Classification of Right Hemisphere Damage Using MFCC Paralinguistic Voice Features

Visar Rraci
Computer Science Department
Louisiana State University Shreveport
Shreveport, USA
rraciv24@lsus.edu

Landrum Anderson
Computer Science Department
Louisiana State University Shreveport
Shreveport, USA
andersonl54@lsus.edu

Subhajit Chakrabarty
Computer Science Department
Louisiana State University Shreveport
Shreveport, USA
subhajit.chakrabarty@lsus.edu

2025 Intermountain Engineering, Technology and Computing (IETC) | 979-8-3315-1282-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/IETC64455.2025.11039385

***Abstract*—Right Hemisphere Damage (RHD) can profoundly impact the melodic and rhythmic qualities of speech—collectively known as prosody—and lead to social communication difficulties. This study proposes a machine learning workflow to identify RHD-related paralinguistic deficits from single-source audio. Audio signals from the NIH-funded TalkBank RHD dataset were denoised, voice activity filtered, and normalized, and Mel Frequency Cepstral Coefficients were extracted. Multiple supervised learning algorithms (Support Vector Machine, Random Forest, Logistic Regression, and k-Nearest Neighbors) were then trained separately and evaluated. Among these models, the Support Vector Machine with a radial basis function kernel achieved the highest accuracy of 0.79, indicating that short, standardized speech samples contain diagnostic cues for RHD. The contribution of this study is that this is the first application of such methods on the dataset for the objective. This approach highlights the feasibility of automated, non-invasive detection of RHD and offers a promising direction for adjunctive clinical assessment tools.**

Keywords— Right Hemisphere Damage, Paralinguistic, Apragmatism, Speech, Classification, Machine Learning

I. Introduction

A. Background

The right hemisphere of the human brain plays an important role in the melodic and rhythmic qualities of speech—often referred to collectively as *prosody*—as well as the ability to grasp context and implied meaning in conversation [1], [2]. When someone experiences right-hemisphere brain damage (RHD), they may develop "aprosodia," where speech becomes monotonous or intonation sounds inappropriate, and they often struggle to interpret emotional tone [2]. About half of individuals with RHD exhibit some level of these communication difficulties, which can include problems understanding non-literal language (e.g., sarcasm, humor, or idioms) and organizing thoughts coherently during conversation [1][3][4].

Traditional clinical tests for RHD (e.g., RHLB, MIRBI-2) can help identify major prosodic or discourse problems, yet they can be somewhat limited in scope and might miss subtle changes [3][4]. These tools typically rely on human scoring and qualitative judgments, which makes the evaluation subjective and potentially time-consuming. By contrast, recent advances in computer-based analysis of speech have shown promise in automatically detecting subtle cues that might indicate specific types of language impairments [5][6].

B. Objective

This study aims to create and evaluate a machine learning approach for identifying RHD-related speech deficits from short audio samples. Specifically, we focus on extracting acoustic features known as Mel-Frequency Cepstral Coefficients (MFCCs) to capture the unique spectral characteristics of speech in individuals with and without RHD. We then compare multiple supervised classifiers (Random Forest, Support Vector Machine, Logistic Regression, and k-Nearest Neighbors) under a cross-validation framework that groups samples by individual, ensuring robust generalization. We experimented with FastICA-extracted features and pitch contours, but they did not yield promising results.

C. Significance

Clinical detection of RHD currently depends on subjective and time-consuming assessments [3][4]. Having an automated tool could reduce evaluation time, streamline patient screening, and help in early intervention planning. Previous studies in stroke and speech-impairment research have indicated that machine learning approaches can reliably detect subtle communication deficits when properly tuned to acoustic markers [5][6]. Extending these methodologies to RHD offers a more comprehensive and non-invasive way to track changes in prosody and pragmatic skills, complementing existing clinical tests and enhancing overall patient care.

