Multi-Task Learning with Acoustic Features for Alzheimer's Disease Detection

Marek Sviderski, Basel Barakat, Becky Allen
Faculty of Technology
University of Sunderland
Sunderalnd, UK
bh37sy@student.sunderland.ac.uk,{basel.barakat,becky.allen}@sunderland.ac.uk

Kate MacFarlane
Department of Computing and Mathematics
The Manchester Metropolitan University
Manchester, UK
k.macfarlane@mmu.ac.uk

***Abstract*—This study explores the potential of acoustic features extracted from speech recordings for detecting Alzheimer's Dementia (AD), employing a comprehensive approach that incorporates binary classification (healthy control vs. dementia), multiclass classification (healthy control, mild cognitive impairment, AD), and regression analyses (predicting MMSE scores). Additionally, demographic information of the participants was integrated to enhance the models' predictive accuracy. Our methodology involved processing each dataset version through a series of machine learning models tailored to each task, starting with a baseline version, followed by hyperparameter optimisation, and finally applying a combination of preprocessing steps (scaling, outlier removal, dimensionality reduction, and skewness correction) to identify the optimal setup for each model.**

The findings indicate that preprocessing steps significantly improve model performance across all tasks, underscoring the importance of data preparation in machine learning workflows for healthcare applications. Notably, the use of acoustic data alone for AD detection shows promising results, suggesting a pathway toward more generalised approaches that could incorporate recordings in various languages without linguistic dependency. This opens up the possibility for scalable, non-invasive screening tools for AD, leveraging the universal nature of acoustic markers in speech for early detection and monitoring of this condition.

***Index Terms*—Alzheimer's dementia, acoustic analysis, machine learning, feature engineering.**

I. Introduction

Alzheimer's Dementia (AD) is a progressive disease impairing memory and behaviour, predominantly diagnosed in individuals over 65 with varying life expectancy. Identifying AD at an early stage is crucial, as it allows for timely intervention with treatments such as Donepezil, which, while not a cure, can alleviate symptoms and potentially improve the quality and duration of life for those affected [1]. Current AD diagnosis primarily relies on neuroimaging (MRI, fMRI, CT, PET), genomics, biomarkers, and speech patterns, with limitations due to sparse data and underdeveloped novel categories [2]. Speech recognition has a critical advantage over neuroimaging i.e., it can detect AD in the initial disease stages, which may not reveal visible brain changes [3]. Despite the established protocols, the quest for non-invasive, cost-effective, and easily accessible diagnostic tools remains critical.

Neuroimaging faces drawbacks like high costs and patient discomfort, whereas genetic tests predict AD risk without confirming disease presence [4]. Against this disadvantages, language and acoustic features emerge as promising avenues. These methods, particularly the analysis of spontaneous speech, offer a less invasive and more accessible way to detect early signs of AD, leveraging the frequent and natural occurrence of speech in daily life. This paper focuses on the potential of acoustic features, dissecting the nuances of speech that could indicate the onset of AD. By focusing exclusively on acoustic parameters, this research sidesteps the complexities and language dependencies inherent in lexical analysis, paving the way for a universally applicable diagnostic tool.

The initial steps in identifying potential neurological conditions often begin with observing the intricacies of speech, usually noted by individuals or their families. Early detection through these observations is key for timely intervention. This discussion highlights how traditional diagnostic approaches and symptoms are increasingly complemented by machine learning techniques in speech analysis. Machine learning in this context is diversified into three main strategies: analysing semantic errors to uncover lexical irregularities, investigating speech rhythm and pitch for acoustic insights, and employing a combined approach to enhance detection accuracy [5]–[7].

In this study, we emphasise the unique advantages of acoustic analysis, particularly its independence from linguistic content, which potentially makes it applicable across languages and cultures. We tested the efficacy of various acoustic features in accurately diagnosing Alzheimer's Dementia, focusing on binary classification, multilabel classification, and regression task analysis. Adopting this approach facilitates the creation of a diagnostic tool that is versatile across different linguistic and cultural contexts, offering a practical solution for early AD detection that can be implemented worldwide.

