Multifractal Voice Analysis for Detection of Vocal Pathologies

Team A-10

Aksaya Kanthan K S
CB.AI.U4AID23003

Kiruthik Nandhan Murthi
CB.AI.U4AID23018

Manasa V
CB.AI.U4AID23022

Rithvik Rajesh
CB.AI.U4AID23032

AID A - Team 10

20.01.2026

Outline

Introduction and Motivation

- The human voice is produced through a complex interaction of airflow, vocal fold vibration, and vocal tract movement.
- Voice disorders, especially structural and neurological pathologies, disturb this process and introduce subtle irregularities in speech.
- Traditional acoustic features like pitch, jitter, shimmer, and MFCCs capture only short-term or local changes and often fail to reflect deeper voice irregularities.
- Since speech is a nonlinear and multiscale signal, tools like fractal and multifractal analysis are better suited to describe its complex behavior.
- This project explores fractal dimension and multifractal features to better distinguish healthy voices from different types of voice pathologies.

Problem Definition

- Detecting voice disorders from speech is difficult because changes in the voice are often subtle and complex.
- Common acoustic features work well to separate healthy and unhealthy voices, but they struggle to differentiate between types of disorders, especially structural and neurological ones.
- Structural and neurological voice disorders can sound very similar, which often leads to incorrect classification.
- Most existing methods focus on short-term or linear features and do not fully represent the nonlinear and multiscale nature of pathological speech.
- Therefore, there is a need for a robust and objective approach that can better capture these complex voice patterns and improve multi-class voice pathology classification.

Literature Review

Sl No	Title and year	Journal/ conference Name	Methodology	Key findings	Limitations
1	*Nonlinear Statistical Analysis of Normal and Pathological Infant Cry Signals* (2022)	Entropy (Basel) .	Multifractal Wavelet Leaders on cepstral domain; statistical comparison	Healthy cries show higher multifractal complexity; spectrum width discriminative	Infant cries only; no ML classifier; manual segmentation
2	*Detection of Voice Pathology Using Fractal Dimension in a Multiresolution Analysis (2015)*	Journal of Medical Systems	Wavelet decomposition + Fractal Dimension at multiple scales	Fractal features distinguish pathological voices; multiscale FD improves detection	Not multifractal; sensitive to preprocessing; sustained vowels only
3	*Fish Sound Characterization Using MFDFA* (2020)	Bioacoustics	MFDFA on biological acoustic signals	Strong multifractality observed; spectrum width discriminative	Not human voice; no pathology task
4	Pathological Assessment of Patients' Speech Signals Using Nonlinear Dynamical Analysis (2010)	Computers in Biology and Medicine	Nonlinear dynamical features including fractal dimension, correlation dimension, Lyapunov exponent; SVM classifier	Nonlinear and fractal features significantly outperform classical perturbation measures; up to 94.44% accuracy	Monofractal only (no multifractal spectrum); sustained vowels; older dataset
5	*Wavelet Leader Multifractal Analysis of Period and Amplitude Sequences from Sustained Vowels* (2010)	Speech Communication	Wavelet Leader Multifractal Analysis applied to pitch-period and amplitude sequences	Both amplitude and period fluctuations exhibit multifractal behavior; multifractal features enable discrimination between healthy and pathological voices	Sustained vowels only; limited pathology classes; moderate dataset size
6	*Fractal Features for Automatic Detection of Dysarthria* (2017)	IEEE (Conference Proceedings)	DFA-based fractal scaling exponent, fractal jitter, multivariate DFA on articulatory data; XGBoost classifier	Fractal features significantly improve dysarthria detection; up to 90.2% accuracy and high sensitivity	Monofractal only (single scaling exponent); focuses on dysarthria (ALS); requires specific speech task (DDK)

