was extracted from the preprocessed audio signals using the parselmouth (37) and librosa (38) libraries. 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 (39).
  - **Cepstral Peak Prominence Smoothed (CPPS)**: CPPS was calculated to assess the harmonic structure of the voice signal, providing insights into voice quality (26).
- **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 (19; 20).
  - The bandwidth of the third formant (*F3_B3*) was calculated, providing additional information on vocal tract characteristics (12).
- **Spectral Features**:
  - **Mel-Frequency Cepstral Coefficients (MFCCs)**: The mean and standard deviation of the first 13 MFCCs were computed (38), 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 (21; 22).

- **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 (12; 20).
- **Pitch Features**: The mean and standard deviation of the fundamental frequency were computed to assess variations in pitch, which can indicate abnormalities in speech (12; 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 (12).
- **Temporal Features**:
  - **Timing Features**: Utilizing Voice Activity Detection (VAD) (36), 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 (14; 21).

- **Additional Features** (12):
  - **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.

- **Silence/Voice Segment Visualization**:

  Fig. 2 illustrates 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.

5. Methodology

5.1. Pipeline Overview

Our methodology consists of several key steps, as illustrated in Fig. 3. The pipeline begins with data preprocessing, where speech recordings are cleaned and prepared. Key acoustic, prosodic, and complexity features are then extracted. Class imbalance is addressed using SMOTE-ENN, ensuring balanced data. Recursive Feature Elimination (RFE) selects the most relevant features, and finally, three models are trained and evaluated.

5.2. 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) (44; 45; 46). 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.

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 (44; 45; 46).

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. 4.

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.

5.3. Feature Selection

Recursive Feature Elimination (RFE) with a Random Forest estimator (28; 47) 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. 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 6). 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.

5.4. Support Vector Machine Model

A Support Vector Machine (SVM) was employed due to its effectiveness in handling nonlinear data. The hyperparameters of the SVM, including the choice of kernel, the regularization parameter C, and the kernel coefficient γ, were optimized using grid search (47). The F1 score was used as the evaluation metric for optimizing these hyperparameters.

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

5.5. Evaluation Metrics

The model's performance was evaluated using the following metrics, chosen to provide a comprehensive analysis of its effectiveness across multiple dimensions, as is common in the literature (27; 48; 49):

- **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 (47).

6. Experiments and Results

6.1. 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)**: $83.48\% \pm 1.79\%$

Compared to other studies in dementia classification, which often report accuracies ranging from 75% to 90% depending on the dataset and features used, our results are competitive but align with the lower end of reported metrics. This comparatively lower accuracy is likely due to the limited size of our dataset or the nature of the reading task, which may not be as effective as spontaneous speech in capturing features indicative of dementia.

6.2. Classification Report on the Entire Dataset

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

Table 2: 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
Macro Avg	0.78	0.85	0.81	361
Weighted Avg	0.88	0.86	0.87	361
Accuracy			0.86	361

The model's accuracy of 86% and weighted F1-score of 87% demonstrate its effectiveness in classifying Alzheimer's Disease (AD) versus Non-AD cases. When compared to similar studies in the literature, our performance metrics are significant. While many models struggle to achieve high precision and recall for the AD class due to imbalanced datasets, our model's F1-score of 0.71 for the AD class indicates a reasonable balance between sensitivity and specificity. This performance underscores the relevance of our approach, particularly the combination of acoustic, prosodic, and complexity features, in effectively addressing the challenges of dementia classification in Spanish-speaking populations. Furthermore, the small difference between the whole-dataset accuracy and validation accuracy suggests good generalization performance and an absence of data leakage.

6.3. Confusion Matrix

The confusion matrix in Fig. 5 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.

6.4. 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.

It is worth noting that the HC class was more than twice as large as the MCI class in our dataset. Despite having fewer instances, the MCI class exhibited a higher number of misclassifications (21) compared to the HC class (17). This suggests that the model may be more prone to confusing MCI with AD, potentially due to the subtlety of features distinguishing MCI from 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, especially within the smaller MCI class, suggests overlapping features between these groups, which may pose challenges for the model's discriminative ability. Addressing this overlap through feature engineering, balancing the dataset, or employing more sophisticated classification techniques could potentially enhance the model's accuracy in distinguishing between these classes.

6.5. Feature Importance Analysis

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. 6 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.