The cookie theft picture description task from the Boston Diagnostic Aphasia Examination (BDAE) is a classic method for AD detection. It assesses how patients describe a picture, analysing their speech for AD indicators [8]. While this and similar tasks have provided valuable insights, they underscore the importance of speech as a diagnostic tool, yet often rely on content analysis. This paper's focus on acoustic features aims to complement these traditional methods, offering a new perspective on early AD detection that is less dependent on subjective interpretation and more grounded in objective, measurable acoustic properties.

This paper is structured as follows: Section II reviews existing research on AD detection, highlighting the gap this study aims to fill by focusing on acoustic features. Section III details our approach, from dataset description and feature extraction to classification strategies and optimisation techniques, emphasising the novel methodology of hyperparameter optimisation and preprocessor selection. Results from binary, multiclass, and regression analyses are presented in Section IV, followed by a discussion comparing these findings with current literature and their implications. The paper concludes in Section VI with final remarks on the potential of acoustic features for AD detection, the contributions of this study to the field, and future research directions aiming for broader language inclusively and further validation of this approach.

II. Related Work

Recent trends in AD detection through speech and language analysis are increasingly leaning towards combining acoustic and linguistic features to boost the effectiveness of diagnostic models. For example, Valsaraj et al. [9] adopted an eclectic mix of acoustic features alongside pre-trained BERT and TF-IDF for their study. They didn't just stop at gathering a wide range of specific features and statistics like Brunet's Index and Honore's Statistic; they also explored the eGeMAPS acoustic feature set and enriched their model with linguistic insights derived from speech transcriptions. Their preprocessing approach was focused, utilising Recursive Feature Elimination (RFE) specifically on the eGeMAPS features. The outcomes were notable—using just the eGeMAPS features, both the Logistic Regression and SVM models hit a 64% accuracy mark. Adding linguistic features bumped the Logistic Regression's accuracy to 66%, and combining the SVM with BERT and TF-IDF features pushed accuracy up to 70%.

Luz's work [10] opens another interesting chapter in the exploration of automatic acoustic feature sets, with a detailed look at emobase, ComParE, eGeMAPS, and MRCG functionals. Luz's methodology was particularly thorough, employing Pearson's correlation test to weed out features correlated beyond an absolute value of 0.2. The study wasn't just about finding the best feature set; it aimed to discern the most effective ones for distinguishing between Healthy Control (HC) and AD subjects, as well as for predicting Mini-Mental State Examination (MMSE) scores in a regression task. For binary classification, the ComParE set stood out, lifting the accuracy for various models—57% for LDA, 53% for DT, and 57% for 1NN. Luz also argued on the importance of addressing data imbalance and including demographic features, which, despite being readily available in many datasets like the Pitt corpus, are often overlooked.

Martinc and Pollak's study [11] includes advanced machine learning techniques, such as XGBoost, integrating acoustic features (eGeMAPS, MFCC) with linguistic elements like embeddings and readability features. Their innovative approach of early feature-level fusion and a strategic grid search for the best feature combinations yielded a 77% accuracy in binary classification and a RMSE score of 4.43%.

In conclusion, these studies collectively underscore that focusing solely on acoustic features could indeed be fruitful for AD detection. Automatic feature sets, despite some criticisms regarding their original non-AD-specific intentions as highlighted by Luz [10], offer a more automated approach to diagnosis. By incorporating specific features known to be effective in AD diagnosis.

III. Methodology

A. Dataset - cookie theft picture descriptions subset

The dataset used is the subset of the Pitt Dataset from the DementiaBank project [12]. The cookie picture cohort consist of 549 recordings. The full dataset consist of additional tests such as the recall, sentence, and fluency task. Although the distribution of the labels for other tasks is heavily favourite for the dementia label, the oppose can be found in the cookie cohort where the binary classification labels are distributed more evenly Healthy Control (HC) 243 vs. Alzheimer's Dementia (AD) 306.

