A Project Report

on

<u>NeuroSpeechBERT</u>:

The fusion of speech analysis and BERT-based deep learning for Parkinson's progression detection.

carried out as part of the ***Minor Project IT3270*** *Submitted*

by

Shikha Chhabra
229302223

in partial fulfilment for the award of the degree of

Bachelor of Technology

in

Information Technology

MANIPAL UNIVERSITY
JAIPUR

INSPIRED BY LIFE

Department of Information Technology

School of CSE & IT

MANIPAL UNIVERSITY JAIPUR

RAJASTHAN, INDIA

2024-2025

CERTIFICATE

Date: 17-04-2025

This is to certify that the minor project titled **NeuroSpeechBERT: The fusion of speech analysis and BERT-based deep learning for Parkinson's progression detection** is a record of the bonafide work done by **SHIKHA CHHABRA** (229302223) submitted in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Information Technology of Manipal University Jaipur, during the academic year 2024-25.

Mr. Venkatesh Gauri Shankar

Assistant Professor (Senior Scale)

Project Guide, Department of Information Technology

Manipal University Jaipur

Dr. Pratistha Mathur

HoD, Department of Information Technology

Manipal University Jaipur

ABSTRACT

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects millions of individuals worldwide. Despite advancements in medical imaging and clinical diagnostics, early detection remains a significant challenge, often delaying intervention until noticeable motor symptoms appear. Among the earliest and most consistent indicators of PD are subtle changes in speech patterns, which typically emerge years before formal diagnosis. Recognizing this, the present project — *NeuroSpeechBERT* — aims to harness the power of speech analysis combined with deep learning to facilitate the early detection and monitoring of Parkinson's Disease.

The project proposes a multi-branch framework that leverages three complementary approaches: numerical analysis of acoustic and prosodic features, spectrogram-based deep learning for visual pattern recognition, and text-based classification using BERT and other transformer models on transcribed speech. Each branch was designed to capture different aspects of Parkinsonian speech — from physical vocal anomalies to underlying linguistic patterns — thereby offering a comprehensive assessment pipeline.

Extensive experimentation on public datasets demonstrated the effectiveness of the proposed system, with BERT-based models achieving notably high accuracy on transcribed text data, while spectrogram-based CNN models and numerical-feature-based classifiers also produced strong results. The combined findings validate the potential of speech as a non-invasive and accessible biomarker for Parkinson's Disease. Furthermore, the modularity of this approach paves the way for future clinical integration, real-time applications, and remote health monitoring systems aimed at improving early diagnosis and patient outcomes.

LIST OF TABLES

LIST OF FIGURES

1.INTRODUCTION

Neurodegenerative diseases are an escalating global health concern, with Parkinson's Disease (PD) ranking as the second most prevalent. Affecting over 8.5 million individuals worldwide, its rising occurrence—especially in aging populations—underscores the need for effective early detection. Parkinson's Disease manifests through motor symptoms such as tremors and rigidity, alongside non-motor issues like cognitive decline and mood disorders. Speech impairment emerges as one of the earliest and most consistent indicators, often appearing years before diagnosis. However, traditional clinical assessments are subjective and may miss subtle early signs, delaying intervention.This project seeks to address this gap by developing an innovative AI framework that integrates speech-derived features with spectrogram-based deep learning models for non-invasive and sensitive early Parkinson's detection.

1.1. INTRODUCTION

Advancements in artificial intelligence and machine learning offer promising solutions in healthcare, particularly for conditions like Parkinson's Disease. Speech analysis stands out as a diagnostic tool due to its ability to detect neuromotor dysfunction characteristic of PD.

This project focuses on creating a scalable and objective diagnostic approach to complement traditional clinical methods. Speech-based analysis is non-invasive, cost-effective, and suitable for remote healthcare, making it an ideal solution in resource-constrained settings.

Practical applications include early diagnosis, remote monitoring, and therapeutic feedback systems. Advantages include reduced reliance on specialists, wider screening capabilities, and improved patient outcomes through earlier intervention.

