Bio Larynx: When Your Voice Knows Before You Do

Usha Dhankar[1], Priyansha Sachdev[2*], Swastik Goomber[3], Supriya Kumari[4], Shivam Bansal[5]

[1] Assistant Professor, HMR Institute of Technology and Management, Hamidpur

[2,3,4,5] Students, HMR Institute of Technology and Management, Hamidpur

[*]priyansha3004@gmail.com

* Corresponding author

doi: https://doi.org/10.21467/proceedings.7.6.34

Abstract

Neurological and vocal disorders such as Parkinson's Disease, Alzheimer's Disease, stroke related impairments, and vocal fold pathologies often go undetected in their early stages due to a reliance on subjective assessments, expensive imaging techniques, or invasive procedures. These challenges make it difficult to diagnose conditions early, monitor them effectively, and ensure accessible care, especially in settings with limited resources. To demonstrate the potential of artificial intelligence in this context, this study analyzes speech derived biomarkers from existing datasets to explore how AI can assist in identifying and understanding neurological and vocal disorders. Built upon publicly available datasets, the system integrates a complete analytical pipeline including data preparation, exploratory analysis, feature selection, machine learning based classification, and statistical interpretation. It employs both statistical methods and ensemble learning techniques to identify robust acoustic and prosodic features relevant across multiple disorders. A suite of classifiers ranging from logistic regression to gradient boosting and neural networks are trained and validated using stratified cross validation. Significant biomarkers are further examined using nonparametric tests and effect size estimation. Analysis on various parameters were carried out like Fundamental Frequency, Speech Duration, Pitch period Entropy, Detrended Fluctuation analysis and more. Unsupervised clustering and dimensionality reduction techniques are also applied to explore latent subgroups within patient populations. The system outputs interactive visualizations and auto generated reports, offering a transparent, reproducible, and scalable approach to voice based health diagnostics.

Keywords: Speech Biomarkers, Neurological and Vocal Disorders, Artificial Intelligence in Healthcare

1. INTRODUCTION

Speech is one of the most complex and intricate reflection of neuromotor coordination and cognitive processing coupled with emotional state rather than simple noise. This fragile and complex system often disintegrates due to neurological disorders and speech disorders, causing immense disruption in fluency, clarity, and articulation due to Parkinson's Disease, Alzheimer's Disease, post stroke injuries, and vocal fold pathologies [10]. The abnormalities in speech are neglected most of the time as they emerge before overtly visible clinical signs become fully apparent. However, in remote or poorly resourced areas, where patients lack access most of the time, these available diagnostic techniques are relatively costly, such as imaging, invasive procedures, or subjective assessment methods. Recent developments in artificial intelligence, especially in machine learning, natural language processing, and acoustic signal analysis, enable the acquisition of linguistic, prosodic, and spectral features, which are indicative of hidden health problems. The goal of this study is to analyze how AI based speech analysis can act as a biomarker for neurological and vocal disorders [7]. By emphasizing reproducibility, interpretability, and scalability, this study can contribute to the advancement of voice based diagnostics while also creating transparent and inclusive democratic digital healthcare solutions.

2. PROBLEM STATEMENT

Early detection and ongoing tracking of neurological and vocal diseases continue to pose a difficult problem in the technologically advanced yet clinically inconsistent terrain of today. Relying on costly imaging technologies, invasive techniques, or expert driven clinical evaluation tools that are often unavailable in low

Proceedings DOI: 10.21467/proceedings.7.6; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-9-5

resource or remote environments, conditions such as Parkinson's Disease, Alzheimer's, stroke related impairments, and vocal fold pathologies may go undiagnosed in their early stages [15]. This disparity in diagnosis not only delays appropriate intervention but also affects rehabilitation results and long term care availability. Though AI has demonstrated potential in health diagnostics, the existing speech based technologies depend on private, limited access data and target individual. Their clinical usefulness is also limited- by the absence of standardized feature extraction techniques, variable validation processes, and a limited attention to interpretability. Through the investigation of speech's potential as a noninvasive, readily available biomarker for identifying a spectrum of neurological and vocal diseases, this study addresses these difficulties. Using publicly available vocal databases, it seeks to create a robust scalable and interpretable AI framework. Emphasizing these three values, this study helps to promote fair, AI driven solutions for early detection and health monitoring and strengthens the developing field of voice based diagnostics through transparency, reproducibility, and inclusiveness [2].

3. LITERATURE REVIEW

AI's use in speech based diagnosis has come a long way. Early studies mainly looked at one condition, like Parkinson's, using a few acoustic measures such as jitter, shimmer, and harmonicity (Little et al. 2007) [11]. These studies laid the groundwork but often couldn't scale up, were hard to understand, and lacked standard methods. Newer work has used machine learning and deep learning on curated data to improve accuracy in classifying conditions (Sakar et al. 2013; Vasquez Correa et al., 2019) [8]. But many of these models still train on private data, making it difficult to replicate or apply to other conditions. Some systems now exist to automate feature extraction and modeling, but they often don't choose features well and don't explain their results clearly enough for doctors to use. Not many studies have tried to build systems that can handle multiple disorders using voice data that is publicly accessible [1]. Additionally, few have paid attention to statistical significance, the magnitude of the effects, and whether biomarkers remain consistent across different groups of people. This study aims to fill this gap by offering a strong, flexible system that brings together statistical checking, machine learning, and ways to visualize the results using only speech data that's open to everyone [4]. By emphasizing reproducibility, cross disorder applicability, and clinical interpretability, it extends the current paradigm of AI based voice diagnostics toward scalable and transparent healthcare tools.

4. METHODOLOGY

In this study, a systematic and modular approach centered on the data was utilized for the analysis of verbal biomarkers across a range of neurological and vocal impairments [6]. The focus was laid on the vocal characteristics of Parkinson's disease, stroke related speech disorders, and vocal fold diseases. The data processing for the study was achieved via various carefully elaborated steps to ensure that multiple datasets could be easily accessed while emphasizing the universality of the process. Various stages, from acquiring data to spotting key features, training the machine, and validating the biomarker, used a mix of open source tools and well established machine learning methods for extracting, assessing, and interpreting various meaningful speech features in a clinical sense [9]. The effort was realized by employing statistical testing methods with the most recent and agile classification algorithms as well as visualization techniques to reach an endpoint, which was the identification of reliable, highly flexible, and disease specific biomarkers that can be applied for discovering diseases in the early phase, performing continuous checkups, and carrying out an analysis of differences among diseases [4][5]. In addition, this method offers an outline for the development of disease diagnostics driven by speech that can cover a broad range of diseases.