The dataset from the DementiaBank project includes a subset of 549 recordings from the cookie theft picture descriptions task, featuring a relatively balanced label distribution (HC 243 vs. AD 306). However, the full dataset exhibits a significant imbalance, particularly for binary classification tasks, with 276 HC recordings versus 1040 AD. This skew could limit model generalizability to varied datasets. We addressed this challenge using stratified k-fold validation to ensure representative class distribution in each fold, enhancing our model's robustness and reducing the risk of overfitting. Future work should consider adaptive resampling and cost-sensitive learning to further mitigate these issues.

B. Feature Extraction and Pre-processing

Drawing upon the comprehensive literature review by Thaler and Gewald [6], our study taps into a nuanced understanding of speech characteristics indicative of Alzheimer's Dementia (AD) at its early stages. Our approach to feature extraction is detailed below, where we segment the process into various phases to highlight the flow from raw data to ready-to-use features for machine learning modelling.

As illustrated in Figure 1, the feature extraction process begins with the division into Temporal Dynamic and Spectral Features, complemented by Demographic Features. Temporal Dynamic Features include measurements such as speech and silence durations, phonation rates, and pause characteristics. Spectral Features are primarily derived from higher-order Mel Frequency Cepstral Coefficients (MFCCs), carefully selected to minimise biases introduced by external recording conditions.

Our study expands on foundational research by integrating a broad range of acoustic features divided into Temporal Dynamic, Spectral, and Demographic categories. Temporal Dynamic Features examine aspects like speech rhythm and pitch, including measurements of speech and silence durations, phonation rates, and pause characteristics. We derived Spectral

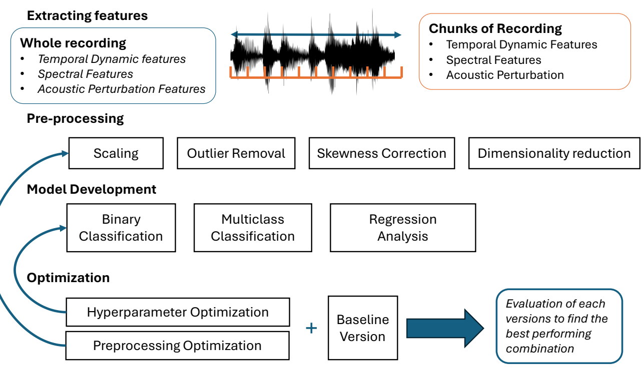

Fig. 1. Flowchart illustrating the feature extraction and preprocessing workflow used in the study, subdivided into Temporal Dynamic, Spectral, and Demographic Features alongside the subsequent preprocessing steps including scaling and outlier removal.

Features from higher-order Mel Frequency Cepstral Coefficients (MFCCs), excluding the first 13 coefficients to avoid biases introduced by recording conditions such as microphone distance and environment. This exclusion focuses our analysis on features that are more robust indicators of neurological changes in Alzheimer's Dementia, such as the mean, kurtosis, and skewness of the remaining coefficients.

Additionally, Demographic Features like age, gender, and education level complement these acoustic indicators, providing context to the speech data and aiding our regression analyses with target labels like the Mini Mental State Examination (MMSE) scores.

This multidimensional approach, integrating nuanced acoustic, spectral, and demographic insights, is designed to capture the complex alterations in speech associated with early-stage AD. By focusing on acoustic and demographic features, our analysis builds on Thaler and Gewald's groundwork, aiming to advance the predictive accuracy of ML models for early AD detection and contribute a novel perspective to the dialogue on speech-based diagnostics in neurodegenerative diseases.

Initially we extracted the Spectral Features to capture the power spectrum characteristics of speech via Mel-Frequency Cepstral Coefficients (MFCCs), along with their first and second-order derivatives. These features, coupled with statistical moments—mean, kurtosis, and skewness—offer a detailed portrayal of the spectral properties. For Temporal Dynamic Features, we quantified the speech time, pause intervals, and phonation rates to evaluate the speech flow and fluency. Acoustic perturbation measures, including jitter and shimmer, were calculated to assess the stability and quality of vocal production.