1.2. PROBLEM STATEMENT

Despite the widespread occurrence of speech impairments in Parkinson's Disease patients, traditional diagnostic methods often overlook these early indicators due to reliance on subjective clinical evaluation. Furthermore, the integration of textual data, such as transcriptions and linguistic features, remains underexplored as a complementary diagnostic tool.

This project addresses the need for an automated, non-invasive, and comprehensive framework that combines vocal signal analysis with BERT-based textual approaches. By utilizing both numerical features, spectrogram-based deep learning, and text data, it aims to improve early Parkinson's detection and bridge existing gaps in diagnostic capabilities.

1.3. OBJECTIVES

The primary objectives of this project are:

i. Design a non-intrusive monitoring system that integrates speech biomarkers with BERT-based deep learning to facilitate early detection and track the progression of Parkinson's disease.

ii. Improve diagnostic accuracy and accessibility, especially in resource-constrained environments, by developing an AI-powered speech assessment tool.

iii. Provide precise disease progression tracking to enhance patient monitoring and enable timely medical interventions.

1.4.SCOPE OF PROJECT

This project focuses on developing and evaluating an AI-driven system for the early detection of Parkinson's Disease based on speech analysis. The study covers both numerical feature extraction from audio samples and spectrogram-based image processing, applying advanced machine learning and deep learning models for classification. The scope encompasses preprocessing pipelines, modeldevelopment, performance evaluation, and comparative analysis, while leaving room for future expansion toward clinical integration and real-time systems.

2.BACKGROUND DETAIL

2.1 CONCEPTUAL OVERVIEW / LITERATURE REVIEW

Parkinson's Disease (PD) is one of the most pressing neurodegenerative disorders worldwide, affecting approximately 8.5 million people. Forecasts suggest that India is on the cusp of becoming a major hub for PD prevalence in the coming years. While traditionally regarded as an age-related condition, recent clinical evidence highlights a concerning rise in early-onset cases, reshaping our understanding of PD's progression and underlining the urgent need for early detection strategies.

Speech disorders are a prominent early sign of PD, with studies showing that up to 90% of patients experience vocal abnormalities long before motor symptoms become evident. These changes, including hypophonia (soft speech), monotonicity (lack of variation in pitch), imprecise articulation, and irregular rhythm—collectively termed hypokinetic dysarthria—often manifest years before formal diagnosis. However, existing diagnostic practices rely heavily on subjective evaluations, which frequently miss subtle, early-stage speech anomalies essential for timely intervention.

This delay in diagnosis has significant repercussions. The gap between the initial onset of the disease and its clinical identification can span 5 to 10 years, during which critical neuroprotective treatments may have been possible. Delayed diagnosis also imposes a steep economic burden, with annual healthcare costs for advanced PD patients reaching over $25,000 compared to just $2,500 for early-stage management.

Research efforts have largely been split between two main avenues: neuroimaging and speech analysis. Techniques like DaTscan, MRI, and PET scans offer valuable insights into brain changes associated with PD but are costly, require specialized expertise, and are impractical for large-scale or frequent monitoring. By contrast, speech analysis provides a non-invasive alternative, focusing on acoustic and prosodic features such as jitter (frequency variations), shimmer (amplitude fluctuations), harmonic-to-noise ratio, and speech timing metrics. Machine learning models applied to these features have achieved classification accuracies as high as 85-95%.

Recent innovations in computational linguistics and signal processing have expanded the field beyond traditional methodologies. Spectrogram-based analysis enables convolutional neural networks to identify subtle acoustic patterns, while natural language processing applied to speech transcriptions reveals cognitive indicators like simplified syntax, semantic deficits, and pragmatic issues, which may indicate early neurodegenerative changes.

Despite advancements in Parkinson's Disease (PD) research, several critical gaps remain. Many studies focus on isolated analytical pathways, overlooking the potential benefits of integrating complementary approaches. Additionally, research often relies on controlled laboratory speech recordings, which lack ecological validity and fail to represent natural conversational speech. Cross-cultural and multilingual variations in PD speech manifestations are also underexplored, despite evidence suggesting these factors influence vocal changes.