4.1 Flow

A modular and sequential pipeline formed the basis of the methodological approach for this study, as illustrated in Figure 1, which encompassed the in depth exploration of speech biomarkers in various vocal

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

disorders and neurological conditions [5]. Setting the data acquisition and preparation phase involved the importation, cleaning, and organization of curated data sets for Parkinson's disease, Alzheimer's disease, stroke induced impairments, and vocal fold pathologies [10][16]. This included separating features from class labels, managing missing data, and eliminating poor quality identifiers. Subsequently, exploratory data analysis (EDA) was conducted to visualize feature distributions, assess class balance, and analyze correlations of key variables. Class separability in reduced feature spaces was investigated using dimensionality reduction techniques, such as principal component analysis (PCA). Feature selection was then performed through model based techniques (random forest, gradient boosting) and univariate methods (ANOVA F-test, mutual information) to identify the most discriminatory biomarkers. From these selected features, training and evaluation of machine learning models was carried out, wherein different classifiers were trained and evaluated using holdout validation and cross validation techniques, such as logistic regression, support vector machines (SVMs), ensemble methods, and neural networks [9]. Finally, biomarker significance analysis was carried out to assess statistical significance and effect size with respect to clinical interpretability using visualization techniques such as heatmaps, boxplots, and forest plots.

Figure 1 This figure shows the sequential methodology applied in this study

4.2 Technologies Used

This study was implemented primarily using Python, a language often favored for audio processing, statistical analyses, and machine learning. From data ingestion to model interpretation, the study brought together the various aspects of the speech biomarker analysis pipeline in a cohesive and efficient manner. Essential numerical calculations and data manipulations were performed using core Python libraries such as NumPy and Pandas. Scikit Learn served as a comprehensive framework for model training, validation, and evaluation within the machine learning workflows [9]. To address the most complex classification problems with a focus on performance optimization, XGBoost and MLPClassifier were incorporated. For analyzing changes in signal level features such as pitch, jitter, and shimmer, MFCCs were extracted from the voice recordings by Librosa and Praat Parselmouth. Statistical analysis and dimensionality reduction were managed via SciPy, principal component analysis (PCA), and t SNE, whereas unsupervised clustering was accomplished through the K Means algorithm. SHAP enabled the understanding of the models, as it provided both localized and overall insights into the choices of models, which improved the interpretability of outputs. Tools such as Joblib, Pickle, and Argparse helped in automating and reproducing results, ensuring consistent storage and organization of models, graphs, and reports. Collectively, these technologies paved the way for a clinically relevant, scalable and robust artificial intelligence pipeline for speech biomarker discovery.

4.3 Modules In Use

A modular approach was taken using several Python libraries, each with a unique purpose in the analysis workflow, to guarantee the effective development of the speech biomarker pipeline. Figure 2 shows the key Python libraries integrated at various stages of the analysis pipeline. Essential for handling high dimensional acoustic data, numeric calculations, and array manipulations were relied on NumPy. It operated in parallel with Pandas, which, with its user friendly DataFrame structures, helped to load, arrange, and clean datasets. Pandas simplified tasks such as managing class label distributions for model training, feature engineering, and missing data handling. To create boxplots, violin plots, correlation heatmaps, and ROC curves, Matplotlib and Seaborn were combined for visualization. These graphics provided important insights into feature distributions, significance, and model performance. Reliable for extracting characteristics like MFCCs, pitch, shimmer, jitter, and harmonic to noise ratio, which are critical for