C. Classification and Analysis Tasks

In our study, we employed a rigorous methodology for splitting our data into training and validation sets to ensure the reliability and validity of our classification tasks. Specifically, for both binary and multiclass classification, we utilized the stratified k-fold validation technique, creating 5 folds for each iteration. This approach was chosen to address the challenge of imbalanced datasets, ensuring that each fold was representative of the overall distribution of classes, thus maintaining the integrity of our analysis and enhancing the robustness of our findings.

1) Binary and Multiclass Classification: For the binary classification task, the objective was to differentiate between Healthy Controls (HC) and individuals diagnosed with Alzheimer's Dementia such as in [13], [14]. This binary classification allows for a straightforward assessment of the model's ability to detect the presence of dementia based on acoustic markers using standalone machine learning pipelines. The models employed for this task included Decision Tree Classifier, K-Nearest Neighbors, Logistic Regression, Support Vector Classifier, and XGBoost Classifier. The performance of these models was evaluated using accuracy and F1 scores to assess both the precision and recall of the classification.

The multiclass classification task extended this binary framework to include an additional category for Mild Cognitive Impairment (MCI), resulting in three classes: HC, MCI, and AD [15], [16]. This classification is crucial for identifying early stages of cognitive decline, which can inform interventions and treatments. The models utilized for this task were Decision Tree, K-Nearest Neighbors, Random Forest, and XGBoost. The multiclass nature of this task poses a greater challenge, requiring the models to discern between not only healthy and diseased states but also between different levels of disease severity.

2) Regression Task: The regression task aimed at predicting the Mini-Mental State Examination (MMSE) score, a quantitative measure of cognitive function, from the acoustic features. The target variable, MMSE score, allows for a continuous assessment of cognitive status, providing a nuanced understanding of the cognitive abilities of the subjects. The models selected for this task included K-Nearest Neighbors Regressor, Lasso Regressor, SVR, and XGBoost Regressor. The performance of these models was evaluated using the R-squared score and Mean Squared Error (MSE) to determine how well the models predict MMSE scores based on speech characteristics.

For each of these tasks, data preprocessing involved adjusting the labels to fit the specific requirements of binary, multiclass, and regression frameworks. The default labelling scheme was designed for binary classification, with subsequent adjustments made to accommodate multiclass labels and continuous outcomes for the regression analysis.

D. Pre-processing

The initial dataset, encompassing both tabular data and complex structures such as arrays and nested arrays. Specifically, the features related to Mel-Frequency Cepstral Coefficients (MFCCs) were encapsulated within nested arrays, identified as 'spectral_first_coefficient' and 'spectral_second_coefficient'. Furthermore, a selection of features including 'spectral_mean', 'spectral_kurtosis', 'spectral_skew', 'acoustic_fundamental_frequency', 'acoustic_max_f0', and 'acoustic_fluctuations_f0' were formatted as simple arrays.

For the preprocessing of simple arrays, we employed a method to extract statistical measures—median, minimum, maximum, standard deviation, kurtosis, and skewness—thereby flattening the array structures. This procedure which operated on each array to generated a comprehensive feature set. Conversely, nested arrays, which contained 13 dimensions of MFCC data, underwent a similar process of feature extraction, albeit through iterative flattening to account for their multi-dimensional nature.

The following step consisted of adding demographic information of sex, age, and educational features. As per handling missing and uniform values, we excluded 'acoustic_max_f0_kurtosis' and 'acoustic_max_f0_skew', because of the amount of missing values and 'acoustic_min_f0', 'acoustic_fundamental_frequency_min', and several others related to 'acoustic_max_f0', which demonstrated no variability and thus, minimal contribution to the predictive modelling process. In the end of this pre-processing phase, our dataset, initially featuring 36 attributes, was expanded into final length of 222 features.