This project addresses these challenges by leveraging a multi-faceted approach that incorporates acoustic, visual, and linguistic data to enhance diagnostic accuracy. By bridging these gaps, it aims to develop a scalable, accessible tool for early PD detection. The ultimate goal is to transform routine speech into a powerful diagnostic medium, democratizing early detection and revolutionizing clinical management across diverse healthcare settings.

3.SYSTEM DESIGN & METHODOLOGY

3.1. SYSTEM ARCHITECTURE

The system architecture for this project is designed to integrate multiple analytical pathways, providing a comprehensive framework for early Parkinson's Disease (PD) detection. The architecture incorporates three core components: numerical feature-based analysis, spectrogram-based analysis, and textual analysis. Each component contributes unique insights, ensuring a holistic diagnostic approach.

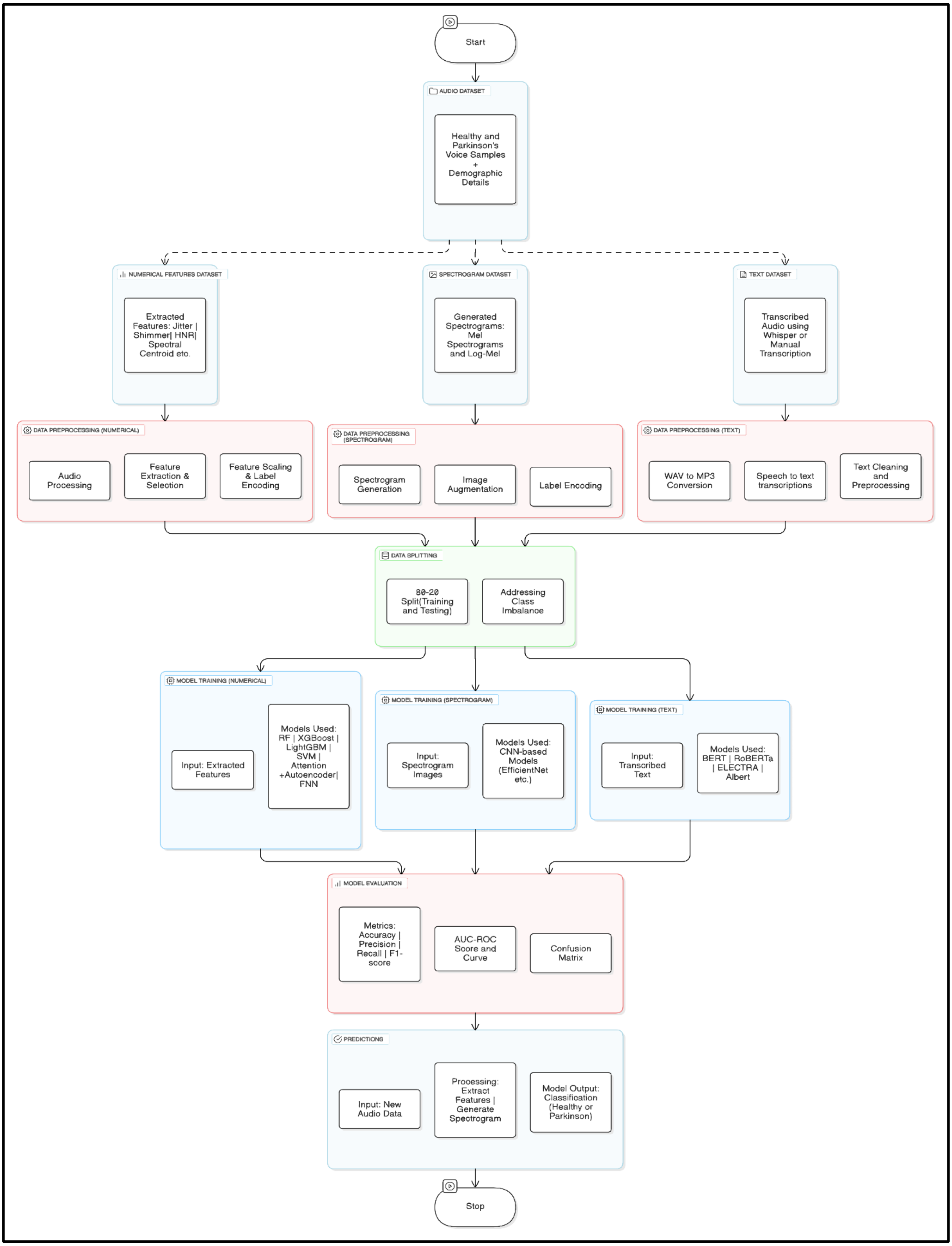


Figure 1:FlowChart: Workflow of the Parkinson's Detection Pipeline

(a) **Numerical Feature-Based Analysis**: Focuses on extracting acoustic and prosodic features, such as jitter and articulation rate, to identify PD-related speech patterns.
(b) **Spectrogram-Based Analysis**: Converts speech signals into spectrograms, enabling CNNs to detect visual patterns linked to neurodegeneration.
(c) **Text-Based Analysis**: Evaluates transcriptions of speech using BERT, highlighting linguistic complexity, coherence, and syntactic structure.

Currently, these approaches are being compared individually to assess their effectiveness, forming the basis for further research and refinement.

3.2 DEVELOPMENT ENVIRONMENT

i. Hardware Infrastructure

(a) Processor: Intel® Core™ i7-11800H (8 cores, 16 threads) with 32GB DDR4 RAM.
(b) GPU: NVIDIA® GeForce® RTX 3050 (16GB VRAM) for accelerated deep learning.
(c) Storage: 1TB NVMe SSD for high-speed data access.

ii. Software Architecture

a) Core Programming Language: Python 3.9 with Anaconda environment management.