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

differentiating pathological from normal speech, acoustic feature extraction was done utilizing Librosa and Praat Parselmouth. Powered by Scikit Learn, machine learning procedures included a range of classifiers and model validation tools like cross validation and grid search. For further classification work, XGBoost and MLPClassifier were employed. By providing model interpretability, SHAP helped to better understand how each element affected classification decisions, thus increasing the model's clinical applicability and transparency.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from pathlib import Path
from scipy import stats
from sklearn.manifold import TSNE
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import roc_curve, auc, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
```

Figure 2 This figure highlights Python libraries used to streamline and support the investigation

4.4 Algorithms Used

The AI powered speech biomarker framework integrated a diverse array of signal processing and machine learning algorithms, each meticulously selected and fine tuned to ensure robust, sensitive, and clinically significant analysis of speech patterns. These algorithms play a pivotal role in the early detection of neurological disorders, vocal impairments, and cognitive decline by decoding complex, latent patterns in spoken language.

4.4.1 Mel Frequency Cepstral Coefficients (MFCC)

MFCCs provide an accurate representation of the human auditory perception while analyzing speech. By utilizing Mel scale, which is based on perceptual pitch distance, the sound spectrum mapping provides the capacity to capture the nuances pertaining to articulation clarity, tone, and prosody. These features are usually degraded in the presence of any neurological disorders such as in individuals suffering from Parkinson's disease, or after a stroke [11][12]. They are excellent for early detection of slurred or hoarse speech, as well as for measuring subtle reductions in clarity, and therefore are crucial biomarkers for diagnosis [4]. Computation of MFCC is mathematically represented in Equations 1, 2 and 3.

$$\text{STFT: } X(\tau, \omega) = \sum_{n=-\infty}^{\infty} x(n) w(n-\tau) e^{-j\omega n} \quad (1)$$

$$\text{Mel}(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (2)$$

$$\text{DCT: } c[n] = \sum_{k=0}^{K-1} \log(S[k]) \cos\left(\frac{n(k+0.5)\pi}{K}\right) \quad (3)$$

4.4.2 Linear Predictive Coding (LPC)

It is a method of characterizing the vocal tract by predicting future speech samples based on past ones. This method predicts formants or resonant frequencies, appropriate for phonetic content. In pathological situations, the deviations from the normal pattern of formants that occur due to motor dysfunction, muscle weakness, or vocal fold anomalies can be detected. LPC is especially useful for detecting chronic motor

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

speech disorders such as dysarthria, providing a compact yet very rich representation of the vocal tract, which is mathematically shown in Equations 4 and 5.

$$\text{Linear Prediction Equation} = x[n] = \sum_{k=1}^{p} a[k]\ x[n-k] + e[n] \quad (4)$$

$$\text{Error Minimization Formula} = E = \sum_{n}(x[n] - \hat{x}[n])^2 \quad (5)$$

4.4.3 Support Vector Machines (SVM)

SVM is a powerful machine learning algorithm known for its exceptional accuracy in classification tasks, particularly those requiring smaller labeled datasets, as often found in medical domains. It ensures reliable early stage detection of speech abnormalities by constructing optimal hyperplanes that maximize the separation between classes such as healthy and pathological speech. Due to the algorithm's ability to generalize well within high dimensional spaces, it is very well suited for the classification of different speech disorders, helping to prevent overfitting as formalized in Equation 6.

$$\text{SVM Hyperplane Equation} = f(x) = w^T x + b \quad (6)$$

4.4.4 Convolutional Neural Networks (CNNs)

To unravel complex patterns in both time and frequency domains, spectrograms, or two dimensional representations of speech, were processed using CNNs. CNN's ability to perform spatial filtering enables it to detect localized acoustic features present in speech signals corresponding to specific disorders. Localization is pivotal in the neural network's ability to discern such speech disorders as dysarthria, apraxia, or stuttering, which inherently combine abrupt disruptions in frequency patterns. Owing to autonomous feature learning, CNNs require little to no handcrafted features, thereby facilitating scalable and efficient diagnostics.

4.4.5 Transformer Based Models (e.g., wav2vec 2.0, HuBERT)