D. General Layout

The remainder of this paper is organized as follows. Section II discusses the previous work in this domain. Section III details the methods used. Section IV presents the results, including confusion matrices, classification reports, and ROC curves for the models, as well as a discussion. Section V concludes the paper with the limitations, and possible future work.

II. Previous work

A. Traditional RHD Classification and Deficits

Clinicians have developed several RHD-specific assessment batteries, such as the Right Hemisphere Language Battery (RHLB), Mini Inventory of Right Brain Injury (MIRBI-2), and RIC Evaluation of Communication Problems in Right Hemisphere Dysfunction (RICE-3) [3]. While useful, these tools have notable limitations. They often focus on select components (for example, metaphors, affective prosody) and may not thoroughly evaluate pragmatic communication or non-verbal cues (gesture, facial expression) [3]. Consequently, subtle deficits can be missed—the tests may not be sensitive to mild pragmatic impairments, and scoring of discourse/pragmatics is often subjective [3]. In practice, clinicians must supplement these batteries with qualitative observation to fully capture RHD-related communication issues.

B. Broader Stroke and Leison Classification

Contemporary machine learning methods offer several tools to detect RHD-related speech profiles. A crucial step is extracting acoustic features that reflect prosody and voice characteristics. Mel-frequency cepstral coefficients (MFCCs) are widely used features that encapsulate the spectral shape of speech; they have proven effective in many voice classification tasks, and are useful for RHD as well [6].

A variety of machine learning models have been employed for speech-based classification, and comparisons of their performance can guide RHD detection efforts. For example, one study found that SVM outperformed other machine learning models with an accuracy of 0.74, even beating a convolutional neural network(CNN) in most metrics [7].

A practical strategy is to begin with interpretable models like SVM or logistic regression using expert-crafted features (pitch statistics, MFCCs, etc.), and later compare against more complex models. As research advances, we expect comparative evaluations of models on RHD speech corpora to identify which algorithms yield the highest diagnostic accuracy. Early indications from related domains show that machine learning can reliably detect even subtle speech impairments, so with the right features and model tuning, it is feasible to classify RHD presence or absence from an individual's speech profile.

III. Methods

A. Dataset Sources and Description

The dataset used in this project was obtained from the NIH-funded TalkBank Right Hemisphere Damage (RHD) database [8]. Our curated set consisted of 32 high-quality audio recordings: 16 from individuals diagnosed with RHD and 16 from healthy control subjects. Recordings were selected to exclude any files containing excessive background noise or significant audio artifacts, such as those introduced by Zoom recording software, thus preserving the integrity of subsequent feature extraction.

To expand the dataset, each recording, which originally contained multiple sentences or conversational turns, was segmented into individual sentence-level clips. Speech was divided into segments of up to 7 seconds. If a continuous segment exceeded this duration, it was split into consecutive 7-second clips. Recordings with less than 7 seconds of speech were zero-padded to reach the full duration. This preprocessing step increased the number of usable audio samples from the 32 original 45-minute interviews to a total of 1,357 segmented clips.

Crucially, speaker identity was preserved as a grouping variable to maintain evaluation integrity. No audio segments from the same individual were included in both training and testing sets, ensuring speaker-independent evaluation even after data expansion.

While this segmentation approach increased the number of samples, it may introduce subtle bias. For instance, if a speaker, particularly of one gender, spoke for a longer duration, they would contribute more samples to the dataset. However, because the RHD and control groups included comparable numbers of male and female speakers, and average recording durations were similar across individuals, any major gender imbalance was minimized. Potential gender-related differences in speech duration or acoustic characteristics may still influence the models, and these effects are examined in the error analysis.

B. Audio Preprocessing

Audio signals were collected and stored in uncompressed WAV format (16-bit PCM). Each file was resampled to 16 kHz using the Librosa library in Python [9]. Prior to feature extraction, the audio underwent noise reduction by applying a spectral subtraction algorithm (via the noisereduce package) that attenuates stationary background noise [10]. Voice activity detection (VAD) was then performed with the WebRTC VAD algorithm , which adaptively filters out non-speech frames [11]. Specifically, 30 ms frames were iteratively classified as speech or silence under a moderate aggressiveness setting (level=2). Only the segments flagged as speech were retained, thereby reducing the influence of silence and background interference.

