[1] **Francis Obinna Oraekwe**

[2] **George Kamucha**

[3] **Shadrack Mambo**

Automated Identification System for Early Detection of Parkinson's Disease Utilizing both Voiced and Unvoiced Source Data

Journal of Electrical Systems

Abstract: - The development of automated techniques for speech analysis-based Parkinson's disease (PD) detection has attracted a lot of in especially because of its possible uses in health tele-monitoring. Due to the drawbacks of the α - Synuclein Seed Amplification Assay tech scientists are looking more closely at speech signals as a potential substitute for PD detection. In order to identify PD, this proposal desc thorough investigation that emphasizes using both voice and unvoiced source material. Acquiring relative speech data is part of the method which is followed by pitch synchronous and block processing for data preprocessing. Cloud computing will also be used for comput requirements and data storage. Feature extraction will be made easier by the Enhanced Simple Inverse Filter Tracking (ESIFT) techniqu classification will be carried out by combining Support Vector Machines, XGBoost, and LightGBM methods. Finally, to attain the best p solution for a voice recognition system, use the Ensemble technique of supervised machine learning with the goal of a detection accuracy o than 80%. When compared to current speech detection systems, the expected result is a notable increase in accuracy.

Keywords: Parkinson's disease (PD), Machine learning (ML), Speech signal processing (SSP), Cloud computing, Multi classific synchronization (PS), Block Signal Processing (BSP), Enhanced Simple Inverse Filter Tracking (ESIFT)

I. INTRODUCTION

Parkinson disease (PD) stands as the second most prevalent neurodegenerative disorder which targets neurons inside the brain's substantial nigra leading to dopamine loss. Clinical diagnosis of these dysfunctions occurs when 50% of dopaminergic neurons already suffer irreversible damage. Early diagnosis of PD during its prodromal phases proves to be essential. Parkinson's disease shows high prevalence among people aged 55 to 90 years old but also affects those younger than that range. Timely discovery of this disease allows for simple treatment yet if it remains undetected until it reaches the chronic phrase it becomes deadly and dangerous. A specialized technique detects abnormal protein deposits tied to Parkinson's disease in cerebrospinal fluid and identifies patients with complete accuracy through a detection rate that exceeds 100%. The diagnostic technique goes by the name "α-Synuclein Seed Amplification Assay" (α-Syn- SAA) which enables classification through genetic and clinical markers. The limitations of this technique motivate researchers to improve Parkinson's detection accuracy through speech signals. Such drawbacks include; the process is prohibitively expensive while patients face the fear of surgical intervention or obtaining fluid samples from the brain or spinal cord.

This study aims to develop a dependable speech recognition system that analyzes voiced and unvoiced speech patterns to enhance early diagnosis accuracy for Parkinson's disease. Achieving this aim requires the application of feature extraction methods as well as data categorization and pre-processing techniques. This research requires obtaining a validated supervised speech dataset of Parkinson's disease that includes both voiced and unvoiced speech signals. The collected speech data requires pre-processing through both pitch synchronous processing and block processing methods before analysis. The essential characteristics of vocal impairments from the illness emerge through the application of the strong feature extraction technique ESIFT alongside cross-validation methods. To design a hybrid machine learning models, both XGBoost and LightGBM techniques are applied with Support Vector Machines (SVM) and ensemble approaches that can use the unimodal dataset to accurately and early diagnose Parkinson's disease. To carry out performance metrics on the final output based on accuracy, precision, recall, sensitivity, specificity, AUC-ROC, and F1 score.

In [7], the author proposes a novel and efficient approach to the Parkinson's disease (PD) diagnosis system that uses the voice samples of the patients who have been diagnosed before. The method is based on multi-level feature selection. Chi-square and L1-Norm SVM algorithms (CLS) were used for feature selection at the first level. An audio-based depression detection system is suggested in [12], where transformation of a "deep learning (DL)" approach is used to make the data not only compact but also very essential. The proposed method increases the accuracy of the depression detection system through the structure suggested as it learns the specially important and

[1]Department of Electrical Engineering, Pan African University (PAUSTI), Juja, Nairobi, Kenya francis.oraekwe@students.jkuat.ac.ke
[2]Department of Electrical and Information Engineering, University of Nairobi (UoN), Waiyaki, Nairobi, Kenya gkamucha@uonbi.ac.ke
[3]Faculty of Engineering and Technology, Walter Sisulu University (WSU), Butterworth, Mbhashe, South Africa smambo@wsu.ac.za

characteristic markers from raw consecutive acoustic information using an end-to-end “Convolutional Neural Network-based Auto-encoder (CNN AE)” technique. In this study [2], a vowel-based artificial neural network (ANN) structure is introduced for PD prediction from a single vowel phonation. To build a voice-based model to predict Parkinson's disease, a new multi-layer neural network is suggested; 48 Parkinson's disease patients and 20 healthy people provide instances of vowels, numbers, words, and short phrases. In the second step they create ANN models from a single type of speech samples rather than merging several bases. [4], in this study proposed a “speech-based machine learning” method to distinguish the participants with healthy controls (HC) from relapsing-remitting subtype of Multiple Sclerosis (MS) using their speech. It is hypothesized that the potential for MS to create motor speech impairment similar to dysarthria could affect the phonetic posterior estimates of a "deep neural network acoustic model." This approach was created by investigating the fundamentals of the biological mechanism behind speech perception [6]. Two cochlear implant types are examined in the study: one uses a bank of optimized gamma-tone filters, while the other uses a conventional bank of band pass filters. The critical center frequencies of such filters are selected to mimic the vibrating patterns produced by auditory waves in the human cochlea. According to [9], this study suggests that traits can be gleaned from transcripts and unplanned speech utterances by fusing techniques from natural language processing and speech analysis. The transcriptions are examined using contemporary word-embedding techniques like Bidirectional Encoder Representations from Transformer (BERT). This work [3] is novel in two respects. First, the proposed assessment methodology was used to continuous speech samples for speech analysis. Second, the suitability of Wiener filters for voice de-noising in Parkinson Speech Context recognition was investigated and evaluated. They contend that speech energy, Mel spectrograms, prosody, articulation, intonation, vibrancy, and phonation are all instances of Parkinsonian traits. This work uses a short-time energy-established structure to discriminate between the voiced and unvoiced components of the speech signal [11]. After extracting the frame-wise mel frequency cepstral coefficients (MFCC) features from the voiced and unvoiced parts of each spoken phrase, the mean, variance, skewness, and kurtosis statistics are used to generate the feature vector for each spoken utterance. The support vector machine (SVM) can be used to assess the efficacy of features extracted from the voiced and unvoiced regions. In [5], he investigates whether deep learning may be used to automatically estimate the Grade, Roughness, Breathiness, Asthenia, and Strain (GRBAS) using fewer data. Eight experts created and evaluated a collection of 300 pathological continuous vowel samples that ended in /a/ (200 for training, 50 for validation, and 50 for testing). To predict the binomial distribution of GRBAS ratings from a waveform with an onset to an offset, a neural network method was suggested. The study suggested a hybrid Parkinson's disease detection system [8]. The two speech datasets used in the system's design were Parkinson's voice and speech in Italian and the datasets for mobile devices used for voice recordings at King's College London. From the voice samples in the datasets, seventeen acoustic features have been produced using the Parselmouth package. Moreover, the eight most significant aspects, determined by the features’ importance, were used in the model’s design. Choosing these features used a genetic algorithm approach: four classifiers were used in the classification step— XGBoost, random forest, logistic regression, and k-nearest neighbors. In the dataset under examination, 85 individuals underwent “open partial horizontal laryngectomy (OPHL)” of type I (22 subjects), II (32 subjects), and III (31 subjects). They had two different ways of pre-processing the available vocal data (reading assignment: extended vowel sound) to remove non-harmonic frames. After the pre-processing, they retrieved a large number of spectral, cepstral, and temporal features from the allocated harmonic frames. In this study [13], the structure of the speech signal is used to distinguish the voiced and unvoiced components. To obtain the information needed for a comprehensive analysis of the speech signal, the signal was first preprocessed using frame-wise mel-frequency cepstral coefficients (MFCC) analysis. An utterance was represented by a feature vector, which was then populated with the first four statistical moments of the speech signal: mean, variance, skewness, and kurtosis. From these statistics, estimates of the type and amount of information carried by a given speech section were made. The computational engine carried out this processing in real-time and was able to function in two different modes. In one mode, it operated on just the voiced section of a speech signal, while in the other, it operated on the section of speech that was unvoiced. According to the findings of this research work [10], a system was developed to detect Parkinson’s disease in patients through their voiced speech with 6% accuracy. The author compared traditional pipelines against end-to-end approaches. A multilayer perception (MLP) algorithm of deep reinforcement learning was used to train the extracted data source.