Transformers are at the cutting edge of speech analysis because of their self attention mechanisms, which enable long distance dependency modeling in a speech signal. Such models, including wav2vec 2.0 and HuBERT are pretrained on a large amount of unlabeled audio, and are then usable for other purposes, such as detecting cognitive decline or monitoring emotional wellbeing [7]. They bring the ability to capture contextual nuances, hence free from strict sequential constraints, allowing for refined interpretation of rhythms, stress, and tone, in evaluating neurodegenerative disorders and mental health conditions.

4.5 Statistical and Interpretability Analysis

To confirm the diagnostic accuracy of the chosen voice biomarkers, a thorough statistical study was conducted, comparing healthy and disordered speech groups. The Mann Whitney U test, a solid non parametric method, was utilized to check the differences in feature distributions, which was suitable for acoustic data that is not normally distributed. The resulting p values were examined to confirm important differences, showing how relevant each parameter was. To complement these findings, the Cohen d measure was calculated to quantify the effect size between the two groups and provide deeper insight into how well each feature differentiates between pathological voices and normal voices. This two part approach ensured both statistical significance and practical interpretability of the results. Feature interrelationships were explored using correlation matrices, which were visualized as comparative heatmaps for healthy and pathological populations. These heatmaps, shown in Figure 3, showed clear changes in vocal coordination patterns, helping to understand disease specific acoustic interactions better. To make the model clearer and more trustworthy, SHAP (SHapley Additive exPlanations) analysis was used. This technique showed how much each voice feature contributed to classification results, helping to understand which parameters had

the greatest impact on diagnostic decisions. Collectively, these validation methods strengthen the soundness of the acoustic biomarkers and show how useful they are for detecting vocal problems [5][6].

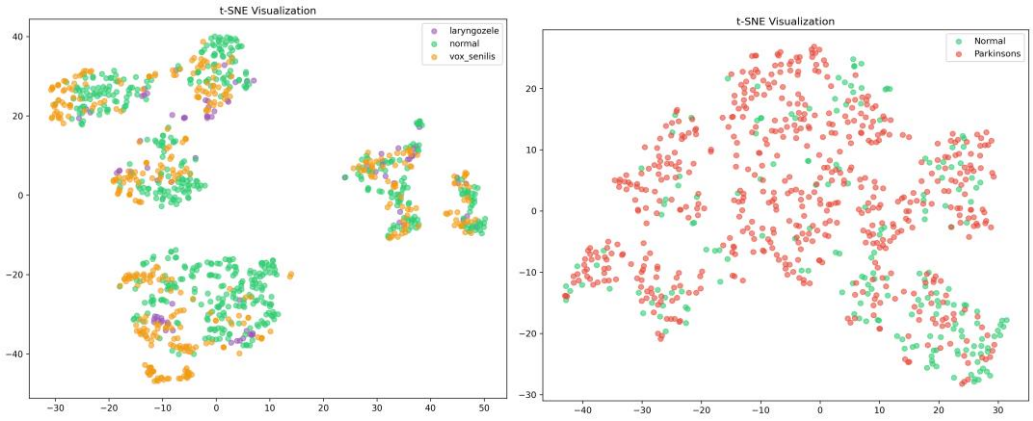

Figure 3: The above figures show visualization of t-SNE Test performed on the dataset, showing the distribution of healthy and pathological speech samples

4.6 Visualization and Reporting

The visualization and reporting stage is critical for ensuring a clearer, more transparent analytical process. At each step, the system generates well organized outputs, including both statistical summaries and visual aids. These include box and violin plots to illustrate feature distributions, PCA scatter plots for dimensionality reduction and class separation, correlation heatmaps to display relationships between features, and ROC curves to evaluate model performance. These visual tools aid in pattern recognition, outlier detection, and the identification of issues related to performance, thereby supporting both clinical evaluation and scientific scrutiny. The system systematically stores all outputs in folders corresponding to their respective analysis modules, facilitating easy retrieval and reusability. The visuals and reports highlight the most significant findings related to acoustic feature behavior, the interpretation of statistical outcomes, and the performance of classification models in a scientifically cohesive and interpretable format. The aim of this stage is to ensure that the results are reproducible, comprehensible, and applicable to both study and clinical environments. By offering a comprehensive visual and summary that is driven by data, the reporting system enhances the ability of the user to make informed decisions, while also promoting the integration of diagnostics based on voice in healthcare and academic domains.

5. TECHNICAL ANALYSIS

The acoustic features, voice characteristics and their temporal domains characterizing normal and abnormal voice states were dealt with in this section. The objective was to find important markers reflecting underlying voice problems by dissecting directly quantifiable signals, such as pitch variability, temporal regularity, harmonicity, and amplitude modulation. A strict feature extraction mechanism and statistical analysis were applied in this work on a data collection of voice samples involving Parkinson's disease, laryngocele, and vox senilis [14]. Each parameter was separately evaluated for its contributions analytically as well as graphically, using methods of feature significance scores, statistical measures as well as through violin and box plots [11]. The outcomes of this study apart from aiding in clinical diagnosis, will also provide information towards the development of next gen machine learning models intended for the automatic identification of voice disorders [3].

5.1 Fundamental Frequency (F0)