To ensure consistent duration, the extracted speech segments were padded or truncated to 7 s per audio file. This standardized input length addresses potential variability in recording durations and prevents bias from longer samples. Each waveform was then energy-normalized by dividing samples by the maximum absolute amplitude, capping the sample values at -1 to 1.

All methods were implemented in Python 3.9, with dependencies managed via pip. A high-level summary of the preprocessing and feature-extraction pipeline is as follows: (1) load and resample using librosa, (2) perform noise reduction, (3) apply VAD, (4) pad or truncate to 7 s, and (5) normalize amplitude.

C. Feature Extraction with MFCCs

Mel-Frequency Cepstral Coefficients (MFCCs) were chosen as the primary acoustic features for classification, given their widespread use in speech processing and robust performance in characterizing spectral envelopes [12]. A set of 13 MFCCs was computed for each processed audio clip using a 25 ms Hamming-windowed frame with a 10 ms hop. The MFCC computation included a discrete cosine transform (DCT-II) on the log-Mel spectra. To obtain a fixed-length

feature vector, the mean and standard deviation of each of the 13 MFCC coefficients across all frames were concatenated, yielding a 26-dimensional representation per sample (13 means and 13 standard deviations). These feature vectors, along with their corresponding class labels, were exported to a CSV file for downstream machine learning classification.

D. Machine Learning Classification

This study employed a group-based cross-validation procedure to train and evaluate multiple machine learning classifiers on the extracted MFCC features. All implementations were performed in Python 3.9 using scikit-learn [13]. A CSV file containing MFCC feature columns, a binary class label, and a file path was loaded into a Pandas DataFrame. Because multiple audio samples could come from the same patient, each file path was parsed to extract a "group_id," ensuring that data from the same individual were never split across training and test folds. The GroupKFold class from scikit-learn partitioned the dataset into k disjoint folds, keeping all samples of a given "group_id" together [13]. All MFCC features were standardized via z-score normalization. Four classifiers were compared under the same cross-validation scheme: (1) Random Forest, a bagged ensemble of decision trees, with n_estimators=100 and random_state=42; (2) Support Vector Machine (RBF), an SVM with a radial basis function kernel; (3) Logistic Regression, a linear method for binary classification; and (4) k-Nearest Neighbors (KNN), a distance-based classifier that labels test samples by the majority vote of their k closest training neighbors [13].

E. Evaluation Metrics

For each fold in the GroupKFold procedure, out-of-sample probability estimates were obtained via cross_val_predict using the method="predict_proba" option. These probabilities were then converted into binary predictions by applying a 0.5 threshold. The performance evaluation focused on four main metrics. First, a confusion matrix was constructed to show counts of true positives, false positives, true negatives, and false negatives, illustrating the distribution of correct and incorrect predictions. Second, a classification report was generated to present precision, recall, and F1-scores for each class, along with the overall macro- and micro-averaged scores. Third, the accuracy of each model was calculated as the proportion of correctly classified samples across all predictions. Finally, receiver operating characteristic (ROC) curves were plotted by comparing the true labels to the predicted probabilities for the positive class, and the area under the ROC curve (AUC) was computed to quantify discriminative ability [14]. A single combined plot was produced to visualize and compare the ROC curves and AUC values of all evaluated models. Once cross-validation was complete, the best-performing classifier (based on accuracy) was retrained on the entire dataset, and both the trained model and the feature-scaling parameters were saved using Joblib for future use.

IV. Results and Discussion

A. Results

Table 1 presents the performance of four classification models evaluated on the segmented speech dataset: Random Forest, Support Vector Machine with Radial Basis Function kernel (SVM RBF), Logistic Regression, and k-Nearest Neighbors (k-NN). Performance metrics are reported separately for the Control and Afflicted classes, including precision, recall, and F1-score. In addition, overall accuracy (Accu) and area under the ROC curve (AUC) are provided for each model.

