Speech-Based Alzheimer's Disease Classification System with Noise-Resilient Features Optimization

Virender Kadyan
Speech and Language Research Centre,
University of Petroleum and
Energy Studies, Dehradun,India
vkadyan@upes.ddn.ac.in

Puneet Bawa
Chitkara University Institute of
Engineering and Technology,
Chitkara University, Punjab, India
puneet.bawa@chitkara.edu.in

Mohd Mujtaba Akhtar
Speech and Language Research Centre,
University of Petroleum and
Energy Studies, Dehradun,India
500084392@stu.upes.ac.in

Muskaan Singh
Cognitive Analytics Research Lab
Intelligent Systems Research Centre,
School of Computing, Ulster University, UK
m.singh@ulster.ac.uk

***Abstract*—Alzheimer's disease is a severe neurological disorder having a major influence on a substantial portion of the population. The prompt detection of this condition is crucial, and speech analysis may play a crucial role in facilitating efficient treatment and care. The main aim of this research has been to investigate the significance of timely identification of speech signal abnormalities associated with Alzheimer's disease in order to provide effective therapy interventions and improve disease management. The study used the Mel Frequency Cepstral Coefficients (MFCC) framework, a well recognized technique for feature extraction known for its versatility across several domains. This research introduces an innovative approach that utilizes both individuals diagnosed with dementia and control participants to detect two unique types of cognitive impairment via the analysis of speech signals. The approach used in this work involves the extraction of acoustic properties from pre-processed speech data obtained from the Pitt Corpus of Dementia Bank. This is achieved by using several feature sets, which include a combination of MFCC, prosodic features, and statistical features. This study examines the attributes of optimum feature optimization in actual and noise-enhanced speech environments using machine learning techniques. The integration of MFCC,Statistical and prosodic features has shown remarkable outcomes, exhibiting a superior accuracy rate of 98.3%. This surpasses the performance of other feature combinations when using the Random Forest classifier.**

***Index Terms*—Alzheimer's Disease, MFCC features, Prosodic Feature, Statistical Feature, Machine Learning, Classification**

I. Introduction

Dementia, a prevalent neurological condition, has a significant impact on the cognitive and behavioural functioning of those afflicted by the condition. Alzheimer's disease (AD) is responsible for the majority of instances of dementia, affecting a population exceeding 50 million individuals [1]. AD is responsible for a substantial percentage, ranging from 60% to 80%, of dementia cases. As the disease advances, persons affected by this disorder may have a significant deterioration in both their general life expectancy and interpersonal relationships [2, 3]. It is expected that this figure may quadruple by the year 2050. AD presents challenges in the realm of verbal communication, resulting in impairments in memory and speaking, especially in the first stages. Automated computational methods that use machine learning and speech analysis play a crucial role in promptly identifying and implementing therapeutic measures for AD [4, 5]. The lack of a definitive treatment for AD amplifies its impact on individuals and their overall state of health. Hence, the use of automated computational methods that integrate machine learning trained models, together with speech analysis, has great importance in the context of early detection, diagnosis, and therapeutic treatments for AD.

Initially, Rentoumi et al. [6] first integrated retrospective linguistic data analysis with robust machine learning classification techniques. The approach included an automated feature selection technique together with machine learning techniques, which yielded a diagnostic accuracy ranging from 67% to 75%. In addition, König et al. [7] employed automated speech analysis as a means of evaluating Mild Cognitive Impairment (MCI) and early-stage AD. The study yielded a 79% accuracy rate in differentiating healthy individuals from those with MCI, an 87% accuracy rate in distinguishing healthy individuals from those with AD, and an 80% accuracy rate in distinguishing between MCI and AD patients. Furthermore, Pappagari et al. [4] conducted a comprehensive analysis of studies that used advanced speaker recognition techniques and language models to include prosody information. The results indicated a collaborative relationship between acoustic and language models, yielding an optimal accuracy rate of 84.51%. Lately, a dataset including transcripts of individuals with AD who speak Nepali and a control group was provided by Karande and Kulkarn [8] via the use of machine learning and deep learning methods.