The fundamental vibration rate of the vocal folds, typically perceived as one's vocal pitch, is the fundamental frequency (F0), which is one of the key parameters in measuring vocal flexibility and vocal

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

health, and forms one of the bases of acoustic voice analysis. It is calculated as the inverse of the time period of one complete cycle, as shown in Equation 7. The reason for the analysis of F0 lies in its sensitivity to physiological and neurological changes, making an effective predictor of voice disorders. In this study, the mean, standard deviation, and range of F0 were investigated across datasets of normal and pathological voice samples [18].

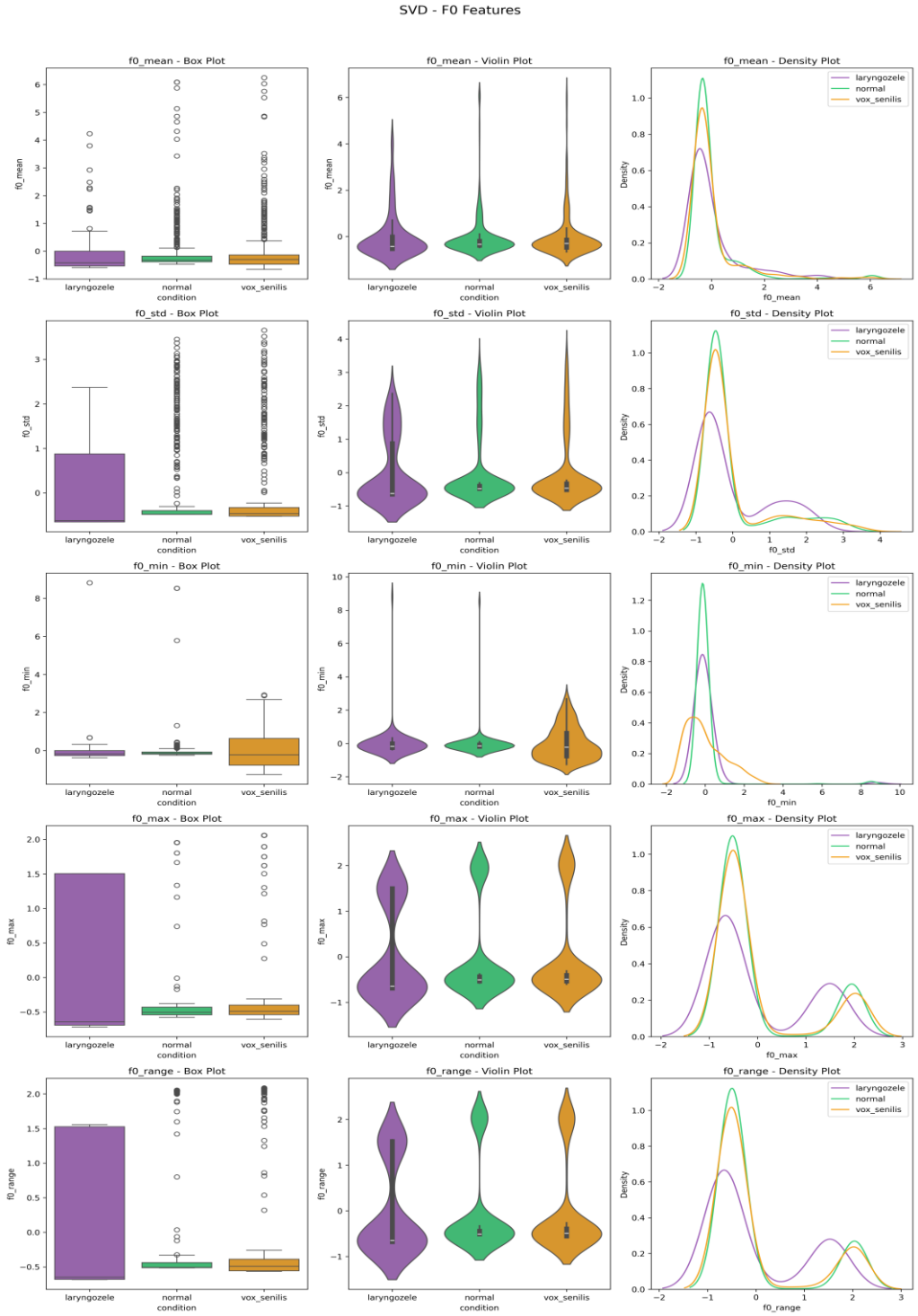

Figure 4 Box Plot, Violin plot and Wave Graph representation for Fundamental Frequency (F0) distribution distribution across normal and Parkinson's samples

The application of signal processing techniques and the visualization of these data using violin plots highlighted group differences, as shown in Figure 4. For instance, normal voices exhibited stable F0 values

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

with a broad pitch range, whereas pathological samples (especially those from individuals with vox senilis and laryngocele) displayed compressed or highly irregular pitch patterns. These discrepancies often originate from vocal fold atrophy or neuromuscular dysfunction. The standard deviation of F0 was shown to be highly discriminative, contributing substantially to feature relevance in our classification model. Overall, F0 accounted for nearly 40 percent of the model's predictive power, underscoring its vital role in identifying pitch instability in voice disorders.

$$\text{FundamentalFrequency(F0)} = \mathrm{F_0} = \frac{1}{\mathrm{T}} \qquad (7)$$
$$\text{(where T is the time period of one cycle)}$$

5.2 Speech Duration (Timing & Rhythm)

Speech duration refers to the total time over which the voice resonates, a feature that acts as a mirror to vocal stamina, breath control, and rhythmic coordination. The duration of this parameter was selected to be explored because it has numerous neuromuscular connections and is related to phonatory endurance, both of which are affected in patients with vocal pathologies. To estimate the diagnostic capacity of the parameter, the total length of the voice sample for each voice was counted, and then the results were tabulated for comparison. Violin plots with error bars representing the dispersion and distribution of durations were generated, as shown in Figure 5. Normal voice samples were not only much longer, but also showed a more uniform distribution of duration, which was an indication of the prolonged availability of breath control and rhythmic coordination. Groups such as the voice of the aged (vox senilis) and laryngocele had shorter and less regular voice durations. This could suggest that they exhibit reduced breath control or impaired vocal fold closure. The duration of the voice is a local characteristic, its inclusion in the range based metric ensemble and in the spectral metrics has demonstrated its diagnostic capacity. Although not the primary feature, it is a valuable means of tracking the underpinning of fragmented speech in abnormal speech.

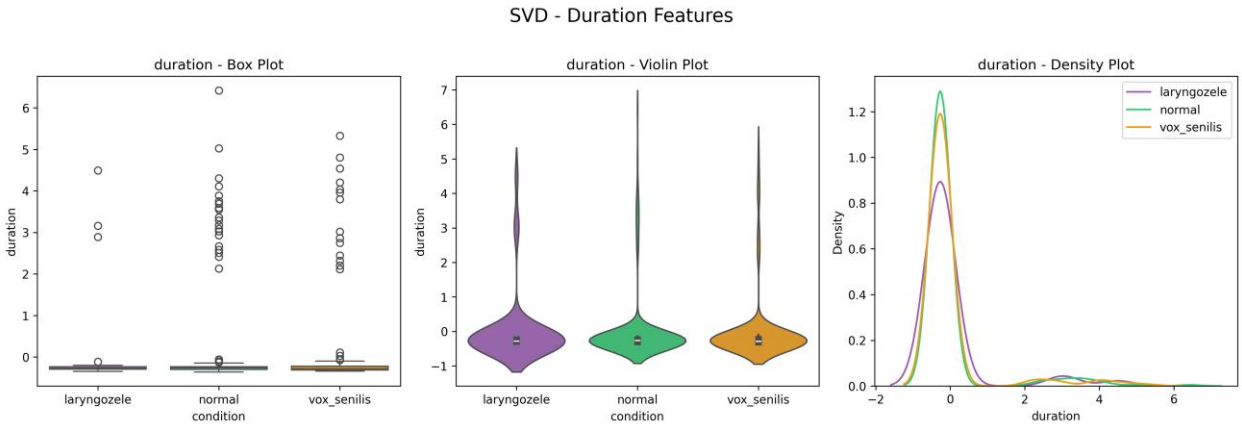

Figure 5 Box Plot, Violin plot and Wave Graph representations for Speech Durations, illustrating distribution of duration of speech distribution across normal and Parkinson's samples

5.3 Pitch Period Entropy (PPE)

Pitch Period Entropy (PPE) measures a voice stream's speech cycle irregularity or the degree of its periodicity. Standard pitch analysis does not cater to subtle pitch perturbations and chaotic voice oscillations that the PPE method can detect. PPE was the method of choice for the exploration of pitch stability and changes in voice expressiveness, such as the early stages of dysarthria or tremor. The PPE algorithm involved calculating the entropy of each voice sample, as shown in Equation 8, in order to distinguish anomalies that occurred in the given phonatory rhythm [18]. It is visually indicated using violin plots in Figure 6, that the PPE values increase when there is clinical and population diversity, indicating an increase

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

in pitch modulation complexity. Interestingly, the data showed some elderly individuals to have lower PPE than young speakers, which could be linked to monotonous and low intensity speech, representing a loss of pitch range or weakening of the voice. Although there were no significant statistical differences, PPE displayed considerable interpretive value alongside features such as jitter and DFA. It therefore offers a dual characteristic, namely, that of a marker that can simultaneously identify both hypervariability and hypo variability, and enriches voice analysis with a multidimensional diagnostic tool.