II. METHODS

Fig 1 displays an overview of the entire methods applied in this research work, from the dataset to pre-processing, then feature extraction and finally to the classification phase.

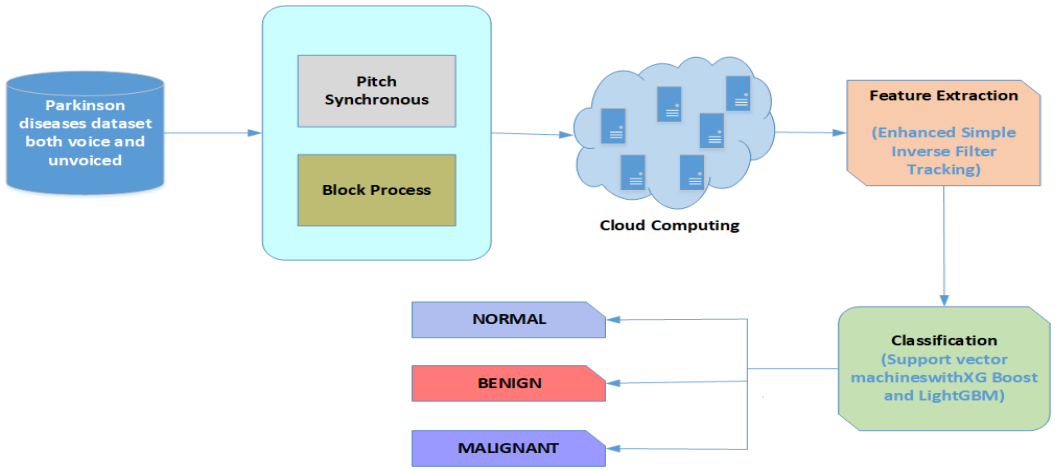
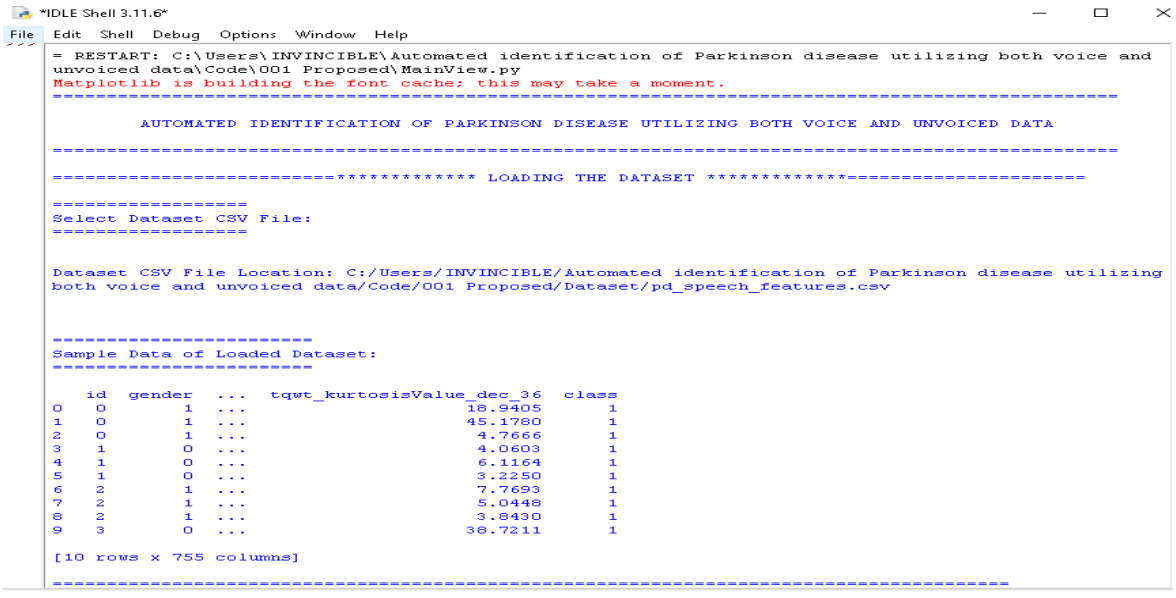


Fig 1: Overall diagram of proposed system.

Dataset

This dataset, which was finished on November 5, 2018, is for sale in the field of speech analysis and Parkinson's disease research training. The study's data were from 758 people, both men and women, with PD and in good health. Their ages ranged from 33 to 87 (65.1 ± 10.9), and they were all from the Department of Neurology at Cerrahpasa University, Faculty of Medicine, Istanbul University, Turkey. The microphone was adjusted to 44.1 KHz during the data collecting process, and each subject's sustained phonation of the vowel /a/ was recorded three times. Parkinson's disease (PD) patients' speech recordings have been subjected to a variety of speech signal processing algorithms, such as Time Frequency Features, Mel Frequency Cepstral Coefficients (MFCCs), Wavelet Transform based Features, Vocal Fold Features, and TWQT features, in order to extract clinically relevant information for PD assessment. There are 758 patients in this dataset, both male and female. The loading of the uploaded dataset is displayed in Fig. 2.



```
*IDLE Shell 3.11.6*
File  Edit  Shell  Debug  Options  Window  Help
= RESTART: C:\Users\INVINCIBLE\Automated identification of Parkinson disease utilizing both voice and
unvoiced data\Code\001 Proposed\MainView.py
Matplotlib is building the font cache; this may take a moment.
================================================================================================

        AUTOMATED IDENTIFICATION OF PARKINSON DISEASE UTILIZING BOTH VOICE AND UNVOICED DATA

================================================================================================

======================================== LOADING THE DATASET ============================================

==================
Select Dataset CSV File:
==================

Dataset CSV File Location: C:/Users/INVINCIBLE/Automated identification of Parkinson disease utilizing
both voice and unvoiced data/Code/001 Proposed/Dataset/pd_speech_features.csv


------------------------
Sample Data of Loaded Dataset:
------------------------

   id  gender  ...  tqwt_kurtosisValue_dec_36  class
0   0       1  ...                    18.9405      1
1   0       1  ...                    45.1780      1
2   0       1  ...                     4.7666      1
3   1       0  ...                     4.0603      1
4   1       0  ...                     6.1164      1
5   1       0  ...                     3.2250      1
6   2       1  ...                     7.7693      1
7   2       1  ...                     5.0448      1
8   2       1  ...                     3.8430      1
9   3       0  ...                    38.7211      1

[10 rows x 755 columns]

================================================================================================
```