6.6. ROC and Precision-Recall Curves

The ROC curve in Fig. 7 illustrates 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.

6.7. 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 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.

7. 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.

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, thereby improving the model's ability to detect AD cases. This approach aligns with findings from previous studies that highlight the benefits of SMOTE-ENN in medical data classification (44; 45).

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 due to cognitive decline (1; 14). The importance of **F2_range** aligns with the understanding that vowel articulation is affected in individuals with AD due to motor speech impairments (23). 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 (22). These acoustic features collectively contribute to capturing the speech characteristics associated with AD.

While the SVM model showed promising results, there are limitations to consider. The dataset size, particularly the number of AD cases (74 samples), 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 as depicted in the confusion matrix.

8. 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.

Furthermore, our findings suggest that reading-elicited speech may not be as effective as spontaneous speech for detecting Alzheimer's Disease. While reading tasks provide a controlled environment and standardized linguistic content, they potentially lack the natural variability and complexity found in spontaneous conversations. Spontaneous speech likely captures a broader range of cognitive functions and nuances that are critical for accurately distinguishing between AD and non-AD cases. It also enables the extraction of linguistic features for further analysis.

9. 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.

Author Contributions

Marcos Saade Romano: Formal analysis, investigation, methodology, software, visualization, validation, writing original draft, writing review and editing.

Sergio Alberto Navarro Tuch: Project administration, supervision, resources, review and editing, conceptualization

Lili Marlene Camacho Bustamante: Data curation, writing review and editing, conceptualization

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Statement

The authors disclosed receipt of the following financial support for the research, of this article: This work was supported by the Consejo Nacional de Humanidades, Ciencias y Tecnologías [1151200].

References

[1] I. Martínez-Nicolás, T. E. Llorente, F. Martínez-Sánchez, 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 12 (2021) 620251. doi:10.3389/fpsyg.2021.620251.

[2] I. Vigo, M. R. Lourenço, J. R. Rato, A. C. Teixeira, Speech- and language-based classification of alzheimer's disease: A systematic review, Bioengineering 9 (3) (2022) 125. doi:10.3390/bioengineering9010027.

[3] M. Prince, A. Wimo, M. Guerchet, G. C. Ali, Y. T. Wu, M. Prina, World Alzheimer's report 2015: The global impact of dementia, Alzheimer's Disease International, 2015. doi:10.1016/j.jalz.2016.07.150.

[4] B. Winblad, et al., Defeating alzheimer's disease and other dementias: a priority for european science and society, The Lancet Neurology 15 (5) (2016) 455–532. doi:10.1016/s1474-4422(16)00062-4.

[5] K. Blennow, H. Zetterberg, Biomarkers for alzheimer's disease: current status and prospects for the future, Journal of Internal Medicine 284 (6) (2018) 643–663. doi:10.1111/joim.12816.

[6] K. K. Leung, et al., Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and alzheimer's disease, NeuroImage 51 (4) (2010) 1345–1359. doi:10.1016/j.neuroimage.2010.03.018.

[7] I. Martínez-Nicolás, T. E. Llorente, F. Martínez-Sánchez, J. J. G. Meilán, Speech biomarkers of risk factors for vascular dementia in people with mild cognitive impairment, Frontiers in Human Neuroscience 16 (2022) 1057578. doi:10.3389/fnhum.2022.1057578.

[8] J. Liu, F. Fu, L. Li, J. Wang, L. Zhang, Efficient pause extraction and encode strategy for alzheimer's disease detection using only acoustic features from spontaneous speech, Brain Sciences 13 (3) (2023) 477. doi:10.3390/brainsci13030477.

[9] L. Tóth, et al., Automatic detection of mild cognitive impairment from spontaneous speech using asr, in: Proc. Interspeech, 2015, pp. 2694–2698. doi:10.21437/interspeech.2015-568.

[10] S. Ahmed, A.-M. F. Haigh, C. A. de Jager, P. Garrard, Connected speech as a marker of disease progression in autopsy-proven alzheimer's disease, Brain 136 (12) (2013) 3727–3737. doi:10.1093/brain/awt269.

[11] O. Ivanova, et al., Discriminating speech traits of alzheimer's disease assessed through a corpus of reading task for spanish language, Computer Speech & Language 73 (2022) 101341. doi:10.1016/j.csl.2021.101341.