Among all models, the SVM RBF classifier achieved the highest performance across multiple metrics. It yielded the highest overall accuracy of 0.79 and the highest AUC of 0.88, indicating superior discriminative capability. It also demonstrated strong performance on class-specific metrics, with F1-scores of 0.80 (Control) and 0.78 (Afflicted), and the highest recall for the Control class (0.83). Logistic Regression also performed well, achieving an AUC of 0.86 and balanced F1-scores of 0.78 (Control) and 0.77(Afflicted). Random Forest and k-NN classifiers performed moderately, with overall accuracies of 0.74 and 0.75, and AUC values of 0.83 and 0.81, respectively.

Figures 1–4 show the confusion matrices for each model. The SVM RBF classifier (Fig. 1) correctly identified 568 control and 500 afflicted samples, achieving the best balance between sensitivity and specificity. The Random Forest model (Fig. 2) had higher misclassification rates, especially among afflicted samples. k-NN (Fig. 3) showed moderate performance, while Logistic Regression (Fig. 4) achieved the lowest number of false negatives, aligning with its strong F1 scores in Table 1.

Fig. 5 shows the ROC curves for all models. SVM RBF achieved the highest AUC (0.88), indicating the best overall discriminative performance. Logistic Regression followed closely (AUC = 0.86), while Random Forest and k-NN had lower AUCs of 0.83 and 0.81, respectively, consistent with their performance in Table 1.

Table 1. Model Performance

Model Name	**Classification Metrics**							
	Control			Afflicted			Accu	AUC
	Prec	Rec	F1	Prec	Rec	F1		
Random Forest	0.73	0.76	0.74	0.74	0.715	0.73	0.74	0.83
SVM RBF	**0.77**	**0.83**	**0.8**	**0.81**	**0.74**	**0.78**	**0.79**	**0.88**
Logistic Regression	0.77	0.79	0.78	0.78	0.76	0.77	0.77	0.86
k-Nearest Neighbors	0.74	0.77	0.76	0.76	0.73	0.74	0.75	0.81