Literature Review

Sl No	Title and year	Journal/ conference Name	Methodology	Key findings	Limitations
7	*Robust and Complex Approach of Pathological Speech Signal Analysis* (2022)	NeuroComputing	Evaluation of 92 features including nonlinear/complexity measures	Nonlinear features improve accuracy when fused with classical features	Multifractals not deeply explored; interpretability limited
8	*A fractal approach to normal and pathological voices (2000)*	Acta Otolaryngol	Discrete Wavelet Transform–based multiresolution analysis; Fractal Dimension computed at multiple scales; classical ML classifiers.	Fractal dimension features differ significantly between normal and pathological voices; multiresolution fractal analysis improves discrimination compared to single-scale measures.	Uses single fractal dimension (not multifractal); sensitive to wavelet and FD choice; primarily sustained vowels
9	*Modified Multifractal Detrended Fluctuation Analysis* (2019)	Water	Improved MFDFA methodology	More robust multifractal estimation in noisy signals	Method paper; no voice application
10	*Combined Use of Nonlinear Measures for Analyzing Pathological Voices (2024)*	*International Journal of Image and Graphics*	Phase-space–based nonlinear analysis combining singularity spectrum coefficients (α_min, α_max, γ_1, γ_2), correlation dimension (D_2) and correlation entropy (K_2); SVM with Gaussian kernel for classification.	Combined nonlinear features achieve very high discrimination between healthy and pathological voices (reported accuracy up to **97%**); fusion of nonlinear measures improves robustness over single features	Not strictly multifractal (no MFDFA/WTMM); small effective dataset size; sustained voice recordings only; limited discussion of physiological interpretability.
11	*Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection* (2007)	Nature Precedings	Recurrence Period Density Entropy (RPDE) for aperiodicity + Detrended Fluctuation Analysis (DFA) for fractal scaling of breath noise; QDA classifier	Fractal scaling and recurrence features outperform classical perturbation measures; ~92% accuracy on large clinical dataset	Uses monofractal scaling only; sustained vowels; not a multifractal analysis

Gaps Identified

- While fractal and multifractal methods have been explored in voice analysis, most studies focus on either fractal or multifractal features in isolation, rather than studying their combined contribution alongside classical acoustic features.

- Existing works often demonstrate high accuracy, but provide limited analysis of how much each feature group (acoustic, fractal, multifractal) contributes to performance improvement.

- Many multifractal-based studies focus on binary classification or a single pathology type, with limited emphasis on multi-class differentiation between structural and neurological disorders.

- Although datasets like SVD are widely used, systematic comparisons across baseline, fractal, and multifractal feature sets on the same dataset are still limited.

Objectives

- To analyze voice signals and identify patterns that distinguish healthy, structural, and neurological voice disorders.
- To build a strong baseline system using traditional acoustic features such as pitch, jitter, shimmer, HNR, and MFCCs.
- To extract fractal dimension features that capture the overall complexity of voice signals across different time scales.
- To explore multifractal characteristics of speech using MFDFA and study their contribution to classification performance.
- To compare the effectiveness of classical, fractal, and multifractal feature sets for multi-class voice pathology classification.

Methodology

- Dataset Preparation
  - Speech recordings are taken from the Saarbrücken Voice Database (SVD).
  - Voices are grouped into three major classes:
    - Healthy – speakers without diagnosed voice disorders
    - Structural disorders – conditions affecting the physical structure of the vocal folds
    - Neurological disorders – conditions affecting neural control of voice production
  - Sustained vowel /a/ recordings are selected for uniform analysis.
  - All signals are resampled to 16 kHz, converted to mono, and amplitude-normalized.

- Baseline Feature Extraction
  - Classical acoustic features are extracted to establish a strong baseline:
    - Fundamental frequency (F0)
    - Jitter and shimmer
    - Harmonics-to-Noise Ratio (HNR)
    - MFCCs (mean and standard deviation)
  - These features capture short-term and perceptual properties of speech.

Methodology

- Fractal Dimension (FD) Analysis
  - Higuchi Fractal Dimension is computed on the time-domain signal to measure overall waveform complexity.
  - Wavelet-based multiresolution FD is performed using a 3-level Discrete Wavelet Transform to capture complexity across frequency bands.
  - FD features describe long-range and scale-dependent irregularities in voice signals.