Fig 2: Dataset loaded in python

Pre-processing

Pitch synchronization and block processing are crucial pre-processing techniques for speech data analysis in Parkinson disease. Both spoken and unvoiced speech can be used using these techniques. Pitch synchronous processing separates data based on their fundamental frequency, allowing for a more thorough analysis of specific speech traits. This method improves the accuracy of collecting relevant data by ensuring alignment with the periodicity of the voice signal. In order to handle voiced and unvoiced portions equally, block processing, on the

other hand, splits the signal into smaller blocks. To overcome memory and computational limitations, it is feasible to train and test the proposed Parkinson's disease model using the entire dataset by utilizing cloud computing resources. The computational load will fall on cloud-based solutions, which offer powerful and scalable computing resources as needed. This eliminates memory limitations and speeds up model training throughout the entire dataset in Parkinson's disease research by enabling parallel processing and efficient use of cloud-based infrastructure.

Block Signal Processing: The Block processing can be conducted by applying either the block convolution or the block recursion. For the purpose of this research, the block convolution method was used during the block processing. This method groups signal samples into blocks, to be processed one after the other.
The operation of a finite impulse response (FIR) filter is described by a finite convolution as;

$$\mathbf{y(n)} = \sum_{k=0}^{I-1} \mathbf{h(k)\, x(n-k)} \qquad \text{..... (1 a)}$$

where x(n) is causal, h(n) is causal and of length LL" role="presentation" style="position:relative;" tabindex="0">L*L*, and the time index LL" role="presentation" style="position:relative;" tabindex="0">n*n*, nn" role="presentation" style="position:relative;" tabindex="0">goes from zero to infinity or some large value.

The delay between the input and output signals is dependent on the number of samples in each block.

$$\mathbf{x_q(n) = x_m\,(n - d_q) = x_m\,(n + t_m - t_q)} \qquad \text{..... (1 b)}$$

t_q is the synthesis pitch marks, d_q is the sequence of delays (= $t_q - t_m$),

Pitch synchronous: The least-square overlap-add synthesis approach can be used to obtain the synthetic signal x(n). By splitting the voice waveform into tiny overlapping segments, this approach alters the pitch and length of a speech signal. The segments are repeated several times or removed to alter duration, and they are moved closer or farther to alter pitch.

$$\mathbf{x_m\,(n) = h_m\,(t_m - n)\;x(n)} \qquad \text{..... (2 a)}$$

t_m represents the pitch marks (set at the pitch-synchronous rate on the voiced portions of the signal, & at a constant rate on the unvoiced portions).
x(n) represents the digitalized speech waveform.
$x_m(n)$ represents the sequences of short-time signals,
$h_m(n)$ is the sequences of pitch synchronous

but first, we calculate the short term energy of the input signal x(n);

$$\mathbf{S} = \sum_{n=-\infty}^{\infty} \mathbf{x(n)} \qquad \text{S = np.sum (frames ** 2, axis = 1)} \qquad \text{..... (2 b)}$$

Feature Extraction

The Enhanced Simple Inverse Filter Tracking (ESIFT) algorithm is a sophisticated technique for feature extraction in the context of Parkinson's disease that records both voice and unvoiced speech components. ESIFT improves the extraction of pertinent information from voice recordings affected by Parkinson's disease by combining inverse filtering and tracking algorithms. By breaking down the speech signal into its source and filter components, this method can capture the key elements of the vocal impairments brought on by the illness. The improved tracking technology guarantees precise monitoring of dynamic changes in voiced and unvoiced regions and offers a thorough depiction of speech disorders. ESIFT's advanced features (id, gender, PPE, DFA, RPDE, numPulses, numPeriodspulses, stdDevPeriod pulses, and Locpct jitter) help improve diagnostic and monitoring processes in both voiced and unvoiced speech domains by identifying subtle variations in voice patterns associated with Parkinson's disease. Cognitive impairment, autonomic dysfunction, sleep disruptions, anxiety and sadness, difficulty speaking and swallowing, exhaustion, pain, dysautonomia, and mood swings are examples of unvoiced traits. The Enhanced Simple Inverse Filter Tracking algorithm is used to classify speech segments according to their voicing and to estimate the pitch period of the speech that has been labeled as voiced.

The decimated speech signal /Wn/ is then inverse filtered by the FIR filter with transfer function

$$\mathbf{A(z) = 1 + \sum_{i=1}^{m} a_i\, z^{-1}} \qquad \text{..... (3 a)}$$

In the sampled data-time domain this equation is equivalent to

$$y_n = w_n + \sum_{i=1}^{4} a_i \, w_{n-1} \qquad \text{..... (3 b)}$$

where {yn} is the spectrally flattened prediction error signal.

Classification

In the analysis of voiced and unvoiced speech in Parkinson's disease, a robust classification framework is produced by combining the XGBoost and LightGBM approaches with Support Vector Machines (SVM). SVM establishes a helpful baseline by dividing data into distinct groups based on identified patterns. The performance of the model is enhanced by combining two powerful gradient boosting algorithms, XGBoost and LightGBM, which handle complex interactions in the data and continuously enhance its predictions.

XGBoost technique comprises of smaller colSample_byTree values which are used to simplify models and avoid over fitting, which has been a significant problem with SVM. As a result, the model performs better during training and increases accuracy during testing.

LightGBM approach offers two primary types of feature importance, which are split scores and gain scores. The number of times a feature is applied to split the data across all of the model's trees is indicated by the Split score. Gain score measures the increase in the model's accuracy attained by employing a certain feature for splitting, while it is helpful for determining which features are most frequently used in the decision-making process.
The capabilities of SVM's discriminative abilities with the boosting algorithms' ability to capture intricate feature interactions, are further combined by the ensemble technique. Because it takes into account the subtle patterns found in speech data that are in both voice and unvoiced segments, combining the two is particularly helpful for identifying Parkinson's disease.

Support Vector Machine Model**:** The number of features was determined according to the cost parameter for the feature selection-based L1-Norm SVM. The dataset with n samples is expressed as:

$$S = \{(xi, yi) \mid xi \in R^n, \; yi \in \{-1, 1\}\}^{k_i} = 1 \qquad \text{..... (4a)}$$

where: xi is the ith sample which has n features and a class label (yi).