The study emphasised that the information processing challenges faced by AD patients are reflected in their speech narratives, particularly during picture description tasks. In this paper, an effort has been made to design and implement a machine learning based system for the effective classification

979-8-3503-6021-9/23/$31.00 © 2023 IEEE

of AD using speech data as the primary form of input. The primary objectives of the study have been

- to examine the significance of selecting features in enhancing the efficacy of predictive ML models for the categorization of Alzheimer's disease.
- to investigate the impact of noise on the efficacy of the categorization system for AD.

The remaining parts of the research paper are organized as follows: Section 2 provides a comprehensive overview of the dataset, including information on data preparation, data augmentation, and the recommended method. Moreover, Section 3 provides an overview of the findings and discussion, while Section 4 presents a comprehensive analysis of the conclusion and future work.

II. Methodology

A. Dataset Details

The present study utilizes the DementiaBank Pitt dataset [9], which comprises many iterations, each including diverse quantities of audio recordings from both male and female speakers. The dataset has a total of 166 voice recordings, all of which were used in the training set. Out of the total number of recordings, 87 were sourced from persons who had received a diagnosis of AD, whilst the other 79 recordings were acquired from individuals who were classified as healthy controls. The content of each recording consists of a visual depiction, first intended for use in the Boston Diagnostic Aphasia Examination [10].

B. Noise Augmentation

The technique of noise augmentation is being used to enhance the size of the Pitt Corpus dataset inside the DementiaBank project. The deliberate injection of background noise ($Noise$) into clean audio set (Sig_{Clean}) has been performed in order to enhance the authenticity and effectively replicate real-world scenarios and enhance the versatility of datasets. The methodology used in this study was the generation of noisy signal (Sig_{Noisy}) replication of noise levels ranging from moderate to high, using a signal-to-noise ratio (SNR) within the range of 15 to 20 dB using equation (1).

$$Sig_{Noisy} = Sig_{Clean} + \frac{Power_{Noisy}}{\sqrt{SNR}} \times Noise \qquad (1)$$

The approach effectively maintains a balance between the clarity of vocal communication and the presence of authentic background noise [11]. However, it should be noted that SNRs that are lower, often below 15 dB, are intentionally excluded from the augmentation procedure as it can obscure the primary speech stream and diminish its overall quality.

C. Proposed System Architecture

The proposed system architecture, as seen in Figure 1, illustrates the inclusion of feature extraction as a foundational component within the proposed architectural framework. In the first step, a total of *13* Mel-Frequency Cepstral Coefficients (MFCC) are computed from a Mel filterbank consisting of *40* channels. This computation is performed using a hamming window with a frame duration of *25ms* and a *50%* overlap ratio [12]. The calculation of features $M(f)$ is based on equation (2), which aims to capture relevant spectral characteristics where f corresponds to the central frequency of an audio signal. Furthermore, the MFCC characteristics that were retrieved provided insights into the articulation of speech and the phonetic content of each speech sample.

$$M(f) = 1125 * ln(1 + \frac{f}{700}) \qquad (2)$$

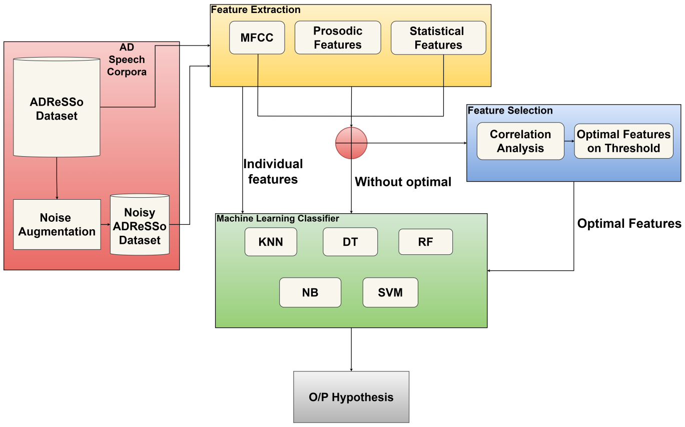

Fig. 1: Block diagram depicting the proposed optimal feature selection-based noise-robust AD classification system using hybrid front end

Moreover, the retrieval of prosodic information is undertaken to encompass elements such as rhythm and intonation, therefore unveiling psychological and expressive dimensions of speech. The statistical data sheds light on speech patterns and variability since it involves the examination of seven statistical variables (namely, IQR, standard deviation, Q25, Q75, min-fun, skewness, and centroid) for further investigation. Similarly, the correlation coefficients of the acquired features were computed for the purpose of conducting correlation analysis facilitating the identification of duplicate or highly correlated variables that may not enhance the predictive capability of the AD classification system for the development of a concise modeling approach as detailed in Algorithm 1.