E. Optimisation Strategies

To enhance the predictive performance of our models for Alzheimer's Dementia detection through acoustic analysis, we employed a systematic approach to optimisation, delineated across three distinct versions: Baseline Version, Hyperparameter Optimised Version, and Pre-processing Optimised Versions. Each version adopted specific strategies aimed at refining the model's accuracy and generalizability.

- **Baseline Version (BV)**: Initially, all models were deployed with their default parameters to establish a performance benchmark. This version served as the foundation, allowing for a clear comparison of the effects of subsequent optimisation strategies. The classification tasks were evaluated using Stratified K-fold cross-validation with five folds to maintain class balance, while regression tasks utilized K-fold validation, also set to five folds, to assess model performance.
- **Hyperparameter Optimised Version (HO)**: Building upon the baseline, this version involved fine-tuning model parameters via GridSearchCV. A curated list of parameters for each model was defined, and GridSearchCV was employed to systematically explore the parameter space, identifying configurations that maximised either accuracy (for classification tasks) or R-squared score (for regression tasks) across the same five-fold cross-validation setup. The best-performing parameters were then applied to re-evaluate the models, ensuring consistency with the baseline validation method.
- **Pre-processing Optimised Versions (PO)**: The final layer of optimisation focused on preprocessing. We explored a combination of techniques to address outlier detection (winsorize, none), scaling (standard, min-max, robust, none), skewness correction (log, square-root, box-cox, yeo-johnson), and dimensionality reduction (PCA with 0.95 variance threshold, SelectKBest with k=10, RFE with a DecisionTreeClassifier and one feature selection, and none). Each model was subjected to all permutations of these preprocessing options, with the objective of identifying the combination that yielded the best average performance across five folds. This exhaustive search culminated in selecting optimal preprocessing steps, which were then applied to the models that demonstrated the highest performance in the hyperparameter optimisation phase.

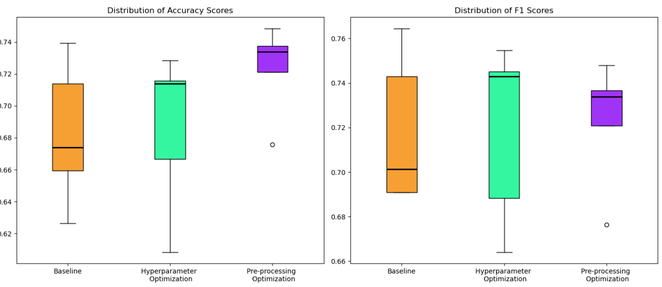

Fig. 2. Accuracy and F1 Score Distribution for Binary Classification Models

This tiered approach to optimisation—spanning model parameters and preprocessing techniques—aimed to rigorously enhance the models' ability to accurately classify and predict Alzheimer's Dementia stages from acoustic data, leveraging both stratified validation (binary and mutliclass classification approach) and standard K-fold validation technique (regression task approach) to split the data into train and tests sets and validate improvements at each stage.

IV. Results and Findings

The effectiveness of the proposed acoustic feature-based approach is evident in the binary and multiclass classification tasks. Our models underwent a series of optimisations, which systematically enhanced their performance.

A. Binary and Multiclass Classification Results

In binary classification, our models underwent a systematic refinement process, as outlined in Section III. This involved adjusting both the models' hyperparameters (HO) and their preprocessing procedures (PO) to improve upon the initial baseline version (BV). The significant improvements across versions can be seen in Table I, particularly for the k-Nearest Neighbors and Logistic Regression models, with the former achieving an increase in accuracy from BV to PO (0.6265 to 0.7211). The highest accuracy was achieved by the XGBoost model in the PO version, reaching an accuracy score of 0.7485 and an F1 score of 0.7479. Figure 2 further showcases the range and consistency of model performance throughout the different stages of optimisation, highlighting the effectiveness of our methodical approach to feature set adjustments.