The SVM in the classification problem with two classes can be re-arranged in the below equation to correct classification errors resulting from the distance from the margin.

$$yi\,(wxi - b) \geq 1 - d, \quad di \geq 0, \; i = 1, \ldots, k \qquad \text{..... (4b)}$$

XGBoost Model: In filling out missing data, in cross validation by stopping the process on time once there's no more improvement.

```
start = time.time()
xg=xgb.XGBClassifier(max_depth=7,learning_rate=0.05,
                     silent=1,eta=1,objective='multi:softprob',
                     num_round=50,num_classes=6)
```

$$\text{Similarity Score (Gain) of XGBoost} = \frac{\sum R^2}{N_R + \lambda} \qquad \text{..... (5)}$$

Where:

(X, Y) = Dataset split to LHS & RHS of leaf

λ = Regularization parameter

γ = Threshold on gain,
(This determines if the tree will split or not)

R = Pseudo (X, Y) of dataset

N_R = Numbers of given R

Light GBM Model: This algorithm is based on decision trees to tackle complex problems. It gives speed, and can easily train a large amount of data by using less memory than other algorithms. This speed and efficiency results to greater accuracy and better performance.

$$\text{Gain}(j) = \frac{G^2_j}{H_j + \lambda} + \frac{(G - G_j)^2}{H - H_j + \lambda} - \frac{G^2}{H + \lambda} \quad \text{..... (6)}$$

where: G = Gradients for the dataset

H = Hessians for the dataset

λ = Regularization terms

eta = Learning rate

Tree Ensemble Model: This uses a lot of decision trees called Random Forest, where each tree is little different from the others. When a new data is obtained, we take the majority vote of ensemble to get a final result.

$$y_i = \text{Ø}(x_i) = \sum_{K=1}^{k} f_k(x_i),\ f_k \in F \quad \text{..... (7)}$$

where $F = \{f(x) = wq(x)\}(q : R^m \rightarrow T, w \in R^T)$ is the space of regression trees (also known as CART)

(**q** represents the structure of each tree that maps an example to the corresponding leaf index)
(**T** is the number of leaves in the tree) (**w** is the leaf weight)

f_k corresponds to an independent tree structure q and leaf weights w

k is the additive functions to predict the output.

x_i is the speech signal (data)

Evaluation metrics

To evaluate the performance of the proposed system, the following parameters have been used:-

Accuracy is simply a ratio of the correctly predicted observations to the total observations,

$$\text{Accuracy} = \frac{TP + TN}{(TP + FP + FN + TN)} \quad \text{..... (8 a)}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The precision is defined by:

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad \text{..... (8 b)}$$

Recall is the ratio of correctly predicted positive observations to all observations in the actual class. It is formulated by:

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad \text{..... (8 c)}$$

F1 score is the weighted average of precision and recall. Therefore, this score takes both false positives and false negatives into account. The F1 score is defined by:

$$F1 = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad \text{..... (8 d)}$$

NPV defines the fraction of the tests that correctly detect healthy individuals.

$$\text{NPV} = \frac{TN}{(TN + FN)} \quad \text{..... (8 e)}$$

Where:
TP is True Positive TN is True Negative FP is False Positive FN is False Negative

III. RESULTS

The diagram in Fig 3 is the GUI of the processes, each click automatically initiates the selected process.

Fig 3: GUI Model based on the specific objectives of this research

Pre-processing

IDLE Shell 3.11.6

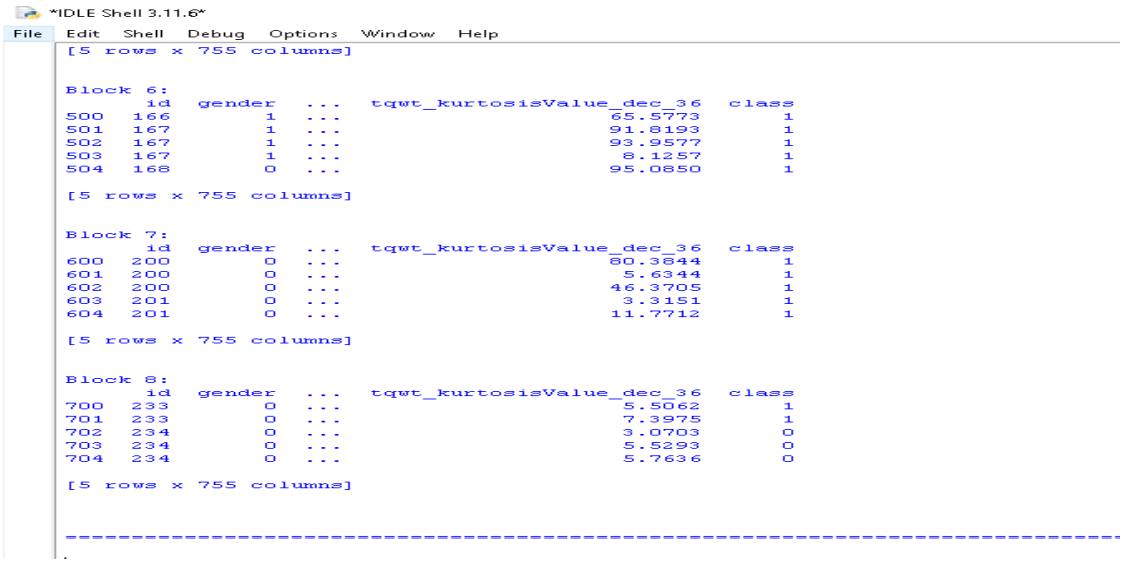
File Edit Shell Debug Options Window Help

```
========================************ PRE-PROCESSING PROCESS ************========================

================
Preprocessed Data:
================

   id  gender       PPE  ...  class  pitch_features  pitch_std
0   0       1  0.85247  ...      1        0.714333   0.140141
1   0       1  0.76686  ...      1        0.667110   0.116105
2   0       1  0.85083  ...      1        0.705563   0.132986
3   1       0  0.41121  ...      1        0.600167   0.192867
4   1       0  0.32790  ...      1        0.552000   0.235712

[5 rows x 757 columns]

================
Block Parameters:
================

Block 1:
   id  gender  ...  tqwt_kurtosisValue_dec_36  class
0   0       1  ...                    18.9405      1
1   0       1  ...                    45.1780      1
2   0       1  ...                     4.7666      1
3   1       0  ...                     4.0603      1
4   1       0  ...                     6.1164      1

[5 rows x 755 columns]


Block 2:
     id  gender  ...  tqwt_kurtosisValue_dec_36  class
100  33       1  ...                     2.7904      0
101  33       1  ...                     4.5031      0
102  34       0  ...                     3.8367      0
103  34       0  ...                     3.3496      0
104  34       0  ...                     3.2542      0

[5 rows x 755 columns]
```