The various machine learning methods including k-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM) have been utilized to learn from characteristic recordings of healthy and AD patients. A systematic approach trains models on MFCC, prosodic, and statistical features to understand their separate contributions. For optimal utilization of their collective knowledge, the models are trained using concatenated features, which combine all three feature extraction methods. Next, the models are trained using the ideal feature subset to create models with enhanced performance. The model's performance has been tested in clean and loud situations.

Algorithm 1 AD Classification with feature optimization using speech analysis

```
Input:
  • DementiaBank Pitt dataset : $AD_{data}$
  • Background Noise: $N_{back}$
  • SNR levels (SNR) : $SNR_{best}$
Output:
  • Best performing model : $M_{best}$
  • Optimal feature set : $F_{best}$
  • Model performance : $PM$

Step 1: Initialize and Load $AD_{data}$ data.
Step 2: Split $AD_{data}$ into $AD_{normal}$ and $NON-AD_{effected}$
Step 3: Perform Noise Augmentation
    For each in $SNR_{best}$ :
        $S_{SNR}(X), N_{back,j} \leftarrow addNoise(x, N_{back,j})$
Step 4: Perform feature selection
    Step 4.1: Extract MFCC
            For each (x) in $AD_{data}$ :
                $MFCC(x) \leftarrow DCT(log(FFT(x)))$
    Step 4.2: Extract prosody features : P(x)
    Step 4.3: Extract statistical features : S(x)
Step 5: Concatenate the features :
$M_{concat} \leftarrow \mathrm{F}_i | F_i$ from MFCC(x),P(x) and S(x)
Step 6: Calculate the correlation analysis on $M_{concat}$:
    Step 6.1: Calculate the correlation matrix :
            $C_{i,j} \leftarrow corr(M_{concat(i)}, M_{concat(i)})$
    Step 6.2: Define the correlation threshold
            $F_{best} \leftarrow (C_{i,j}, || \, C_{i,j}) \, | > T_{value}$
Step 7: Train ML model to obtain ($F_{best}$) on performance matrices
$PM \leftarrow Accu, Precision, Recall, F1-Score$
$mdl1 \leftarrow KNN(\mathrm{F}_{best})$ //train K-nearest-neighbor
$mdl2 \leftarrow DT(\mathrm{F}_{best})$ //train Decision Tree
$mdl3 \leftarrow RF(\mathrm{F}_{best})$ //train Random Forest
$mdl4 \leftarrow SVM(\mathrm{F}_{best})$ //train Support Vector Machine
$mdl5 \leftarrow NB(\mathrm{F}_{best})$ //train Naïve Bayes
Step 8: Obtain best model on best PM
$M_{best} \leftarrow best(Accu, (mdl1, mdl2, mdl3, mdl4, mdl5))$
```

By simulating real-life conditions with ambient noise, we can assess the robustness and flexibility of each model. A comprehensive examination of the model's usefulness in various scenarios involves examining performance indicators, including accuracy, precision, recall, and F1-score.

III. Results and Discussions

The experiments conducted assess the performance of several machine-learning approaches, including KNN, DT, RF, NB, and SVM, in both clean and noisy environments. The dataset had been divided into an 80-20% ratio, with 80% of the data used for training and 20% used for testing. The division was performed by dividing the dataset into k subsets, where $k = 5$. The model was trained five times, each time utilizing a distinct subset as the set used for testing and the remainder of the data for training. Furthermore, the experiments investigated the impact of different feature selection methods on the performance of these techniques where the process of MFCC feature extraction was done using MATLAB programming environment. Further, Open Smile Toolkit [13] has been used for the process of prosodic feature extraction presenting a comprehensive set of 31 prosodic components used to ascertain physiological voice characteristics. Lastly, statistical features are extracted using the Librosa library [14].