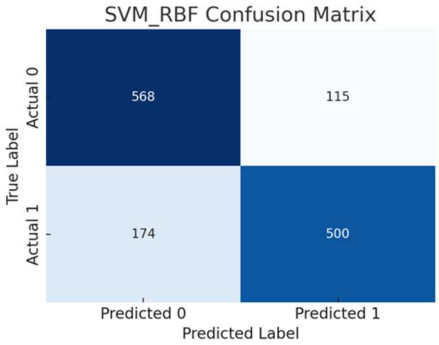

Fig. 1. SVM_RBF Confusion Matrix

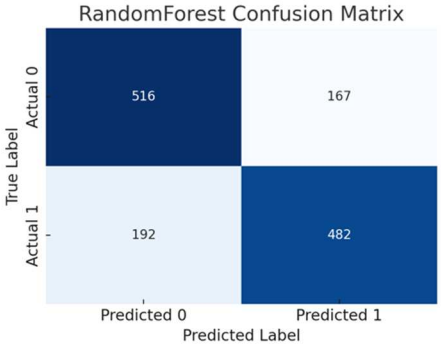

Fig. 2. Random Forest Confusion Matrix

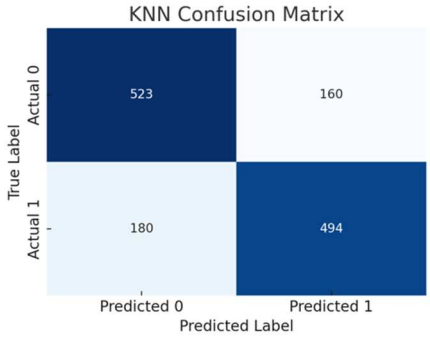

Fig. 3. KNN Confusion Matrix

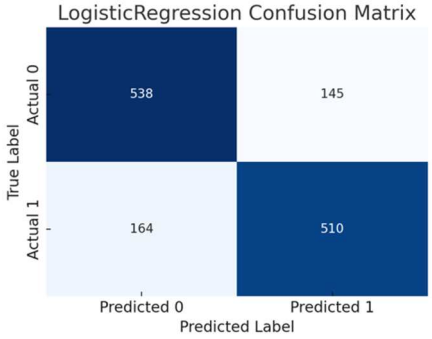

Fig. 4. Logistic Regression Confusion Matrix

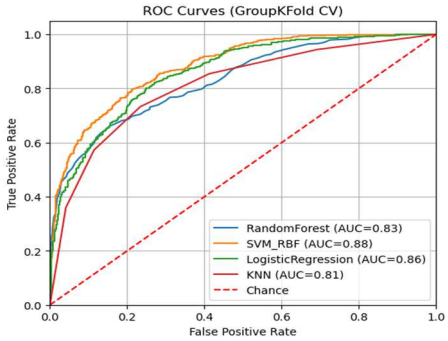

Fig. 5. ROC Curve

B. Discussion

In this study, we demonstrated the feasibility of using MFCC extracted features and machine learning to detect right hemisphere damage (RHD) from paralinguistic speech characteristics. Using a consistent preprocessing pipeline and group-based cross-validation, we evaluated four classifiers. The Support Vector Machine (SVM) with an RBF kernel achieved the highest accuracy (~79%) and AUC (0.88), indicating MFCCs effectively capture acoustic markers of RHD. The observed accuracy (~74–79%) notably surpasses random chance (~50%), suggesting meaningful discrimination between RHD and non-RHD speech.

Precision and recall were balanced for both classes, indicating no significant bias toward healthy controls. For instance, SVM showed a recall of 0.74 and precision of 0.81 for the RHD class. Logistic Regression offered slightly higher recall (0.76) but lower precision, providing options for clinical scenarios based on acceptable false-negative or false-positive rates.

Our error analysis, stratified by gender (the only available demographic), revealed consistent performance across male (~80%) and female (~78%) speakers, with no systematic gender bias in classification errors. Maintaining gender balance during training likely encouraged the learning of gender-invariant features. Still, subtle gender-related acoustic differences, such as pitch, might influence model decisions. Future research should explicitly assess gender as a potential confounder or incorporate pitch-related features.

Clinically, these findings highlight the potential of a lightweight machine learning workflow as an adjunct tool for identifying individuals at risk for RHD-related communication impairments. Such automated screening, analyzing brief speech samples, could quickly flag patients who warrant detailed clinical assessments. This technology is intended as supportive decision-making assistance, objectively quantifying acoustic deviations and reinforcing clinical judgment rather than replacing expert evaluations.

V. LIMITATIONS AND FUTURE WORK

Despite these positive outcomes, several limitations must be acknowledged. First, our dataset included a relatively small

number of unique speakers (32), limiting model generalizability and the diversity of captured speech characteristics. Expanding the dataset with additional individuals and more varied speech samples per speaker is essential for broader representation. Second, our analysis mainly focused on paralinguistic features (MFCCs) and did not explicitly capture other potentially informative linguistic or contextual cues relevant to RHD symptomatology. Notably, our preliminary attempts to incorporate pitch contours and features extracted via FastICA were not promising, possibly due to methodological issues or data limitations. Thus, future studies should revisit these prosodic features—such as pitch range, intensity variation, and rhythmic patterns—with refined extraction methods. Lastly, despite careful preprocessing steps (noise reduction, VAD, normalization), residual variability from recording conditions or microphone differences could still affect classification results.