IDLE Shell 3.11.6

File Edit Shell Debug Options Window Help

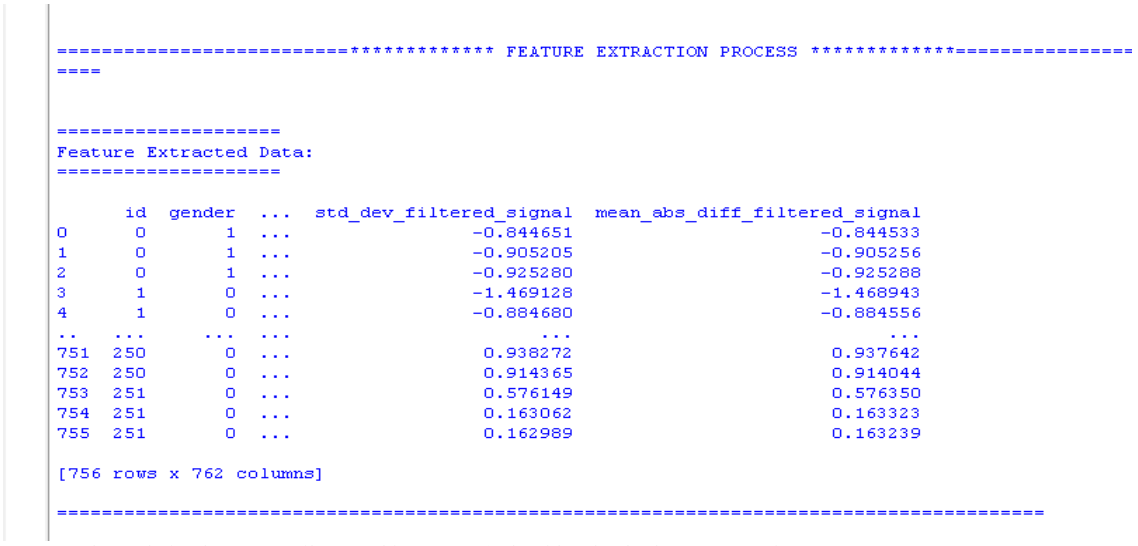
```
Block 3:
       id  gender  ...  tqwt_kurtosisValue_dec_36  class
200   66       0  ...                     9.1055      1
201   67       0  ...                     2.5336      1
202   67       0  ...                    13.1466      1
203   67       0  ...                     3.7013      1
204   68       0  ...                     2.7978      1

[5 rows x 755 columns]


Block 4:
       id  gender  ...  tqwt_kurtosisValue_dec_36  class
300  100       1  ...                   107.8370      1
301  100       1  ...                    68.8642      1
302  100       1  ...                     6.0970      1
303  101       1  ...                    10.3949      1
304  101       1  ...                    75.2606      1

[5 rows x 755 columns]


Block 5:
       id  gender  ...  tqwt_kurtosisValue_dec_36  class
400  133       0  ...                    14.0905      1
401  133       0  ...                     4.7872      1
402  134       1  ...                    73.2197      1
403  134       1  ...                    92.2488      1
404  134       1  ...                     2.6109      1

[5 rows x 755 columns]
```



```
*IDLE Shell 3.11.6*
File  Edit  Shell  Debug  Options  Window  Help
[5 rows x 755 columns]

Block 6:
      id  gender  ...  tqwt_kurtosisValue_dec_36  class
500  166       1  ...                   65.5773      1
501  167       1  ...                   91.8193      1
502  167       1  ...                   93.9577      1
503  167       1  ...                    8.1257      1
504  168       0  ...                   95.0850      1

[5 rows x 755 columns]

Block 7:
      id  gender  ...  tqwt_kurtosisValue_dec_36  class
600  200       0  ...                   80.3844      1
601  200       0  ...                    5.6344      1
602  200       0  ...                   46.3705      1
603  201       0  ...                    3.3151      1
604  201       0  ...                   11.7712      1

[5 rows x 755 columns]

Block 8:
      id  gender  ...  tqwt_kurtosisValue_dec_36  class
700  233       0  ...                    5.5062      1
701  233       0  ...                    7.3975      1
702  234       0  ...                    3.0703      0
703  234       0  ...                    5.5293      0
704  234       0  ...                    5.7636      0

[5 rows x 755 columns]

================================================================================
```

Fig 4: Pre-processed data using pitch synchronous and block processing.

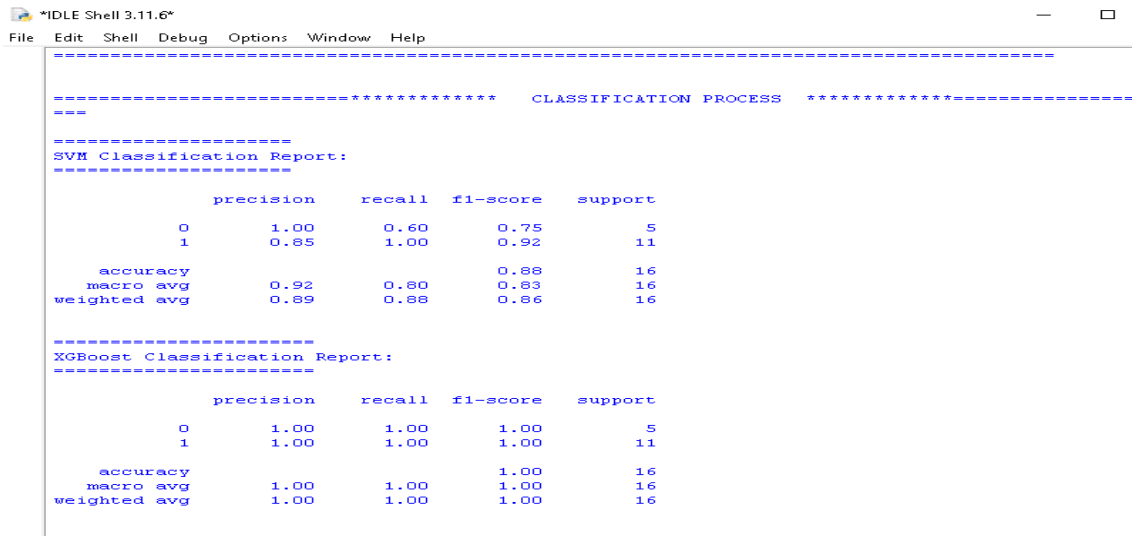
Feature Extraction



```
========================================== FEATURE EXTRACTION PROCESS ==========
====


====================
Feature Extracted Data:
====================

      id  gender  ...  std_dev_filtered_signal  mean_abs_diff_filtered_signal
0      0       1  ...                -0.844651                      -0.844533
1      0       1  ...                -0.905205                      -0.905256
2      0       1  ...                -0.925280                      -0.925288
3      1       0  ...                -1.469128                      -1.468943
4      1       0  ...                -0.884680                      -0.884556
..   ...     ...  ...                      ...                            ...
751  250       0  ...                 0.938272                       0.937642
752  250       0  ...                 0.914365                       0.914044
753  251       0  ...                 0.576149                       0.576350
754  251       0  ...                 0.163062                       0.163323
755  251       0  ...                 0.162989                       0.163239

[756 rows x 762 columns]

================================================================================
```

Fig 5: Enhanced Simple Inverse Filter Tracking (ESIFT) algorithm for the feature extraction