The best performance of the XGBoost model can be attributed to its robust handling of non-linear relationships

TABLE I
ACCURACY AND F1 SCORE PROGRESSION TABLE FOR BINARY CLASSIFICATION MODELS

Model	BV	HO	PO
Accuracy Score			
Decision Tree	0.6594	0.6667	0.6758
k-Nearest Neighbors	0.6265	0.6083	0.7211
Logistic Regression	0.6739	0.7157	0.7376
Support Vector Classifier	0.7139	0.7139	0.7339
XGBoost Classifier	0.7395	0.7285	0.7485
F1 Score			
Decision Tree	0.6910	0.6884	0.6764
k-Nearest Neighbors	0.6908	0.6640	0.7209
Logistic Regression	0.7014	0.7452	0.7367
Support Vector Classifier	0.7431	0.7431	0.7339
XGBoost Classifier	0.7646	0.7548	0.7479

within the acoustic features, which are critical in distinguishing between stages of Alzheimer's Dementia. Additionally, Logistic Regression's performance improvement suggests its effectiveness in linearly separable data after optimal preprocessing, highlighting the importance of feature scaling and outlier management.

The multiclass classification task extends the complexity of AD detection by differentiating between healthy control, mild cognitive impairment, and Alzheimer's dementia. The improvements in accuracy and F1 scores from BV to PO are reported in Table II. Here, models like Decision Tree and k-Nearest Neighbors notably benefited from the optimisation process, emphasising the capability of acoustic features in fine-grained differentiation of cognitive states. The best score for this classification is achieved with the XGBoost model in the BV version for both accuracy and F1 score of 0.7215 and 0.6957 respectively. The box-plot of accuracy scores for the multiclass models is shown in Figure 3, providing a statistical overview of the models' performance.

TABLE II
ACCURACY OF MULTICLASS CLASSIFICATION MODELS

Model	BV	HO	PO
Accuracy Score			
Decision Tree	0.5756	0.5867	0.6273
k-Nearest Neighbors	0.6071	0.6071	0.6735
Random Forest	0.6828	0.6772	0.6974
XGBoost	0.7215	0.6975	0.7196
F1 Score			
Decision Tree	0.5768	0.5819	0.6301
k-Nearest Neighbors	0.5874	0.5874	0.6530
Random Forest	0.6539	0.6484	0.6722
XGBoost	0.6957	0.6709	0.6933

In the multiclass setting, the enhanced performance of Decision Trees following optimisation suggests an increased model capacity to manage overfitting through depth and leaf constraints, crucial in more granular classifications like distinguishing between MCI and AD.