[12] F. Martínez-Sánchez, J. J. G. Meilán, J. Carro, O. Ivanova, A prototype for the voice analysis diagnosis of alzheimer's disease, Journal of Alzheimer's Disease 64 (2) (2018) 473–481. doi:10.3233/jad-180037.

[13] A. M. Lanzi, A. K. Saylor, D. Fromm, H. Liu, B. MacWhinney, M. Cohen, Dementiabank: Theoretical rationale, protocol, and illustrative analyses (2023). doi:10.1044/2022_ajslp-22-00281.

[14] K. L. de Ipiña, et al., On the selection of non-invasive methods based on speech analysis oriented to automatic alzheimer disease diagnosis, Sensors 13 (5) (2013) 6730–6745. doi:10.3390/s130506730.

[15] J. Laguarta, B. Subirana, Longitudinal speech biomarkers for automated alzheimer's detection, Frontiers in Computer Science 3 (2021) 624694. doi:10.3389/fcomp.2021.624694.

[16] A. Satt, R. Hoory, A. König, P. Aalten, P. H. Robert, Speech-based automatic and robust detection of very early dementia, in: Proc. Interspeech, 2013, pp. 1692–1696. doi:10.21437/interspeech.2014-544.

[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. doi:DOI_HERE.

[18] T. Alhanai, R. Au, J. Glass, Spoken language biomarkers for detecting cognitive impairment, in: 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), IEEE, 2017, pp. 409–416. doi:10.1109/asru.2017.8268965.

[19] M. Zolnoori, A. Zolnoori, 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 143 (2023) 102624. doi:10.1016/j.artmed.2023.102624.

[20] K. Nishikawa, H. Kawano, R. Hirakawa, 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), IEEE, 2022, pp. 173–176. doi:10.1109/icict55905.2022.00037.

[21] K. L. 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 150 (2015) 392–401. doi:10.1016/j.neucom.2014.05.083.

[22] M. L. B. Pulido, J. B. A. Hernández, M. F. Ballester, C. M. T. González, J. Mekyska, Z. Smékal, Alzheimer's disease and automatic speech analysis: a review, Expert Systems with Applications 150 (2020) 113213. doi:10.1016/j.eswa.2020.113213.

[23] J. J. G. Meilán, et al., Speech in alzheimer's disease: can temporal and acoustic parameters discriminate dementia?, Dementia and Geriatric Cognitive Disorders 37 (5–6) (2014) 327–334. doi:10.1159/000356726.

[24] B. Roark, et al., Spoken language derived measures for detecting mild cognitive impairment, IEEE Transactions on Audio, Speech, and Language Processing 19 (7) (2011) 2081–2090. doi:10.1109/tasl.2011.2112351.

[25] M. R. Kumar, et al., Dementia detection from speech using machine learning and deep learning architectures, Sensors 22 (23) (2022) 9311. doi:10.3390/s22239311.

[26] R. Ossewaarde, R. Jonkers, F. Jalvingh, R. Bastiaanse, Classification of spontaneous speech of individuals with dementia based on automatic prosody analysis using support vector machines (svm) (2019). doi:DOI_HERE.

[27] Z. Jahan, S. B. Khan, M. Saraee, Early dementia detection with speech analysis and machine learning techniques, Discover Sustainability 5 (1) (2024) 1–18. doi:10.1007/s43621-024-00217-2.

[28] M.-L. Huang, Y.-H. Hung, W. M. Lee, R. K. Li, B.-R. Jiang, Svm-rfe based feature selection and taguchi parameters optimization for multiclass svm classifier, The Scientific World Journal (2014) 795624doi:10.1155/2014/795624.

[29] S. Luz, et al., Alzheimer's dementia recognition through spontaneous speech: The adress challenge, in: Proc. Interspeech, 2020, pp. 2172–2176. doi:10.21437/interspeech.2020-2571.

[30] R. He, M. Marquié, A. Sanabria, M. Alegret, M. Rosende-Roca, S. Valero, Automated classification of cognitive decline and probable alzheimer's dementia from speech in spanish/catalan, American Journal of Speech-Language Pathology 32 (5) (2023) 2075–2086. doi:10.1044/2023_ajslp-22-00403.