Classification



```
*IDLE Shell 3.11.6*
File  Edit  Shell  Debug  Options  Window  Help
================================================================================


======================================   CLASSIFICATION PROCESS  ==============
===

======================
SVM Classification Report:
======================

              precision    recall  f1-score   support

           0       1.00      0.60      0.75         5
           1       0.85      1.00      0.92        11

    accuracy                           0.88        16
   macro avg       0.92      0.80      0.83        16
weighted avg       0.89      0.88      0.86        16


=======================
XGBoost Classification Report:
=======================

              precision    recall  f1-score   support

           0       1.00      1.00      1.00         5
           1       1.00      1.00      1.00        11

    accuracy                           1.00        16
   macro avg       1.00      1.00      1.00        16
weighted avg       1.00      1.00      1.00        16
```

```
==========================
LightGBM Classification Report:
==========================

              precision    recall  f1-score   support

           0       1.00      0.80      0.89         5
           1       0.92      1.00      0.96        11

    accuracy                           0.94        16
   macro avg       0.96      0.90      0.92        16
weighted avg       0.94      0.94      0.94        16


==========================
Ensemble Classification Report:
==========================

              precision    recall  f1-score   support

           0       1.00      0.80      0.89         5
           1       0.92      1.00      0.96        11

    accuracy                           0.94        16
   macro avg       0.96      0.90      0.92        16
weighted avg       0.94      0.94      0.94        16


============================
Ensemble Model Accuracy:0.9954
============================


================================================================================================
```

Fig 6: Classification process using XGBoost and LightGBM techniques with Support Vector Machine (SVM) and Ensemble technique, for the Model Accuracy.

The proposed work is compared with evaluation metrics such as:-

Accuracy: This is the state of be precise and correct, with insignificant errors.
Sensitivity: The quality to quickly detect essential features in the signals
Specificity: The quality of relating uniquely only to the unessential features.
AUC-ROC: To provide aggregate measure of performance across classifications threshold.
Recall: To measure the performance of classifier in binary & multiclass classifications.
F1-score: This integrates precision and recall into a single metric, for better understanding.

```
================================================================================================


==================************** PERFORMANCE METRICS **********************=====

====================
  1. Accuracy Graph [%]
====================


=================
  2. Recall Graph [%]
=================


====================
  3. Specificity Graph [%]
====================


====================
  4. Sensitivity Graph [%]
====================


====================
  5. AUC ROC Graph [%]
====================


====================
  6. F1-Score Graph [%]
====================


================================================================================================
```

Fig 7: Evaluation metrics running in python

From the above evaluation metrics, the following graphs were obtained;