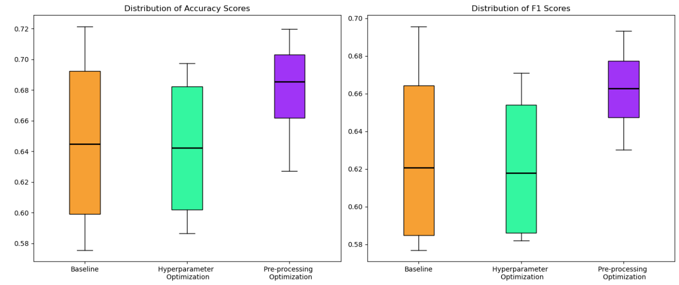

Fig. 3. Accuracy and F1 Score Distribution for Multiclass Classification Models

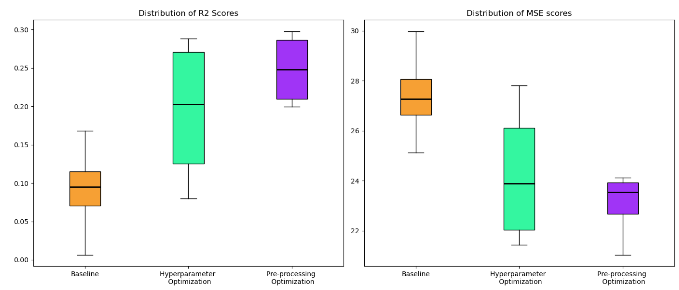

Fig. 4. R-squared and MSE Score Distribution for Regression Task Models

B. Regression Task Results

The regression tasks highlight the quantitative assessment of the predictive power of the optimised models. Table III showcases the R-squared and MSE scores, indicating the models' abilities to forecast the progression of AD based on acoustic features. The table also demonstrates an increased predictive accuracy from BV through to PO. The best evaluation scores were achieved with the Lasso Regressor being with R-squared value of 0.2979 in the PO version and lowest MSE score with 21.0307 as well in the PO version.

TABLE III
R-SQUARED VALUES FOR REGRESSION TASK MODELS

Model	BV	HO	PO
R-squared Score			
k-Nearest Neighbors Regressor	0.0063	0.0801	0.1995
Lasso Regressor	0.1681	0.1401	0.2979
SVR	0.0972	0.2646	0.2130
XGBoost Regressor	0.0922	0.2878	0.2822
MSE Score			
k-Nearest Neighbors Regressor	29.9835	27.8024	24.1243
Lasso Regressor	25.1327	25.5378	21.0307
SVR	27.4243	22.2256	23.8667
XGBoost Regressor	27.1283	21.4323	23.2108

The regression analysis particularly benefited from the Lasso Regressor, which effectively reduced model complexity by eliminating non-contributory features, a crucial factor when working with high-dimensional data like acoustic features. This model's ability to focus on significant predictors likely enhanced its performance in predicting MMSE scores.

C. Summary of Findings

In summary, the optimisation of acoustic feature-based machine learning models has yielded significant improvements in their ability to diagnose and quantify AD. The binary and multiclass classification results indicate that our models can effectively differentiate between the various cognitive states associated with AD, while the regression models demonstrate the capability to accurately predict the degree of cognitive decline. These findings validate the premise that acoustic features are robust indicators for AD detection and progression assessment, paving the way for future research to enhance the early diagnosis and monitoring of this condition.

V. Discussion and Limitations

This study's focus on acoustic features in detecting Alzheimer's Dementia (AD) demonstrates substantial diagnostic potential without the complexity of linguistic features. Research by Valsaraj et al. [9] supports this, showing that acoustic features alone can offer significant accuracy. Similarly, Luz [10] underscores the importance of feature selection to enhance model reliability, while Martinc and Pollak [11] reveal that combining acoustic with linguistic features can increase accuracy but also complexity, which may not be ideal in clinical applications.

Despite promising findings, our study confronts limitations with the DementiaBank dataset, which may not fully represent the diversity in education level, gender, and age. These demographic factors can significantly influence speech patterns and, therefore, the extraction and interpretation of acoustic features. For example, variations in education levels can affect linguistic abilities and speech complexity, which are crucial in diagnosing AD from acoustic data. Additionally, age and gender have been shown to influence voice and speech dynamics, potentially affecting the accuracy of acoustic diagnostics. Future research should not only expand to datasets that encompass a wider demographic range but also include a more rigorous control and analysis of these factors to enhance the robustness and utility of acoustic diagnostics in AD.

VI. Conclusion

In conclusion, this study demonstrates the effectiveness of using acoustic features for early detection of Alzheimer's Dementia (AD). Our results show that with the right optimisation techniques, machine learning models can use these features to accurately distinguish between different stages of cognitive impairment.

The most important insight from this work is that acoustic features alone have the potential to contribute significantly to AD diagnosis, potentially simplifying and improving current diagnostic procedures.

Future research should continue to refine these acoustic features and explore their applications in a broader range of datasets to confirm the findings. Additionally, leveraging advancements in deep learning could offer new methods for feature analysis, which may further improve diagnostic accuracy.

Ultimately, this study suggests that acoustic feature analysis is a valuable area of research with the potential to enhance AD detection methods, making it a promising direction for future study.