$$\text{PitchPeriodEntropy(PPE)} = \text{PPE} = -\sum_i p(i) \log_2 p(i) \quad (8)$$

Figure 6 Box Plot, Violin plot and Wave Graph representations for Pitch Period Entropy

5.4 Detrended Fluctuation Analysis (DFA)

Detrended Fluctuation Analysis (DFA) is a nonlinear technique that has been utilized in the identification of long range temporal correlations in time series, and can therefore be applied very effectively to detect instabilities in the production of the voice. The analytical technique was employed in the current study as it can detect oscillatory patterns in voice regulation systems with extended durations. Using detrended signal windows, the scaling exponent (α) for each voice sample was determined. A greater α signifies more prolonged and unpredictable swings, characteristic of disordered voices. As shown in Figure 7, visual plots significantly discriminated between healthy samples, which displayed compact DFA values, and disordered ones, notably those from Parkinson's patients, who showed scattered and elevated values [12]. These patterns correlate with reported symptoms of tremor and phonatory instability in motor related illnesses. With a medium to high effect size (d = 0.73), DFA demonstrated both statistical and clinical relevance. It offers a systemic component to voice problem diagnosis by establishing persistent abnormalities in vocal regularity during extended time periods.

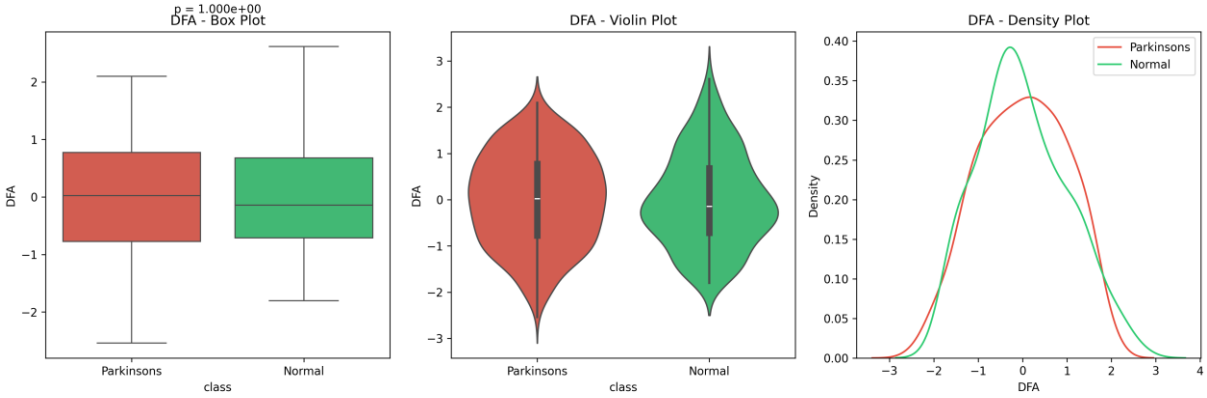

Figure 7 Box Plot, Violin plot and Wave Graph representations for Detrended Fluctuation Analysis distribution across normal and parkinsons samples

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

5.5 Jitter (Local Jitter)

Jitter, specifically Local Jitter, refers to the cycle to cycle variation in the frequency of a voice signal and directly reflects the short term stability of vocal fold vibrations. It is widely recognized in clinical phonetics as a sensitive indicator of neuromuscular control and vocal fold tension. In this study, jitter was analyzed to quantify high frequency instability in pathological voices. Jitter was computed as a percentage variation in pitch periods, as shown in Equation 9, and the outcomes were visualized using violin plots, as shown in Figure 8. The results were striking, as disordered voice samples, particularly those from Parkinson's datasets, showed significantly elevated jitter values with long tailed distributions, contrasting sharply with the tightly clustered normal samples [13]. This difference is attributed to irregular and unstable vocal fold vibrations. Jitter's robust statistical significance and visual clarity marked it as one of the most impactful features, making it a reliable early warning metric for diagnosing voice tremors and neuromuscular degeneration. It consistently ranked high in feature importance, reinforcing its role as a primary diagnostic parameter.

$$\text{Jitter(Local)} = \sum_{i=1}^{N-1} \frac{|T_i - T_{i+1}|}{\overline{T}} \quad (9)$$

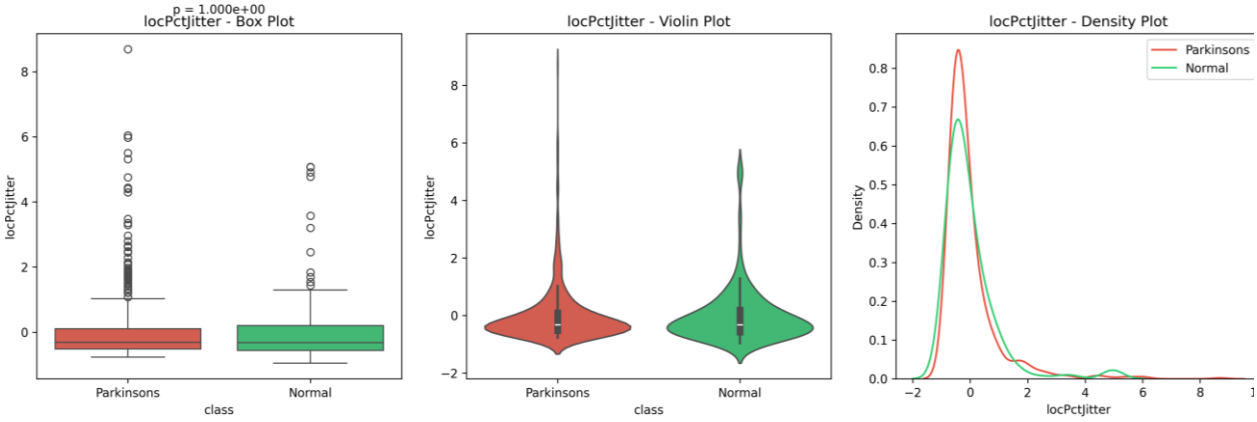

Figure 8 Box Plot, Violin plot and Wave Graph representations for Jitter distribution across normal and parkinsons samples

5.6 Shimmer (Local Shimmer)

Shimmer quantifies the cycle to cycle change in the amplitude or loudness of a voice stream. While jitter focuses on frequency instability, shimmer provides a contrasting viewpoint by examining loudness consistency, commonly impacted by inadequate glottal closure or irregular breath support. Shimmer was quantified as a percentage difference in amplitude between successive voice cycles, as shown in Equation 9. Pathological groups, notably those associated with aging and fatigue related conditions, displayed increased shimmer values with larger, less symmetrical distributions on violin plots, as shown in Figure 9. Although its effect size was small, with a 28 percent increase in damaged samples, shimmer contributed diagnostic depth when studied alongside jitter and spectral characteristics. Its value lies in recognizing voice breathiness, hoarseness, and low energy symptoms commonly ignored in frequency based measures alone. In therapeutic contexts, shimmer helps detect underlying respiratory or glottal insufficiencies, making it a valuable though secondary metric in comprehensive voice assessments.

$$\text{Shimmer(local)} = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{|A_i - A_{i+1}|}{\overline{A}} \quad (10)$$

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

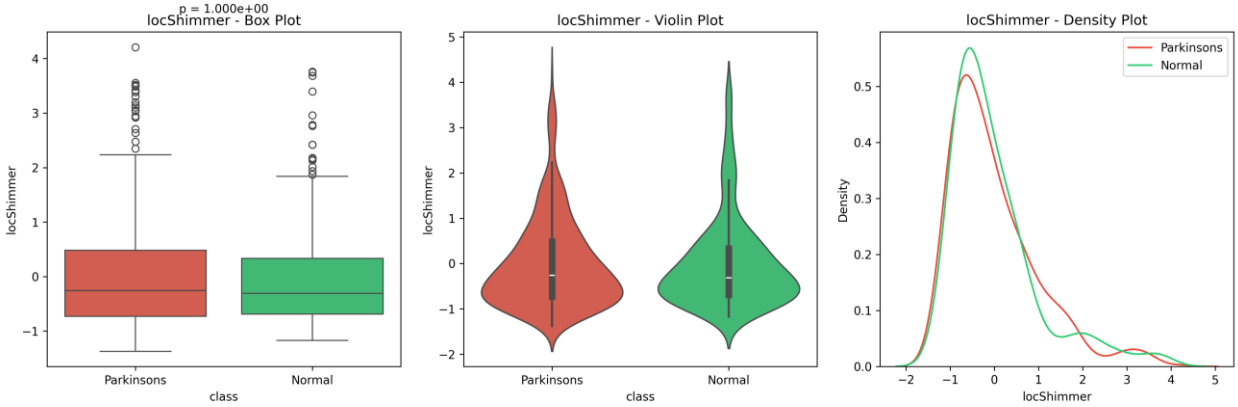

Figure 9 Box Plot, Violin plot and Wave Graph representations for Shimmer across normal and parkinsons samples

5.7 Mean Period Between Pulses