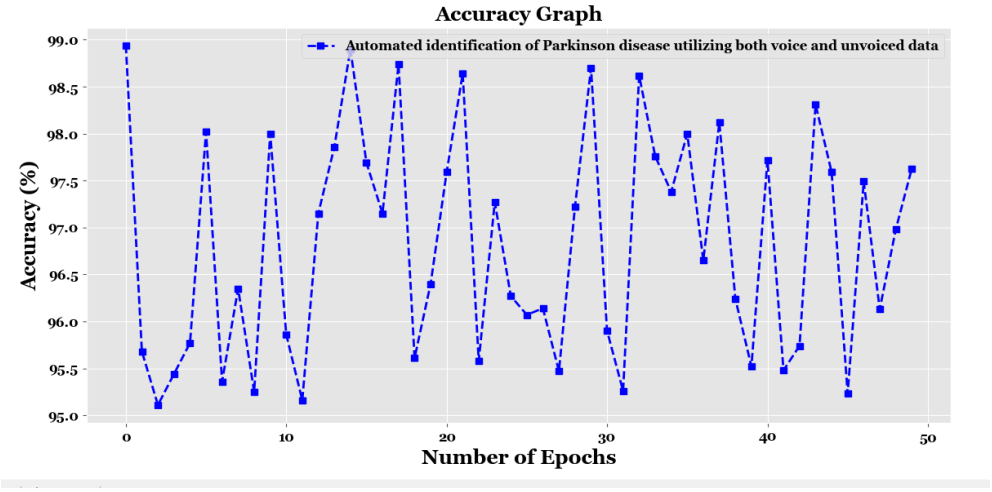


Fig 8: Accuracy graph

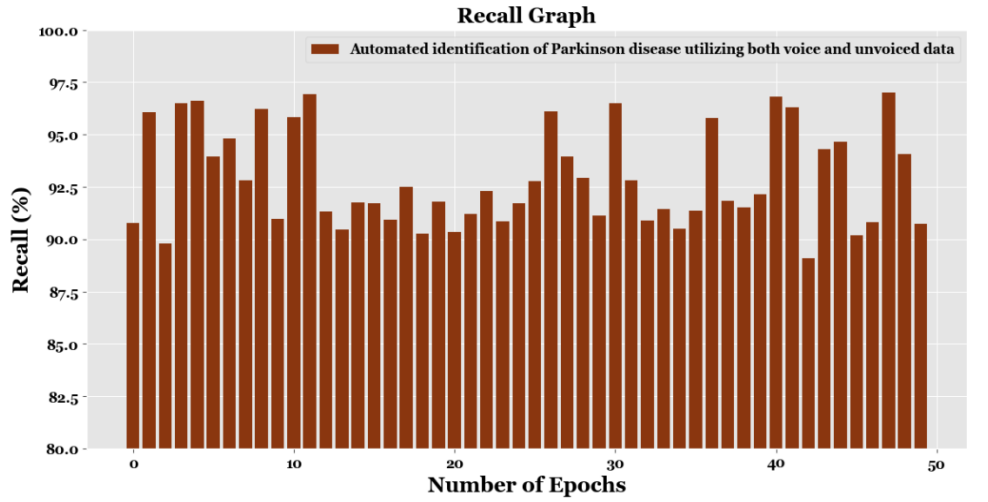


Fig 9: Recall graph

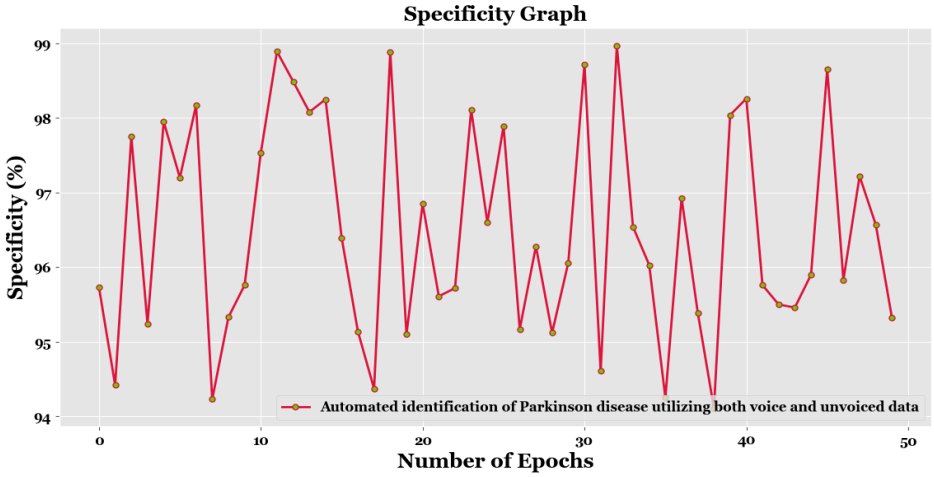


Fig 10: Specificity graph

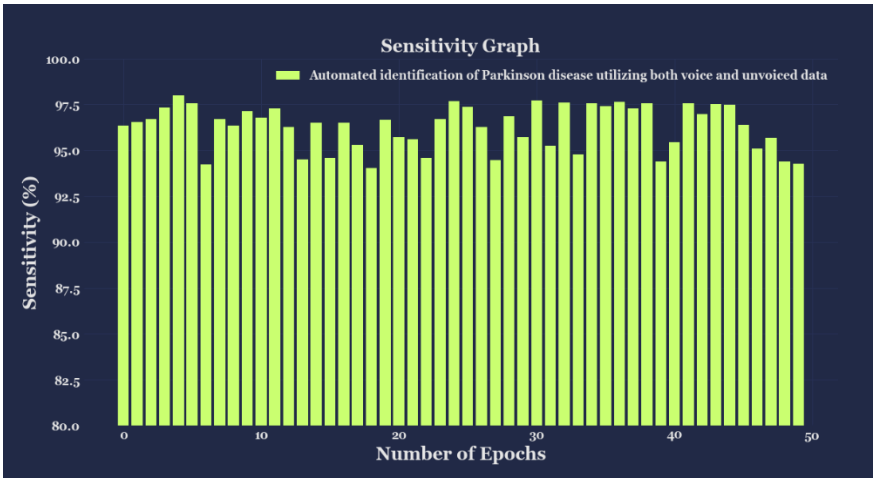


Fig 11: Sensitivity graph

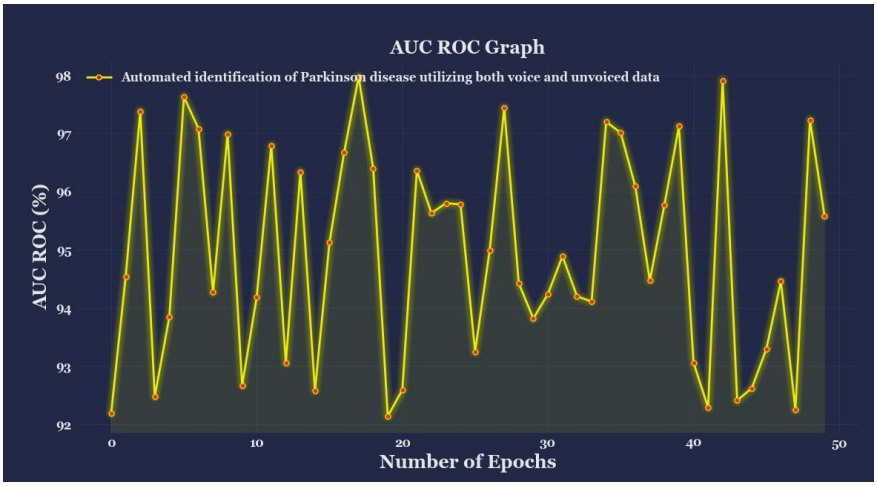


Fig 12: AUC ROC graph

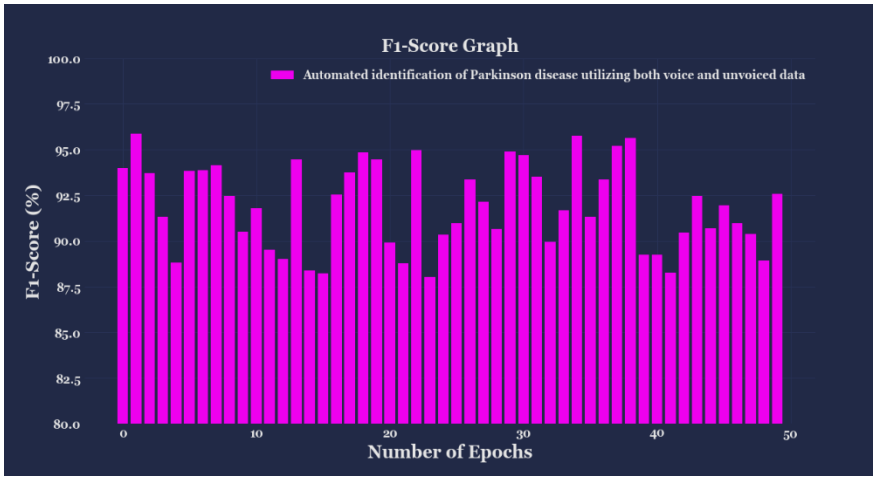


Fig 13: F1-Score graph

IV. DISCUSSION

The process begins by clicking on the Dataset button on the GUI model, this process allows the dataset to be uploaded into the IDLE of python, which takes a short while to load in, before the pre-processing phase can be

initiated by clicking on the Pre-processing button on the GUI model. The Fig.4 shows how the data are pitch synchronized and divided into block parameters during the pre-processing. Once this step is completed, the feature extraction phase can be initiated by clicking on the Feature extraction button. The Fig.5 shows how relevant speech features are extracted from the pre-processed data, applying the Enhanced Simple Inverse Filter Tracking method in the process. The next step is by initialling the classification process, as seeing in (Fig.6), which adapts the SVM model together with the XGBoost and LightGBM classifiers, and combined with the Ensemble model to give a resultant output of 99.54% accuracy rate which is key in detection of early Parkinson's disease. The final step can now be initiated by click on the Performance metrics button, this allows the evaluation process to commence in (Fig.7). The outcome of the evaluation produces the following graphs;

The Accuracy graph in (Fig.8) displays the rate at which the model performs, it shows us how correct a prediction is made on dataset on a percentage scale. The X-axis shows the steps of the training progression, while the Y-axis shows the level of accuracy. From this graph, the maximum level of attainable accuracy on the Parkinson' disease detection is 99.54%, which becomes the workable level.

The Recall graph in (Fig.9) known as the Precision-Recall curve displays the relationship between the values of precision and recall for the classification models across different settings of threshold. The higher the threshold, the higher the precision but lower the recall, and vice-verse.

The Specificity graph in (Fig.10) known as the Receiver Operating Characteristics (ROC) displays a correct identification of individuals with absence of a particular disease, this illustrates the trade-off between correct negative cases and incorrect negative cases across multiple cut-off points on the scale of test result. A better specificity test is as a result of higher curves.

The Sensitivity graph shows how a change in one input variable can affect the outcome of a model, which helps to notify the most significant variables in arriving to a final resultant. Thus it can visualise which factors are more critical in decision making, the input variable is plotted on the x-axis while output variable is plotted on the y-axis. The graph in (Fig.11) has tornado charts which display its variables vertically with bars showing changes within possible output values, thus allows fast visual comparison of most significant factors.

The AUC ROC graph is known as Area under the Receiver Operating Characteristics Curve which represents the performance of binary classification model across several thresholds, whereby the rate of true positive is plotted against false positive. This graph in (Fig.12) is important because it sets balance between sensitivity and specificity.

The F1-Score graph helps classification model in balancing between precision and recall. The F1 score displays values between 0 and 1 across datasets and model variations. A higher F1 score as shown in (Fig.13) indicates a better balanced and model performance.

V. CONCLUSIONS

Parkinson disease dataset is pre-processed by applying pitch synchronous and block processing Segmentation is essential to improve the accuracy of identifying diseased and normal voices and to detect the disease's progression.