b) Audio Processing Framework:
   - Librosa 0.9.2 — feature extraction and visualization.
   - OpenSMILE 3.0 — acoustic feature extraction.
   - Parselmouth 0.4.3 — Praat-based voice analysis.
   - Pydub 0.25.1 — audio manipulation.
c) Deep Learning Frameworks:
   - PyTorch 1.12.0 — core deep learning model development.
   - TensorFlow 2.9.0 — alternative for model experiments.
d) Natural Language Processing Stack:
   - Hugging Face Transformers 4.21.0 — used for BERT, ALBERT, RoBERTa, and ELECTRA.
   - OpenAI Whisper — for speech-to-text transcription.
e) Data Science and Visualization Utilities:
   - Pandas, NumPy — data manipulation.
   - Matplotlib, Seaborn — visualizations.
f) Machine Learning Toolkit:
   - Scikit-learn — preprocessing and evaluation metrics.
   - XGBoost, LightGBM — gradient boosting experiments.

3.3. METHODOLOGY: ALGORITHM / PROCEDURES

The proposed system adopts a robust multi-branch pipeline designed to leverage speech signal analysis for the early-stage detection of Parkinson's Disease. The methodology integrates three complementary perspectives: **Numerical Feature-Based Learning**, **Spectrogram-Based Deep Learning**, and **Text-Based Transformer Models**. This multi-modal strategy aims to capture both the acoustic impairments and the linguistic anomalies commonly associated with Parkinsonian speech.

(i) Dataset Preparation

(a) **Numerical Feature Dataset**:

- Speech recordings were obtained from the Figshare Parkinson's Voice Dataset, consisting of a total of 81 audio samples — 41 from Healthy Controls (HC) and 40 from Parkinson's Disease (PD) patients.
- Alongside the audio data, a metadata file was provided, containing corresponding demographic details such as age, sex, and class label (Healthy or Parkinson's Disease) for each recording.

(b) **Spectrogram Dataset:** The same audio recordings were transformed into Mel Spectrograms and Log-Mel Spectrograms to enable deep learning-based image classification.