A. Performance Evaluation on Clean Conditions

In the first set of experiments, the performance evaluation for developing the AD classification system has been performed on a clean environment. MFCC is commonly utilized in speech analysis due to its ability to maintain audio spectrum information under clean conditions. However, the MFCC may not capture all of the disorder's speech problems. The findings in Table I showcased the best accuracy on the SVM classifier outperforming the other classifiers with an accuracy of 80%. In this set of experiments, MFCC elements are incorporated into the recommended task utilizing prosodic and statistical features where the feature space of MFCC is enlarged by including prosodic and statistical properties. The amalgamated features have aided machine learning models in capturing Alzheimer 's-related speech patterns improvising the model's ability to generalize and identify AD related speech patterns.

TABLE I: Accuracy Obtained on Heterogenous Classification Technique using MFCC Approach on Clean Conditions.

Classifiers	Accuracy	Precision		Recall		F1-Score	
		0	1	0	1	0	1
DT	0.65	0.56	0.73	0.62	0.67	0.59	0.70
RF	0.55	0.47	0.80	0.88	0.33	0.61	0.47
NB	0.60	0.50	0.67	0.50	0.67	0.50	0.67
SVM	**0.80**	0.75	0.83	0.75	0.83	0.75	0.83
KNN	0.65	0.56	0.73	0.62	0.67	0.59	0.70

B. Performance Evaluation using Optimal Feature Selection approach

In the next set of experiments, a feature selection technique based on the co-relation analysis has been employed to ascertain the optimal subset of features. The identification of a specific subset comprising three features that consistently exhibited superior performance compared to the whole concatenated set of features in both the RF and SVM models holds considerable importance. These findings in Table II implies that the amalgamated features of the optimal selection have been found to be more efficient for the classification of AD. The RF and SVM algorithms with accuracy of 90% and 91% respectively have been shown to be durable and capable of effectively handling intricate feature spaces. The importance of feature selection is underscored by the advantage gained by these models via the use of a smaller feature set, particularly in situations involving high-dimensional data such as sound analysis.

TABLE II: Accuracy obtained on fusion of conventional MFCC with various combinations of prosodic and statistical features

Classifiers	Accuracy	Precision		Recall		F1-Score	
		0	1	0	1	0	1
DT	0.70	0.60	0.80	0.75	0.67	0.67	0.73
RF	0.90	0.80	1.00	1.00	0.83	0.89	0.91
NB	0.80	0.75	0.83	0.75	0.83	0.75	0.83
SVM	**0.91**	0.80	1.00	1.00	0.83	0.89	0.91
KNN	0.80	0.75	0.83	0.75	0.83	0.75	0.83

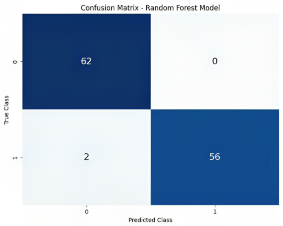

Fig. 2: Confusion matrix optimal features using Random Forest

C. Performance Evaluation on Noisy Conditions

In this section, noise augmentation is used to provide real-world variability to test datasets. This method captured genuine scenario challenges more accurately. The significance of feature selection and model resilience is shown by the rising effectiveness of RF-Optimal (RF-O) and SVM-Optimal (SVM-O) techniques in environmental noise. This shows that these approaches with optimal feature selection can manage noisy real-world datasets. The high accuracy rate of 98.3% demonstrates the effectiveness of the feature selection approach and machine learning models in generalizing and adapting to various scenarios in comparison to non-optimal features based classifiers RF and SVM, as shown in Table III. Figure 2 displays the confusion matrix for selecting features in noisy situations for developing a noise-robust AD classification system.