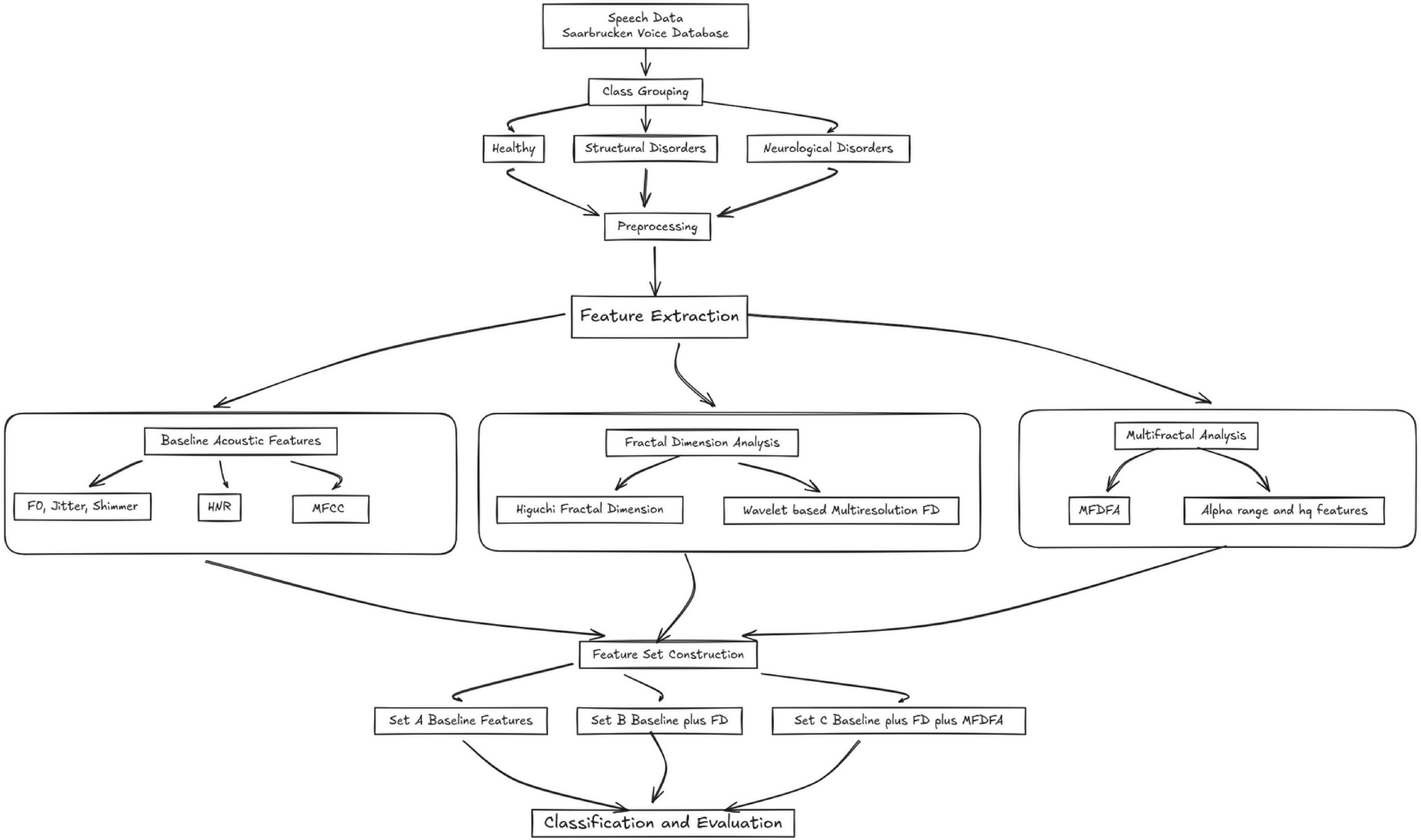
- Multifractal Detrended Fluctuation Analysis (MFDFA)
  - MFDFA is applied to the steady vowel segment to analyze scale-dependent fluctuations.
  - Multiple moment orders q are used to capture different fluctuation strengths.
  - Compact multifractal features are extracted, including:
    - Singularity spectrum width ($\Delta\alpha$)
    - Generalized Hurst exponents h(q) at selected qqq values
  - These features represent heterogeneity and nonlinearity in voice dynamics.