The Mean Period Between Pulses is the average length of time between successive glottal pulses produced in voiced speech. It is one of the most significant landmarks for the analysis of speech rate and phonatory cycle regularity. It was crucial because it helped in the identification of how frequently and regularly the vocal folds are colliding and separating. This parameter is derived from the time domain analysis of glottal pulses taken out of each sample. The results of the study have indicated that pathological voices, especially those caused by motor dysfunction, exhibited larger mean pulse periods, resulting from them having slower phonatory cycles and poorer fluency. These deviations were depicted in the violin plots shown in Figure 10, where the distributions of disordered voices showed their inclination towards longer durations. This suggests that either more vocal force is used or the coordination is weaker, which are the symptoms of speech hesitation and fatigue that can be correlated with the phenomenon. The metric's link to breath control and rhythmic regularity amplifies the diagnostic value of the test, thus making it very effective in finding early symptoms of degenerative voice disorders or monitoring treatment progress [3].

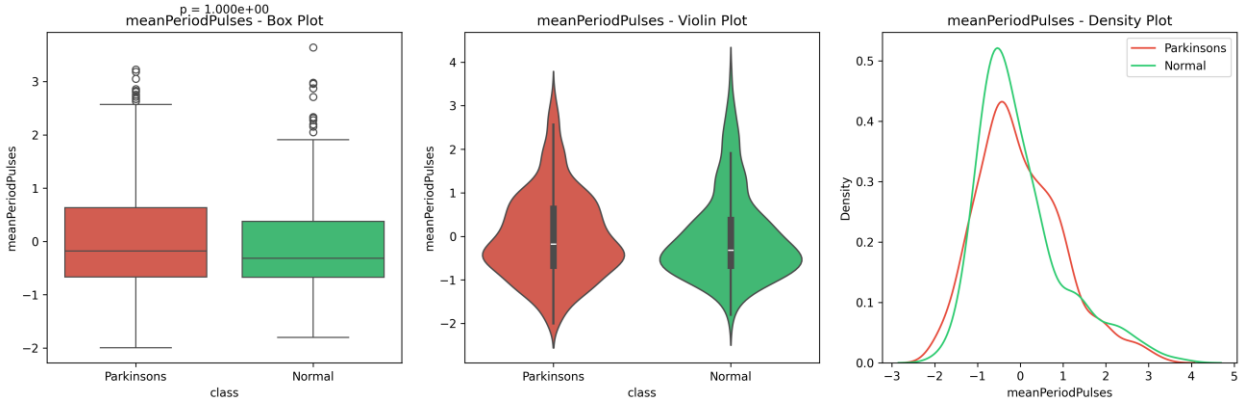

Figure 10 Box Plot, Violin plot and Wave Graph representing the Mean Period Between Pulses of normal and parkinsons voice samples

5.8 Standard Deviation of Periods

The Standard Deviation of Periods assesses the volatility of the time interval between two successive voice tones. It is a crucial element for the control of clean speech rhythm. The measure clearly shows the state of the vocal cords, i.e., if they are vibrating in an orderly or chaotic manner during the process of speech production. The investigation was carried out to extract the most common feature of rhythm that was irregularly involved in the speech of a patient or the vocal folds of the elderly. A simple method used in this study was that, based on the time shot, the standard deviation (SD) was estimated by calculating the

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

difference of pulses. As illustrated in Figure 11, the graph of the values of the parameter showed the change in speech regularity with the enhancement of the illness. It was even observed that affected patients' distributions were more dispersed and stretched over a broader range in the violin plots hence, the pace of their voice was harder to coordinate. In the case of perceptive use alone, patients did not experience any significant effect compared to seizures induced by the breathing disorder or compared to the fundamental frequency (F0). Among jitter and frequency (F0), the standard deviation of the period was found to be more efficient among the timing based measures. The new version of the time analysis model allows more accurate identification of speech abnormalities related to timing information, which might not be observable if only pitch or amplitude parameters are followed [17].

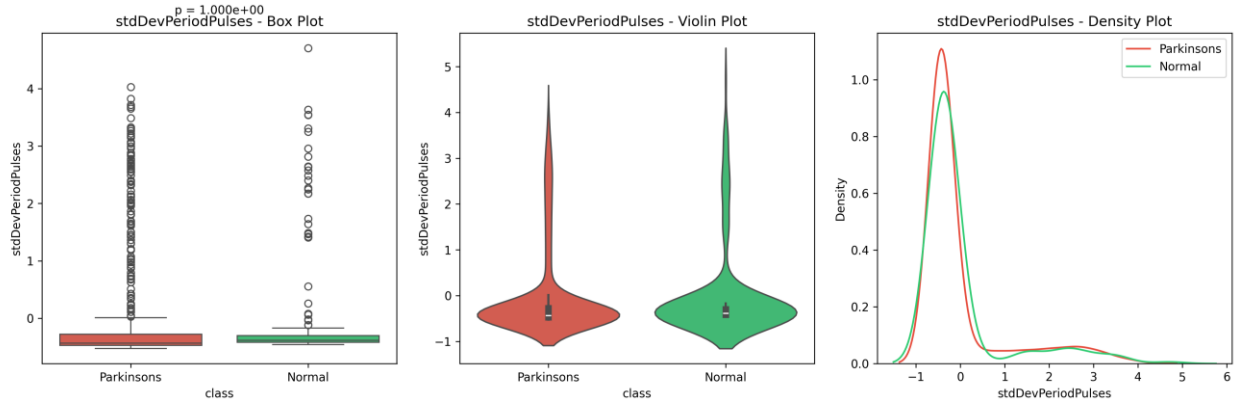

Figure 11 Box Plot, Violin plot and Wave Graph representations for Standard Deviation of Periods for for normal and Parkinson's voice samples

5.9 Number of Pulses

The number of glottal pulses that occur during phonation provides a quantitative measure of vocal effort, phonation endurance, and airflow control. This parameter was measured to assess the ability of the subjects to produce voices under different conditions. It was calculated by counting the number of glottal closures across all samples of the phonation range. Healthy voices had more pulses with a compact and dense distribution on the violin plots, as shown in Figure 12, demonstrating a steady airflow and stable vocal fold vibration. Whereas pathological voices, especially those in people with Parkinson's disease, had far fewer pulse counts with a dispersed distribution, indicating decreased phonation endurance and weak breath control. The beauty of this measure is in its simple interpretation, with fewer pulses meaning either a shortened speech burst or one suggesting voice fatigue. Despite being less sensitive than jitter or F0, it remains a useful tool for evaluating endurance, particularly in the context of long term monitoring or rehabilitation assessment.

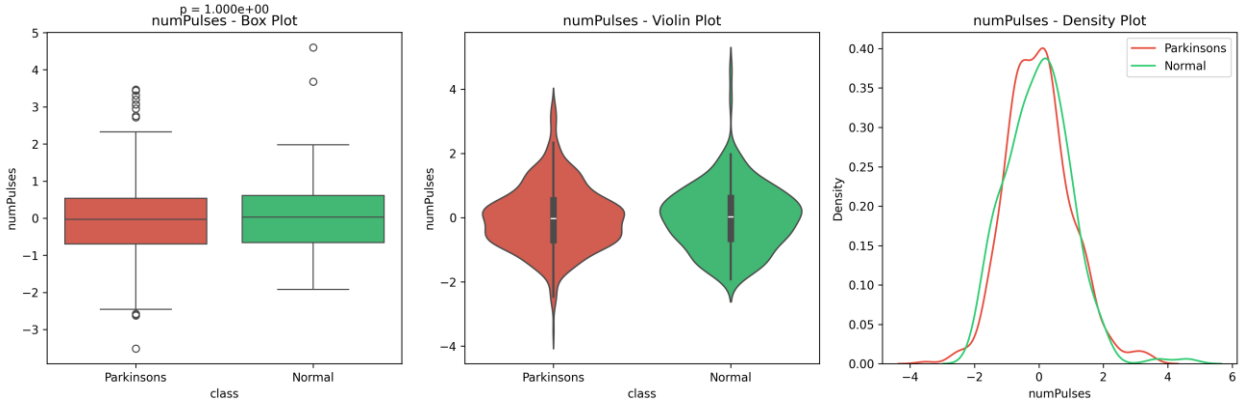

Figure 12 Box Plot, Violin plot and Wave Graph representations for Number of Pulses

Proceedings DOI: 10.21467/proceedings.7.6
ISBN: 978-81-989164-9-5

Series: AIJR Proceedings
ISSN: 2582-3922

6. CONCLUSION