Utilizing cloud computing resources is a workable method to get beyond memory and computational limitations in training and testing the proposed model for Parkinson's disease utilizing the whole dataset. The computational burden will be moved to cloud-based platforms with on-demand access to scalable and potent computer resources.

Improving the Simple Inverse Filter Tracking (SIFT) algorithm is a practical way to solve the method's lack of learnable parameters and make it work with less demanding hardware for clinical situations. A balance is maintained between hardware needs and model complexity in the SIFT method with the integration of lightweight learnable components or adaptable features.

Enhanced Simple Inverse Filter Tracking (ESIFT) method with a multilingual phonetic set is an efficient way to overcome the language-specific limitation in the proposed feature extraction strategy for Chinese, Spanish, and English speakers.

An ensemble technique using Support Vector Machines (SVM) in conjunction with XGBoost and LightGBM was used to overcome classification issues resulting from limited pathological sample numbers and gender imbalance. This group minimizes the effects of sparse data and gender inequality by using the advantages of each method to improve classification performance.

In conclusion, this research successfully demonstrated the effectiveness of utilizing a dataset that includes voiced and unvoiced supervised speech signals for detecting Parkinson's disease. By employing advanced techniques such as Enhanced Simple Inverse Filter Tracking (ESIFT), Support Vector Machines (SVM), XGBoost, LightGBM, and Ensemble Machine Learning algorithms, we achieved an impressive accuracy rate of 99.54%. This significantly exceeds the existing benchmark of 75% in this field. Moreover, our results are comparable to the α-Synuclein Seed Amplification Assay, which reports accuracy rates exceeding 100%. Given the minimal difference of only -0.46% accuracy, this research alleviates concerns regarding the invasive procedures typically required for

cerebrospinal fluid collection to determine PD status. Our findings underscore that speech signal analysis for Parkinson's disease detection is superior to traditional methods, offering simplicity, affordability, flexibility, and reliability advantages.

LIST OF ABBREVIATIONS

AE	Auto-Encoder
ANN	Artificial Neural Networks
AUC-ROC	Area Under the ROC Curve
BERT	Bidirectional Encoder Representations from Transformer
CAS	Complete Active Speech
CLS	Chi-square and L1-Norms algorithms
CNN	Convolutional Neural Networks
DT	Decision Tree algorithm
ESIFT	Enhanced Simple Inverse Filter Tracking
GMM-UBM	Gaussian Mixture Model-Universal Background Model
GRBAS scores	Grade, Roughness, breathiness, asthenia, strain
GUI	Graphic User Interface
HNR	Harmonic-to-Noise Ratio
INFVo	Intelligibility, Fluency, Voicing and noise
LightGBM	Light Gradient-Boosting Machine
MFCC	Mel frequency Cepstral coefficient
MLP	MultiLayer Perceptron
PD	Parkinson's disease
QCP	Quasi-Closed Phase
RNN	Recurrent Neural Networks
SIFT	Simple Inverse Filter Tracking
SVM	Support Vector Machine
XGBoost	eXtreme Gradient Boosting

DECLARATIONS

Availability of data and materials
The speech dataset used in this research is attached as the supplementary material for further assessment.

Supplementary Description: The attachment in the supplementary section contains a dataset of extracted features from mixed speech signals including patients with Parkinson's disease and healthy individuals as well.

Competing interests
The authors declare no conflict of interest

Funding
This research received no specific grant from any funding agency, commercial or non-profit organizations.

Authors' Contributions
F O O carried out 85% of the entire research, which comprising of the acquirement of speech dataset used in the research, feature extraction of data, classification and training of the dataset using machine learning techniques, evaluation metrics of the outcome and writing of the reports. G K carried out the pre-processing analysis and validation of data. S M carried out the literature review and editing of this research work. All authors read and approved the final manuscript.

Acknowledgements
The authors would like to express their gratitude to all those who contributed to the development of this research. We appreciate the constructive comments from the reviewers and editorial team, which helped improve the clarity and quality of the manuscript.

Ethical statement
This research did not required ethical approval.

REFERENCES

[1] Carullo, A., Vallan, A., Fantini, M., & Succo, G. (2023). Vocal-Feature Based Classification of Post-Laryngectomy Patients for Rehabilitation Monitoring. *IEEE Transactions on Instrumentation and Measurement*. *PP (99): 1-1*

[2] Demir, F., Siddique, K., Alswaitti, M., Demir, K., & Sengur, A. (2022). A simple and effective approach based on a multi-level feature selection for automated Parkinson's disease detection. *Journal of Personalized Medicine*, *12*(1), 55.

[3] Faragó, P., Ștefănigă, S. A., Cordoș, C. G., Mihăilă, L. I., Hintea, S., Peștean, A. S., ... & Ileșan, R. R. (2023). CNN-Based Identification of Parkinson's disease from Continuous Speech in Noisy Environments. *Bioengineering*, *10*(5), 531.

[4] Gosztolya, G., Svindt, V., Bóna, J., & Hoffmann, I. (2023). Extracting Phonetic Posterior- Based Features for Detecting Multiple Sclerosis From Speech. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

[5] Hidaka, S., Lee, Y., Nakanishi, M., Wakamiya, K., Nakagawa, T., & Kaburagi, T. (2022). Automatic GRBAS Scoring of Pathological Voices using Deep Learning and a Small Set of Labelled Voice Data. *Journal of Voice*.

[6] Huckvale, M., Liu, Z., & Buciuleac, C. (2023). Automated voice pathology discrimination from audio recordings benefits from phonetic analysis of continuous speech. *Biomedical Signal Processing and Control*, *86*, 105201.

[7] Islam, R., Abdel-Raheem, E., & Tarique, M. (2022). A novel pathological voice identification technique through simulated cochlear implant processing systems. *Applied Sciences*, *12*(5), 2398.

[8] Lamba, R., Gulati, T., Jain, A., & Rani, P. (2023). A Speech-Based Hybrid Decision Support System for Early Detection of Parkinson's disease. *Arabian Journal for Science and Engineering*, *48*(2), 2247-2260.

[9] Liu, W., Liu, J., Peng, T., Wang, G., Balas, V. E., Geman, O., & Chiu, H. W. (2023). Prediction of Parkinson's disease based on artificial neural networks using speech datasets. *Journal of Ambient Intelligence and Humanized Computing*, *14*(10), 13571-13584.

[10] N.P. Narendra, Bjorn Schuller, and Paavo Alku, (2021). Detection of Parkinson's disease from speech using voice data information, *IEEE/ACM Transactions on Audio, Speech, and Language Processing Language Processing*.

[11] Pérez-Toro, P. A., Arias-Vergara, T., Klumpp, P., Vásquez-Correa, J. C., Schuster, M., Noeth, E., & Orozco-Arroyave, J. R. (2022). Depression assessment in people with Parkinson's disease: The combination of acoustic features and natural language processing. *Speech Communication*, *145*, 10-20.

[12] Sardari, S., Nakisa, B., Rastgoo, M. N., & Eklund, P. (2022). Audio based depression detection using Convolutional Autoencoder. *Expert Systems with Applications*, *189*, 116076.

[13] Warule, P., Mishra, S. P., & Deb, S. (2023). Significance of voiced and unvoiced speech segments for the detection of common cold. *Signal, Image and Video Processing*, *17*(5), 1785- 1792.