(c) **Text Dataset (Transformer-Based Models):**
Since the Figshare dataset primarily consisted of prolonged vowel sounds rather than natural speech, it was unsuitable for meaningful speech-to-text transcription. Therefore, a new dataset was created for the text-based classification approach by compiling:

- **Healthy Controls:** Speech samples sourced from the **Mozilla Common Voice Dataset**.
- **Parkinson's Disease Patients:** Speech samples sourced from the **SJTU Parkinson's Speech Dataset** (GitHub source).

(ii) Data Preprocessing

Audio Preprocessing: Each raw audio file was standardized through normalization, silence trimming, and resampling to maintain consistency across samples. Waveform and spectrogram visualizations were generated to validate audio integrity prior to feature extraction.

(iii) Feature Extraction & Selection (Numerical Approach)

(a) Acoustic and prosodic features were extracted using Librosa, OpenSMILE, and Parselmouth, yielding a total of 67 features.

(b) A subset of 24 statistically significant features was selected based on their relevance to Parkinsonian speech characteristics. These included:

- Perturbation Metrics: Jitter, Shimmer, Harmonic-to-Noise Ratio (HNR).
- Spectral Attributes: Spectral Centroid, MFCC.
- Prosodic Elements: Speaking Rate, Intensity, Duration.
- Pitch and Voice Quality Features.

(iv) Data Preparation (Numerical Features)

(a) Class labels were encoded: Healthy Controls (HC = 0) and Parkinson's Disease (PD = 1).

(b) Feature values were standardized using Min-Max Scaling and Z-Score Normalization.

(c) An 80:20 Train-Test Split was employed for model evaluation.

(d) SMOTE (Synthetic Minority Over-sampling Technique) was applied to address class imbalance by synthetically generating minority class samples**.**

(v) Spectrogram Generation & Preparation

(a) Raw audio files were converted into Mel Spectrograms and Log-Mel Spectrograms using Librosa.
(b) Spectrograms were exported as .png images to facilitate training using convolutional neural networks.

(vi) Audio-to-Text Conversion (Textual Models)

(a) Audio samples were converted from .wav to .mp3 formats for compatibility with the OpenAI Whisper ASR model.
(b) Whisper was fine-tuned to improve transcription accuracy, especially for the nuances of Parkinsonian dysarthric speech.

(vii) Text Preprocessing

(a) Text transcripts were standardized by:
  - Converting all text to lowercase.
  - Removing punctuation and special characters via regular expressions.
  - Normalizing whitespace for clean tokenization.

(b) Class labels were encoded using LabelEncoder() for compatibility with ML pipelines**.**

(viii) Model Training & Evaluation

(a) Numerical Feature-Based Approach:
  - Random Forest Classifier — hyperparameter tuning was performed via RandomizedSearchCV.
  - Feedforward Neural Network (FNN) — trained on the standardized feature set.

(b) Spectrogram-Based Deep Learning Approach:
  - Deep Convolutional Networks including EfficientNet-B0, InceptionV3, and ResNet50 were trained on spectrogram images for visual pattern recognition.

(c) Text-Based Transformer Approach:
  - Multiple transformer architectures were fine-tuned using Hugging Face Transformers, including:BERT, RoBERTa, ELECTRA and Albert.
  - These models were evaluated for their ability to capture subtle linguistic markers and cognitive impairments in transcribed speech.

(ix) Evaluation Metrics

All models were benchmarked across multiple performance criteria:

(a) Accuracy (Training and Testing Phases)
(b) Confusion Matrix
(c) Precision, Recall and F1-Score
(d) AUC-ROC Curve and Score

4. IMPLEMENTATION AND RESULT

4.1. MODULES / CLASSES OF PROJECT

The system is modularly designed to handle three parallel data streams — numerical features, spectrogram images, and transcribed text — each optimized for machine learning and deep learning pipelines:

MODULE NAME	DESCRIPTION
AudioPreprocessor	Handles raw audio input by applying noise reduction, normalization, silence trimming, and resampling for dataset consistency. Also supports waveform visualization for analysis.
FeatureExtractor	Extracts acoustic and prosodic features using Librosa, OpenSMILE, and Parselmouth. Outputs numerical data for use in machine learning and deep learning models.
SpectrogramGenerator	Converts audio signals into Mel and Log-Mel spectrogram images, enabling image-based classification with models like EfficientNet and ResNet.
TextTranscriber	Uses OpenAI's Whisper ASR to convert audio to text, enabling the application of NLP models for Parkinson's detection.
TextPreprocessor	Cleans transcribed text by applying lowercasing, punctuation removal, and label encoding to prepare input for NLP models like BERT.
ModelTrainer	Manages training for Random Forest, Neural Networks, CNNs (EfficientNet, ResNet, Inception), and transformer models like BERT.
Evaluator	Calculates evaluation metrics including Accuracy, Precision, Recall, F1-Score, and ROC-AUC, and generates performance comparison reports.

Table 1- Overview of NeuroSpeechBERT Project Components

4.2. IMPLEMENTATION DETAIL

(i) **Numerical Feature-Based Model**

The numerical approach was centred on processing handcrafted acoustic and prosodic features extracted from audio recordings.

(a) **Data Preprocessing**

- 24 relevant features were extracted, including Jitter, Shimmer, Spectral Centroid, and others.
- Features were standardized using z-score normalization.
- Labels were encoded using LabelEncoder (0 for Healthy, 1 for Parkinson's).

(b) **Dataset Handling**

- The dataset was split into 80% training and 20% testing using stratified sampling for balanced class distribution.
- SMOTE was applied to the training set to address class imbalance by generating synthetic samples for the minority class.

(c) **Model Architecture**

For classification, both a Feedforward Neural Network (FNN) and a Random Forest model were employed.

- The FNN architecture consisted of an input layer with 24 features, two hidden layers (128 neurons each) with ReLU activation and a dropout rate of 0.3, and an output layer with a single neuron and sigmoid activation for binary classification.
- In parallel, a Random Forest classifier was optimized using Randomized Search and evaluated as a baseline ensemble learning method.

(d) **Training Configuration:**

- Optimizer: Adam (learning rate: 0.001) | Loss function: Binary Cross-Entropy.
- Epochs: 50, with early stopping based on validation loss to avoid overfitting.

(ii) **Spectrogram-Based Deep Learning**

This approach leveraged the visual representation of speech signals through spectrograms for image-based classification.

(a) **Data Preprocessing**

- Waveform visualization to inspect signal quality.
- Spectrogram images were generated using Librosa to capture frequency and temporal information.

(b) **Model Architectures**

- **EfficientNet**: Lightweight yet powerful model for spatial feature extraction.
- **ResNet50**: Deep residual network to overcome vanishing gradient issues.
- **InceptionV3**: Multi-scale feature extraction through parallel convolution filters.

(c) **Training Parameters**

- Epochs: 10 | Batch size: 32
- Optimizer: Adam (lr: 0.0001) | Loss: Categorical Cross-Entropy
- Learning Rate Scheduler: ReduceLROnPlateau based on validation loss.
- Dataset Split: ~70-80% training / ~20-30% validation with early stopping.

(iii) **Text-Based Transformer Model**

The text-based approach converted transcribed speech samples into input for pre-trained language models.

(a) **Data Preprocessing**

- Speech samples were transcribed and cleaned (lowercased, punctuation removed).
- Labels were encoded for binary classification.
- The dataset was split into 80% training and 20% testing.
- Text was tokenized using each model's tokenizer to generate input IDs and attention masks.

(b) **Model Application**

- Multiple Transformer models were fine-tuned:
  - ✓ BERT (bert-base-uncased)
  - ✓ Roberta (roberta-base)
  - ✓ Electra (google/electra-base-discriminator)
  - ✓ Albert (albert-base-v2)
- Each model was extended with a fully connected classification head.
- Optimized using AdamW and a learning rate scheduler.
- Epochs: 5, using binary cross-entropy loss.