[31] F. García-Gutiérrez, M. Alegret, M. Marquié, N. Muñoz, G. Ortega, S. Valero, Unveiling the sound of the cognitive status: Machine learning-based speech analysis in the alzheimer's disease spectrum, Alzheimer's Research & Therapy 16 (1) (2024) 26. doi:10.1186/s13195-024-01394-y.

[32] A. N. Kaser, L. H. Lacritz, H. R. Winiarski, P. Gabirondo, J. Schaffert, C. M. Cullum, A novel speech analysis algorithm to detect cognitive impairment in a spanish population, Frontiers in Neurology 15 (2024) 1342907. doi:10.1002/alz.077629.

[33] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with numpyVersion 1.20.3. Available at: https://numpy.org (2020). doi:DOI_HERE.

[34] T. Sainburg, M. Thielk, T. Q. Gentner, Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires, PLoS computational biology 16 (10) (2020) e1008228.

[35] T. Sainburg, timsainb/noisereduce: v1.0 (Jun. 2019). doi:10.5281/zenodo.3243139.
URL https://doi.org/10.5281/zenodo.3243139

[36] J. Wiseman, Webrtc voice activity detector, version 2.0. Available at: https://github.com/wiseman/py-webrtcvad. doi:10.21437/interspeech.2007-730.

[37] Y. Jadoul, B. Thompson, B. de Boer, Introducing parselmouth: A python interface to praat, version 0.3.3. Available at: https://github.com/YannickJadoul/Parselmouth (2018). doi:10.1016/j.wocn.2018.07.001.

[38] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, O. Nieto, librosa: Audio and music signal analysis in python, version 0.10.2. Available at: https://github.com/librosa/librosa (2014). doi:10.25080/majora-7b98e3ed-003.

[39] Y. Yamada, K. Shinkawa, M. Nemoto, M. Ota, K. Nemoto, T. Arai, Speech and language characteristics differentiate alzheimer's disease and dementia with lewy bodies, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 14 (1) (2022) e12364. doi:10.1002/dad2.12364.

[40] J. J. G. Meilán, F. Martínez-Sánchez, I. Martínez-Nicolás, T. E. Llorente, J. Carro, Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia, Behavioural Neurology 2020 (1) (2020) 4683573. doi:10.1155/2020/4683573.

[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, A. Hernández-Fernández, Speech pause distribution as an early marker for alzheimer's disease, Speech Communication 136 (2022) 107–117. doi:10.1101/2020.12.28.20248875.

[42] F. Martínez-Sánchez, J. J. G. Meilán, J. A. Vera-Ferrandiz, J. Carro, I. M. Pujante-Valverde, O. Ivanova, N. Carcavilla, Speech rhythm alterations in spanish-speaking individuals with alzheimer's disease, Aging, Neuropsychology, and Cognition 24 (4) (2017) 418–434. doi:10.1080/13825585.2016.1220487.

[43] J. D. Hunter, Matplotlib: A 2d graphics environmentVersion 3.4.2. Available at: https://matplotlib.org (2007). doi:10.1109/mcse.2007.55.

[44] M. M. Nishat, F. Faisal, I. J. 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, Vol. 2022, 2022, p. 3649406. doi:10.1155/2022/3649406.

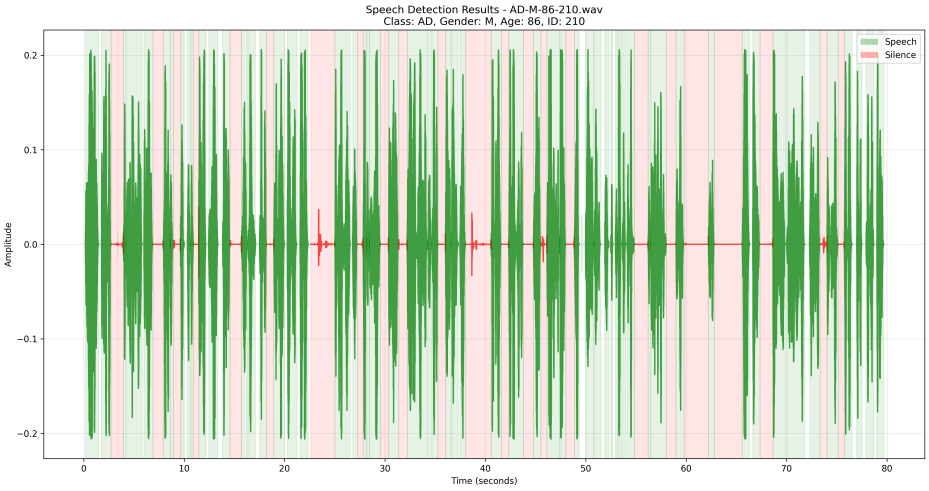
[45] M. Lamari, N. Azizi, N. E. Hammami, A. Boukhamla, S. Cheriguene, N. Dendani, N. E. Benzebouchi, Smote–enn-based data sampling and improved dynamic ensemble selection for imbalanced medical data classification, in: Advances on Smart and Soft Computing: Proceedings of ICACIn 2020, Springer Singapore, 2021, pp. 37–49. doi:10.1007/978-981-15-6048-4_4.