TABLE III: Accuracy obtained on optimal and non-optimal features using noise augmentation approaches

Classifiers	Accuracy	Precision		Recall		F1-Score	
		0	1	0	1	0	1
RF	0.97	0.97	0.97	0.97	0.97	0.97	0.97
SVM	0.93	0.93	0.95	0.92	0.95	0.93	0.93
RF-O	**0.98**	0.97	1.00	1.00	0.97	0.98	0.98
SVM-O	0.93	0.96	0.89	0.89	0.97	0.92	0.93

IV. Conclusion

This study uses conventional MFCC methodologies to propose three distinct feature pairings. In order to achieve optimal feature selection, dominant and high-energy information is obtained by retrieving prosodic and statistical features. Several experiments have been conducted to examine three different forms of integrated MFCC features in both clean and noisy situations. In order to address the problem of AD classification in the picture description job, five classifiers were used to assess the robustness of the proposed system. Nevertheless, it is worth noting that basic models exhibit comparable performance to more complex models when applied to both unaltered and improved speech samples. The results indicate that the amalgamation of MFCC, prosodic, and statistical optimal features, when used with an RF classifier, has the ability to accurately detect AD with an accuracy of 98.3%. The integration of audio and MRI images, via the use of data augmentation methods, in conjunction with advanced front and back-end approaches, has the potential to broaden the scope of research in this field.

References

[1] K. Ritchie, I. Carriere, L. Su, J. T. O'Brien, S. Lovestone, K. Wells, and C. W. Ritchie, "The midlife cognitive profiles of adults at high risk of late-onset alzheimer's disease: The prevent study," *Alzheimer's & Dementia*, vol. 13, no. 10, pp. 1089–1097, 2017.

[2] T. Behl, G. Kaur, A. Sehgal, S. Bhardwaj, S. Singh, C. Buhas, C. Judea-Pusta, D. Uivarosan, M. A. Munteanu, and S. Bungau, "Multifaceted role of matrix metalloproteinases in neurodegenerative diseases: Pathophysiological and therapeutic perspectives," *International Journal of Molecular Sciences*, vol. 22, no. 3, p. 1413, 2021.

[3] T. Bhattacharya, G. A. B. e. Soares, H. Chopra, M. M. Rahman, Z. Hasan, S. S. Swain, and S. Cavalu, "Applications of phyto-nanotechnology for the treatment of neurodegenerative disorders," *Materials*, vol. 15, no. 3, p. 804, 2022.

[4] R. Pappagari, J. Cho, S. Joshi, L. Moro-Velázquez, P. Zelasko, J. Villalba, and N. Dehak, "Automatic detection and assessment of alzheimer disease using speech and language technologies in low-resource scenarios.," in *Interspeech*, vol. 2021, pp. 3825–3829, 2021.

[5] Y.-W. Chien, S.-Y. Hong, W.-T. Cheah, L.-H. Yao, Y.-L. Chang, and L.-C. Fu, "An automatic assessment system for alzheimer's disease based on speech using feature sequence generator and recurrent neural network," *Scientific Reports*, vol. 9, no. 1, p. 19597, 2019.

[6] V. Rentoumi, L. Raoufian, S. Ahmed, C. A. de Jager, and P. Garrard, "Features and machine learning classification of connected speech samples from patients with autopsy proven alzheimer's disease with and without additional vascular pathology," *Journal of Alzheimer's Disease*, vol. 42, no. s3, pp. S3–S17, 2014.

[7] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, *et al.*, "Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 1, no. 1, pp. 112–124, 2015.

[8] S. Karande and V. Kulkarni, "Automated prognosis of alzheimer's disease using machine learning classifiers on spontaneous speech features," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 245–251, 2023.

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

[10] J. C. Borod, H. Goodglass, and E. Kaplan, "Normative data on the boston diagnostic aphasia examination, parietal lobe battery, and the boston naming test," *Journal of Clinical and Experimental Neuropsychology*, vol. 2, no. 3, pp. 209–215, 1980.

[11] P. Bawa and V. Kadyan, "Noise robust in-domain children speech enhancement for automatic punjabi recognition system under mismatched conditions," *Applied Acoustics*, vol. 175, p. 107810, 2021.

[12] V. Kadyan, A. Mantri, and R. Aggarwal, "A heterogeneous speech feature vectors generation approach with hybrid hmm classifiers," *International Journal of Speech Technology*, vol. 20, pp. 761–769, 2017.

[13] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, pp. 1459–1462, 2010.

[14] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, pp. 18–25, 2015.