References

[1] (2018) Alzheimer's disease - treatment. Accessed: 2022-11-22. [Online]. Available: https://www.nhs.uk/conditions/alzheimers-disease/treatment/

[2] A. Javeed, A. Dallora, J. Berglund *et al.*, "Machine learning for dementia prediction: A systematic review and future research directions," *J Med Syst*, vol. 47, p. 17, 2023. [Online]. Available: https://doi.org/10.1007/s10916-023-01906-7

[3] Q. Zuo, B. Lei, Y. Shen *et al.* (2021) Multimodal representations learning and adversarial hypergraph fusion for early alzheimer's disease prediction. [Online]. Available: https://doi.org/10.48550/arXiv.2107.09928

[4] A. Bartoletti-Stella, M. Tarozzi, G. Mengozzi *et al.*, "Dementia-related genetic variants in an italian population of early-onset alzheimer's disease," *Frontiers in Aging Neuroscience*, vol. 14, 2022. [Online]. Available: https://doi.org/10.3389/fnagi.2022.969817

[5] J. Tuijl, "Screening for cognitive impairment in older general hospital patients: Comparison of the six-item cognitive impairment test with the mini-mental state examination," *International Journal of Geriatric Psychiatry*, vol. 27, pp. 755–762, 2012. [Online]. Available: https://doi.org/10.1002/GPS.2776

[6] F. Thaler and H. Gewald, "Language characteristics supporting early alzheimer's diagnosis through machine learning - a literature review," *Health Informatics - An International Journal*, vol. 10, pp. 5–23, 2021. [Online]. Available: https://doi.org/10.5121/HIIJ.2021.10102

[7] N. Cummins, Y. Pan, Z. Ren *et al.*, "A comparison of acoustic and linguistics methodologies for alzheimer's dementia recognition," in *Interspeech 2020*. ISCA, 2020, pp. 2182–2186.

[8] L. Cummings, "Describing the cookie theft picture," *Pragmatics and Society*, vol. 10, pp. 153–176, 2019. [Online]. Available: https://doi.org/10.1075/PS.17011.CUM

[9] A. Valsaraj, I. Madala, N. Garg, and V. Baths, "Alzheimers dementia detection using acoustic & linguistic features and pre-trained bert," arXiv.org, 2021.

[10] S. Luz, F. Haider, S. de la Fuente *et al.*, "Alzheimer's dementia recognition through spontaneous speech: The adress challenge," arXiv.org, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2004.06833

[11] M. Martinc and S. Pollak, "Tackling the adress challenge: A multimodal approach to the automated recognition of alzheimer's dementia," in *Interspeech*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:226206210

[12] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The Natural History of Alzheimer's Disease: Description of Study Cohort and Accuracy of Diagnosis," *Archives of Neurology*, vol. 51, no. 6, pp. 585–594, Jun. 1994. [Online]. Available: https://doi.org/10.1001/archneur.1994.00540180063015

[13] R. B. Ammar and Y. B. Ayed, "Speech processing for early alzheimer disease diagnosis: Machine learning based approach," *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–8, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:58669778

[14] P.-Y. Chen, S.-Y. Hong, W.-T. Cheah, L.-C. Fu, and Y.-L. Chang, "An assessment system for alzheimer's disease based on speech using a novel feature sequence design and recurrent neural network," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3289–3294, 10 2018.

[15] S. Mirzaei, M. El Yacoubi, S. Garcia-Salicetti, J. Boudy, C. Kahindo, V. Cristancho-Lacroix, H. Kerhervé, and A.-S. Rigaud, "Two-stage feature selection of voice parameters for early alzheimer's disease prediction," *Irbm*, vol. 39, no. 6, pp. 430–435, 2018.

[16] G. Gosztolya, V. Vincze, L. Tóth, M. Pákáski, J. Kálmán, and I. Hoffmann, "Identifying mild cognitive impairment and mild alzheimer's disease based on spontaneous speech using asr and linguistic features," *Computer Speech & Language*, vol. 53, pp. 181–197, 2019.