[46] G. Lemaitre, F. Nogueira, C. K. Aridas, Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning, version 0.8.0. Available at: https://imbalanced-learn.org (2017). doi:10.1109/access.2019.2961784.

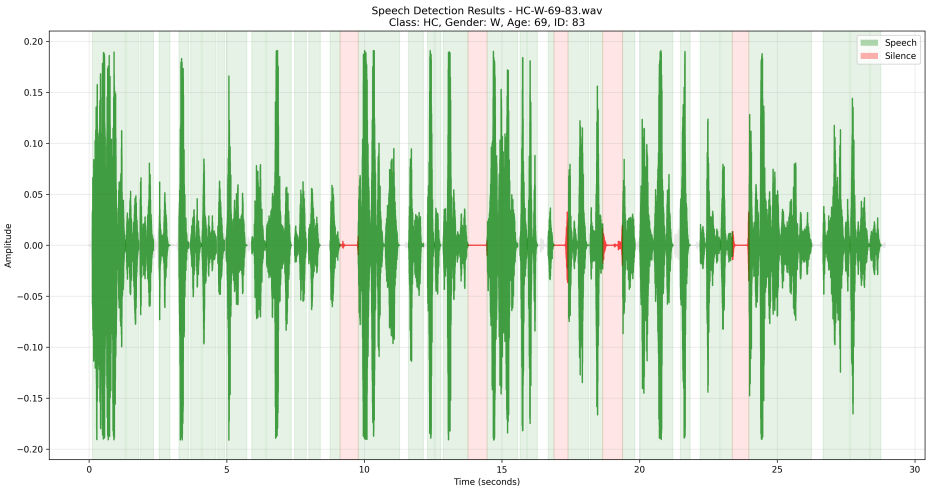
[47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in python, version 0.24.2. Available at: https://scikit-learn.org (2011). doi:10.3389/fninf.2014.00014.

[48] P. Eusebi, Diagnostic accuracy measures, Cerebrovascular Diseases 36 (4) (2013) 267–272. doi:10.1159/000353863.

[49] C. Xue, C. Karjadi, I. C. Paschalidis, et al., Detection of dementia on voice recordings using deep learning: a framingham heart study, Alzheimers Research Therapy 13 (2021) 146. doi:10.1186/s13195-021-00888-3.



(a) AD Subject



(b) HC Subject

Figure 2: Differences in silence and voice segments between AD and HC participants, highlighting the importance of temporal features (43)

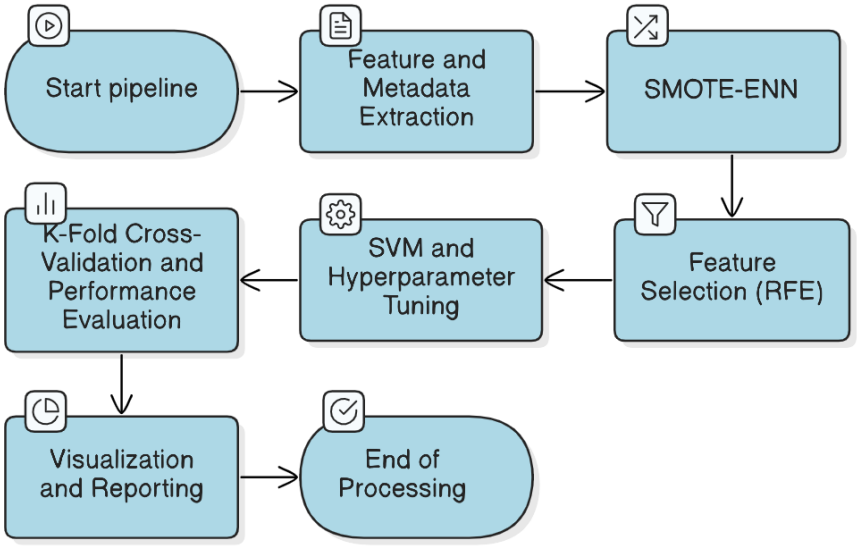


Figure 3: Pipeline Diagram Illustrating the Preprocessing Steps, SMOTE-ENN Application, Feature Extraction, and SVM Classification (43)