In summary, this study is an innovative step towards bringing spoken diagnosis to a broader audience through artificial intelligence. By coupling machine learning with statistical rigor and interpretability in a transparent and modular pipeline, it establishes speech as an objectively valid biomarker that has scalable and noninvasive possibilities for neurological and vocal disorders. Unlike conventional diagnostics, which usually depend on costly, invasive, or inaccessible methods, this study captures the multidimensional richness of the human voice, a signal so common yet deeply informative. This work not only fortifies diagnostic sound attributes such as jitter, shimmer, and MFCC features, but also demonstrates how to strengthen this evidence by combining statistical testing with machine learning models, such as SVM, XGBoost, and MLP, for robust classification. The use of SHAP values and dimensionality reduction techniques guarantees that interpretability does not fall by the wayside, thus creating a framework that is not only effective but also explainable and clinically meaningful. The findings are only the tip of the iceberg, future studies promise elements of longitudinal assessments, real time voice monitoring, multilingual assessments, and telehealth and mobile integration. Moreover, such a framework would serve as a basis for its extension in evaluating mental health, instruction in speech therapy, and early prediction of cognitive decline. In doing so, it brings healthcare closer to accessibility and accuracy by giving voice to silent symptoms. It is indeed a stride toward the future, in which your voice does not only express how you feel but also tells how well you are, even before you know it.

REFERENCES

[1] A. Idrisoglu, A. L. Dallora, P. Anderberg, and J. S. Berglund, "Applied Machine Learning Techniques to Diagnose Voice-Affecting Conditions and Disorders: Systematic Literature Review," *J. Med. Internet Res.*, vol. 25, p. e46105, Jul. 2023. https://www.jmir.org/2023/7/e46105/

[2] Z. Zhang, "Mechanics of human voice production and control," *J. Acoust. Soc. Am.*, vol. 140, no. 4, p. 2614, Oct. 2016. https://asa.scitation.org/doi/10.1121/1.4964509

[3] K. Verdolini, C. Rosen, and R. Branski, Eds., *Classification Manual for Voice Disorders*. New York, NY: Psychology Press, 2005. https://www.routledge.com/Classification-Manual-for-Voice-Disorders/Verdolini-Rosen-Branski/p/book/9781841690416

[4] J. Robin, J. E. Harrison, L. D. Kaufman, F. Rudzicz, W. Simpson, and M. Yancheva, "Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations," *Digit. Biomark.*, vol. 4, no. 3, pp. 99–108, 2020. https://www.karger.com/Article/FullText/510820

[5] G. Fagherazzi, A. Fischer, M. Ismael, and V. Despotovic, "Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice," *Digit. Biomark.*, vol. 5, no. 1, pp. 78–88, 2021. https://www.karger.com/Article/FullText/515346

[6] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. Hosseini Ghomi, "Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease," *J. Biomed. Inform.*, vol. 104, p. 103362, Apr. 2020. https://www.sciencedirect.com/science/article/pii/S1532046419302106

[7] K. López-de-Ipiña *et al.*, "On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis," *Sensors*, vol. 13, no. 5, pp. 6730–6745, May 2013. https://www.mdpi.com/1424-8220/13/5/6730

[8] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019. https://www.nature.com/articles/s41591-018-0316-z

[9] S. A. Syed, M. Rashid, and S. Hussain, "Meta-analysis of voice disorders databases and applied machine learning techniques," *Math. Biosci. Eng.*, vol. 17, no. 6, pp. 7958–7979, Nov. 2020. https://www.aimsciences.org/article/doi/10.3934/mbe.2020404

[10] U. Petti, S. Baker, and A. Korhonen, "Areview of automatic Alzheimer's disease detection from speech and language," *J. Am. Med. Inform. Assoc.*, vol. 27, no. 11, pp. 1784–1797, Nov. 2020. https://academic.oup.com/jamia/article/27/11/1784/5907371

[11] C. O. Sakar *et al.*, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Appl. Soft Comput.*, vol. 74, pp. 255–263, Jan. 2019. https://www.sciencedirect.com/science/article/pii/S1568494618305639

[12] B. E. Sakar *et al.*, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 4, pp. 828–834, Jul. 2013. https://ieeexplore.ieee.org/document/6516742

[13] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012. https://ieeexplore.ieee.org/document/6146196

[14] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective Automatic Assessment of Rehabilitative Speech Treatment in Parkinson's Disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 181–190, Jan. 2014. https://ieeexplore.ieee.org/document/6616469

[15] M. Shen, P. Mortezaagha, and A. Rahgozar, "Explainable artificial intelligence to diagnose early Parkinson's disease via voice analysis," *Sci. Rep.*, vol. 15, no. 1, p. 11687, Apr. 2025. https://www.nature.com/articles/s41598-025-96575-6

[16] M. A. Little *et al.*, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009. https://ieeexplore.ieee.org/document/4760677

[17] A. Tsanas *et al.*, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, Apr. 2010. https://ieeexplore.ieee.org/document/5352700

[18] A. Iyer *et al.*, "A machine learning method to process voice samples for identification of Parkinson's disease," *Sci. Rep.*, vol. 13, no. 1, p. 20615, Nov. 2023. https://www.nature.com/articles/s41598-023-47568-w