Looking ahead, future research should prioritize dataset expansion by including additional samples from the broader TalkBank RHDBank corpus, covering diverse speakers and varied linguistic tasks. Further exploration of advanced feature extraction methods, particularly revisiting explicit prosodic cues and refining techniques like pitch contours or FastICA, may provide more comprehensive speech representations. Additionally, employing deep learning architectures (e.g., convolutional or recurrent neural networks) could potentially enhance performance given sufficient data. Pairing these models with explainability techniques—such as identifying critical frequency bands or prosodic elements indicative of RHD—will be crucial for clinical application. Pursuing these avenues promises improved diagnostic accuracy, greater clinical utility, and deeper insights into how paralinguistic signals reflect right hemisphere dysfunction.

REFERENCES

[1] Ferré P, Ska B, Lajoie C, Bleau A, Joanette Y. Clinical Focus on Prosodic, Discursive and Pragmatic Treatment for Right Hemisphere Damaged Adults: What's Right? Rehabil Res Pract. 2011;2011:131820. doi: 10.1155/2011/131820. Epub 2011 Feb 16. PMID: 22110970; PMCID: PMC3200269.

[2] S. M. Sheppard, M. D. Stockbridge, L. M. Keator, L. L. Murray, and M. L. Blake, "The Company Prosodic Deficits Keep Following Right Hemisphere Stroke: A Systematic Review," *Journal of the International Neuropsychological Society*, pp. 1–16, Jan. 2022, doi: https://doi.org/10.1017/s1355617721001302.

[3] A. Parola *et al.*, "Assessment of pragmatic impairment in right hemisphere damage," *Journal of Neurolinguistics*, vol. 39, pp. 10–25, Aug. 2016, doi: https://doi.org/10.1016/j.jneuroling.2015.12.003.

[4] C. L. Johns, K. M. Tooley, and M. J. Traxler, "Discourse Impairments Following Right Hemisphere Brain Damage: A Critical Review," *Language and Linguistics Compass*, vol. 2, no. 6, pp. 1038–1062, Nov. 2008, doi: https://doi.org/10.1111/j.1749-818x.2008.00094.x.

[5] M. Thye and D. Mirman, "Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia," *NeuroImage: Clinical*, vol. 20, pp. 1129–1138, 2018, doi: https://doi.org/10.1016/j.nicl.2018.10.017.

[6] B. Vimal, M. Surya, Darshan, V. Sridhar, and A. Ashok, "MFCC Based Audio Classification Using Machine Learning," Jul. 2021, doi: https://doi.org/10.1109/icccnt51525.2021.9579881.

[7] B. Gutiérrez-Serafín, J. Andreu-Perez, H. Pérez-Espinosa, S. Paulmann, and W. Ding, "Toward assessment of human voice biomarkers of brain lesions through explainable deep learning," *Biomedical Signal Processing and Control*, vol. 87, p. 105457, Sep. 2023, doi: https://doi.org/10.1016/j.bspc.2023.105457.

[8] Minga, J., Johnson, M., Blake, M. L., Fromm, D., & MacWhinney, B. (2021). Making sense of right hemisphere discourse using RHDBank. *Topics in Language Disorders, 41*(1), 99-122.

[9] B. McFee *et al.*, "librosa/librosa: 0.11.0," *Zenodo*, Mar. 2025, doi: https://doi.org/10.5281/zenodo.15006942.

[10] T. Sainburg, "timsainb/noisereduce: v1.0," Jun. 2019. https://doi.org/10.5281/zenodo.3243139

[11] John Wiseman, "py-webrtcvad/LICENSE at master · wiseman/py-webrtcvad," *GitHub*, 2016. https://github.com/wiseman/py-webrtcvad/blob/master/LICENSE (accessed Mar. 14, 2025).

[12] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980, doi: https://doi.org/10.1109/tassp.1980.1163420.

[13] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, Available: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

[14] M. Majnik and Z. Bosnić, "ROC analysis of classifiers in machine learning: A survey," *Intelligent Data Analysis*, vol. 17, no. 3, pp. 531–558, May 2013, doi: https://doi.org/10.3233/ida-130592