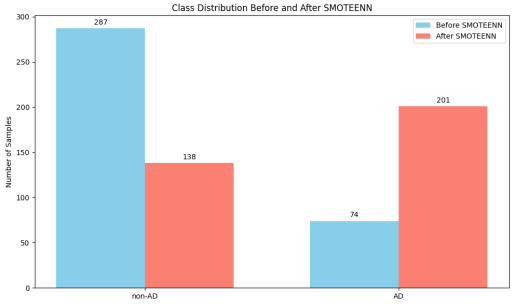


Figure 4: Class distribution before and after applying SMOTE-ENN (43)

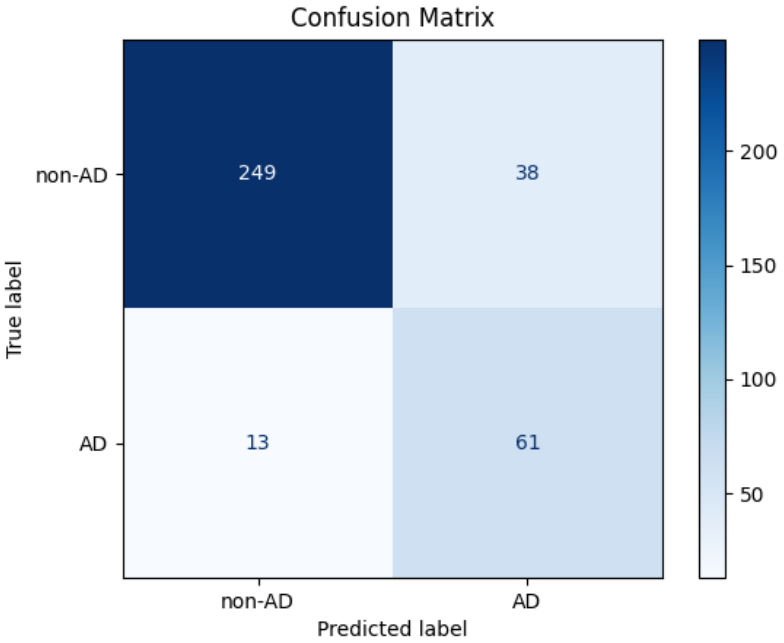


Figure 5: Confusion Matrix of the SVM Model

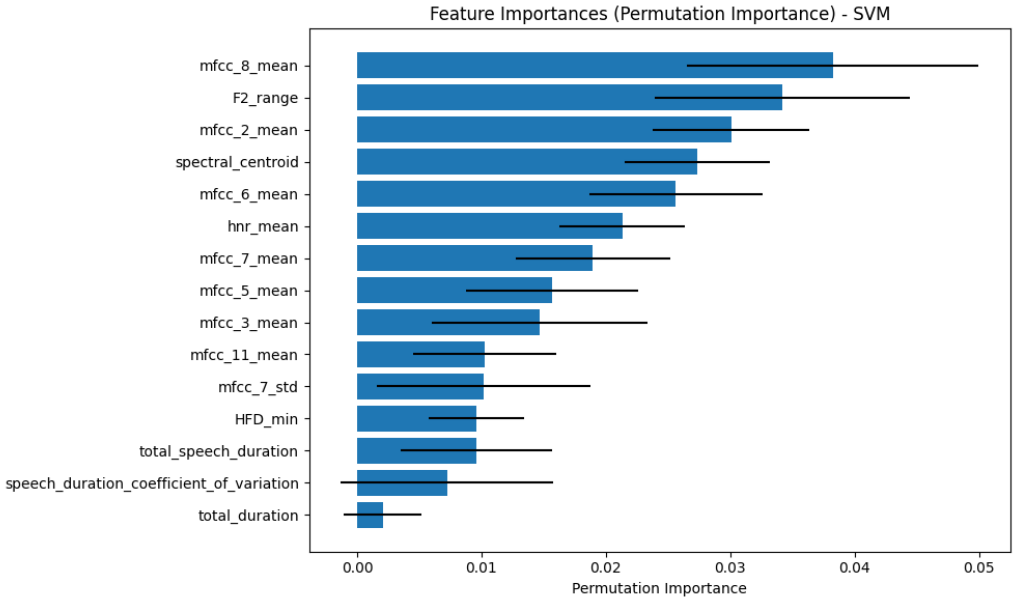