Methodology

- Feature Set Construction
  - Set A: Baseline acoustic features
  - Set B: Baseline + Fractal Dimension features
  - Set C: Baseline + Fractal Dimension + Multifractal (MFDFA) features

- Classification and Evaluation
  - The extracted feature sets will later be used to train multiple machine learning models.
  - Different feature combinations will be tested to study their impact on classification.
  - Performance will be evaluated using standard metrics such as: Accuracy, Precision, Recall, F1-score,Confusion matrices
  - The goal is to analyze the incremental contribution of fractal and multifractal features.

Methodology



Dataset Description

- Voice recordings are taken from the Saarbrücken Voice Database (SVD), a clinically validated dataset widely used in voice pathology research.
- The data is organized into three categories:
  - Healthy voices:
    - Speakers without any diagnosed voice disorder
    - 687 sustained vowel /a/ recordings
  - Structural voice disorders
    - Disorders affecting the physical structure of the vocal folds
    - Included pathologies: Phonationsknötchen (vocal nodules), Stimmlippenpolyp (vocal polyps), Reinke Ödem
    - 125 sustained vowel /a/ recordings
  - Neurological voice disorders
    - Disorders affecting neural control of voice production
    - Included pathologies: Rekurrensparese (recurrent laryngeal nerve paralysis), Morbus Parkinson, Spasmodische Dysphonie
    - 270 sustained vowel /a/ recordings
- The final dataset consists of 1082 speech samples.

Preliminary Results and Discussion

- Baseline acoustic features (F0, jitter, shimmer, HNR) were extracted for all selected voice samples.
- Descriptive statistics show systematic differences between healthy, structural, and neurological voices:
  - Healthy voices exhibit higher average F0 and lower jitter, indicating more stable phonation.
  - Pathological voices show higher jitter and lower HNR, reflecting increased vocal irregularity and noise.
- Higuchi Fractal Dimension (FD) analysis reveals consistent trends in signal complexity:
  - Healthy voices show lower average FD values and smaller variability.
  - Structural and neurological voices show higher FD values and wider spread, indicating increased irregularity.
- Neurological voices exhibit greater FD variability, consistent with unstable neural control of voice production.
- These observations suggest that fractal features capture complexity information not fully represented by classical acoustic measures.

Preliminary Results and Discussion

- Multifractal Feature Observations
  - Multifractal analysis using MFDFA has been implemented for all voice samples.
  - The singularity spectrum width ($\Delta\alpha$) shows noticeable differences across voice categories:
  - Healthy voices generally exhibit narrower spectra, indicating more uniform signal dynamics.
  - Structural and neurological voices show wider spectra, reflecting higher heterogeneity and irregular fluctuations.
- Generalized Hurst exponents h(q)h(q)h(q) also show class-dependent trends, suggesting differences in fluctuation behavior at multiple scales.
- These results indicate that pathological voices exhibit stronger multifractality compared to healthy voices.

Interim Summary/Conclusion

- Data preparation, preprocessing, and feature extraction up to multifractal analysis have been completed.

- Baseline, fractal, and multifractal features exhibit consistent and interpretable patterns across voice categories.

- Fractal and multifractal features capture different aspects of voice signal behavior compared to traditional acoustic features.

Completion plan for the next phase

- Experiment with different feature combinations, including:
    - Baseline features alone
    - Fractal features alone
    - Multifractal features alone
    - Combined feature sets

- Train and evaluate multiple machine learning models to study how different features influence classification performance.

- Analyze class-wise performance, with special focus on distinguishing structural and neurological voice disorders.

- Conduct comparative analysis to understand the added value of fractal and multifractal features over classical acoustic features.

References

- H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, 2004.
- T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," Physica D: Nonlinear Phenomena, vol. 31, no. 2, pp. 277–283, 1988.
- J. W. Kantelhardt et al., "Multifractal detrended fluctuation analysis of nonstationary time series," Physica A: Statistical Mechanics and its Applications, vol. 316, no. 1–4, pp. 87–114, 2002.
- D. A. Torres, L. A. Da Silva, and M. A. Batista, "Detection of voice pathology using fractal dimension in a multiresolution analysis," Journal of Medical Systems, vol. 39, no. 12, pp. 1–9, 2015.
- P. Abásolo et al., "Nonlinear analysis of voice signals: Application to pathological voices," Computers in Biology and Medicine, vol. 40, no. 2, pp. 132–139, 2010.

Thank you