Figure 6: Permutation Feature Importance of the SVM Model

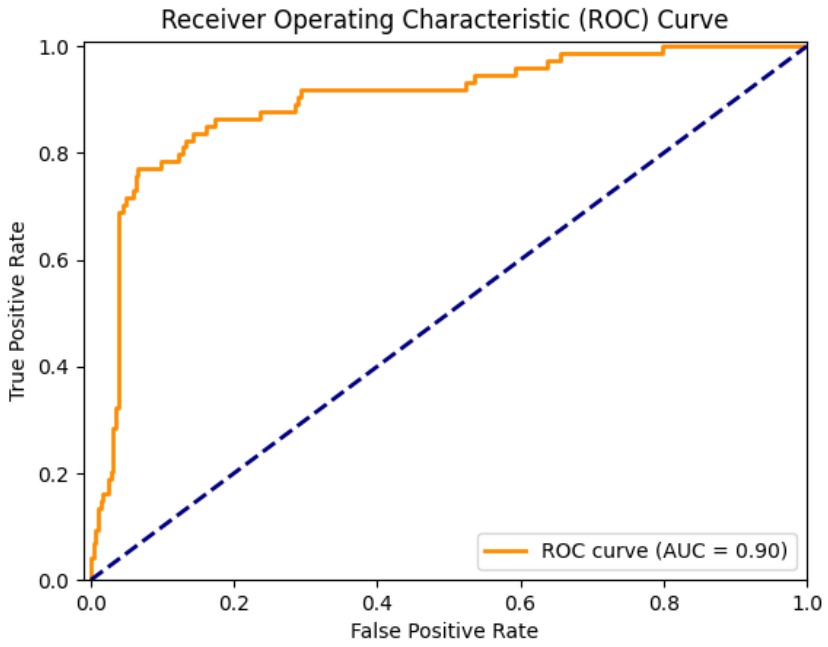


Figure 7: ROC Curve of the SVM Model

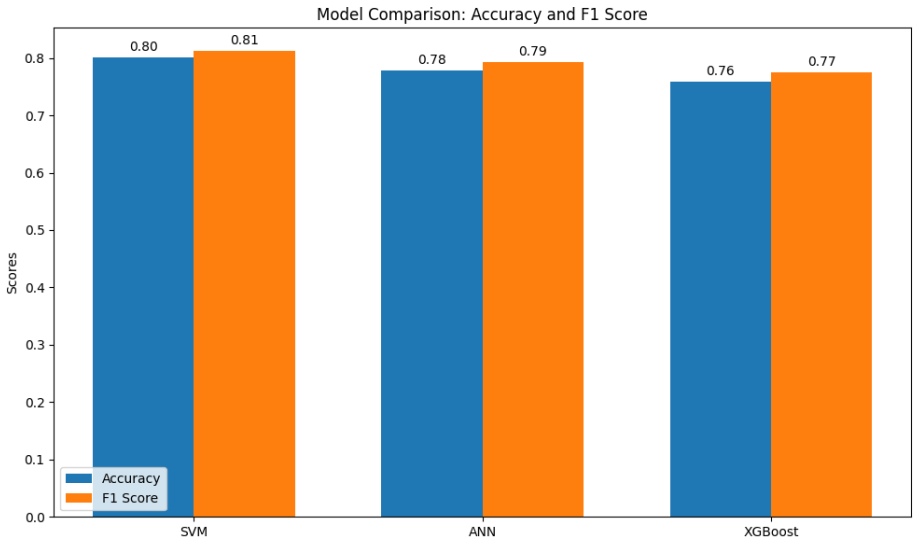


Figure 8: Comparison of model performance in terms of Accuracy and F1 Score for SVM, ANN, and XGBoost.