4.3. RESULTS AND DISCUSSION

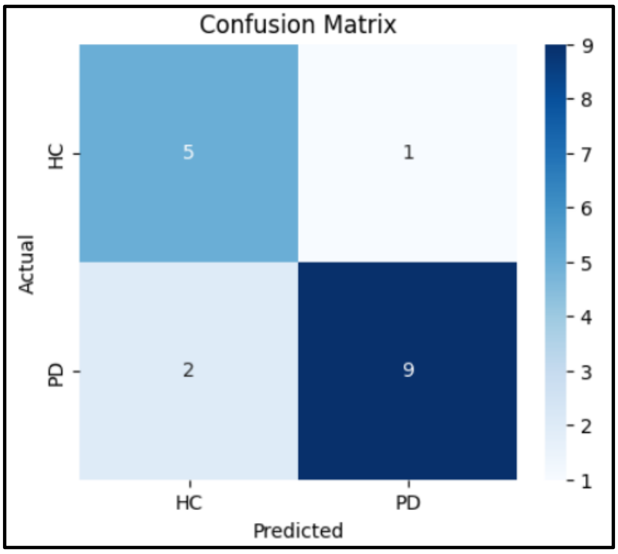
After training and testing on the respective datasets, the following performance was observed:

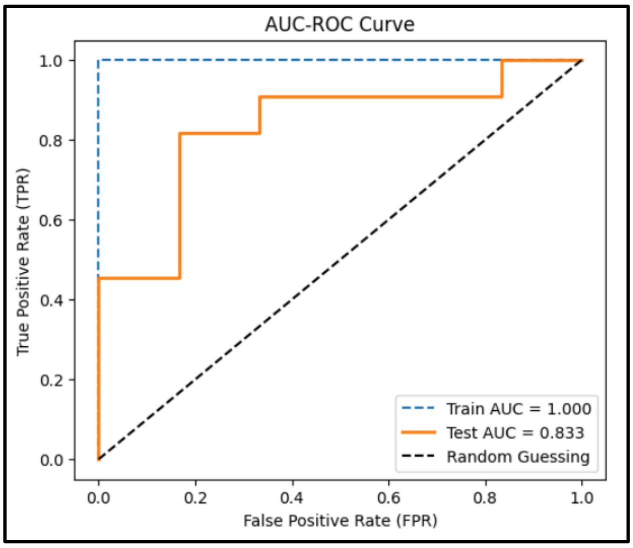
(a) Numerical feature-based

Model	Training accuracy	Testing accuracy	Precision	Recall	F1- Score	AUC-ROC Score
Random Forest	96.88 %	82.35%	0.81	0.83	0.81	0.74
Feedforward Neural Network (FNN) Classifier	85.71%	70.59%	0.8	0.73	0.76	0.74

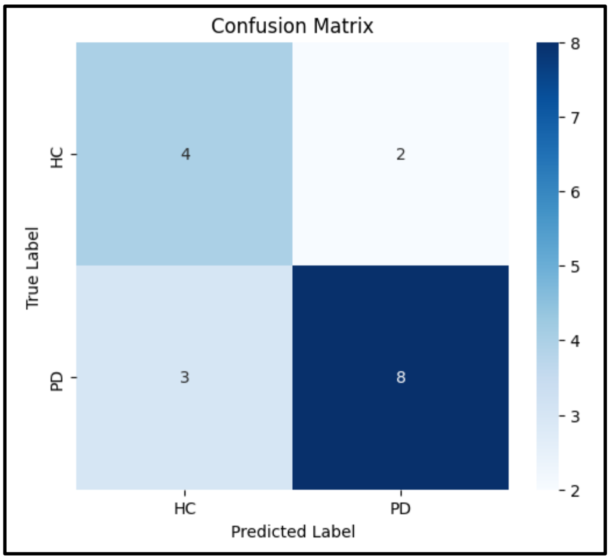
Table -2: Performance of Numerical Feature-Based Models

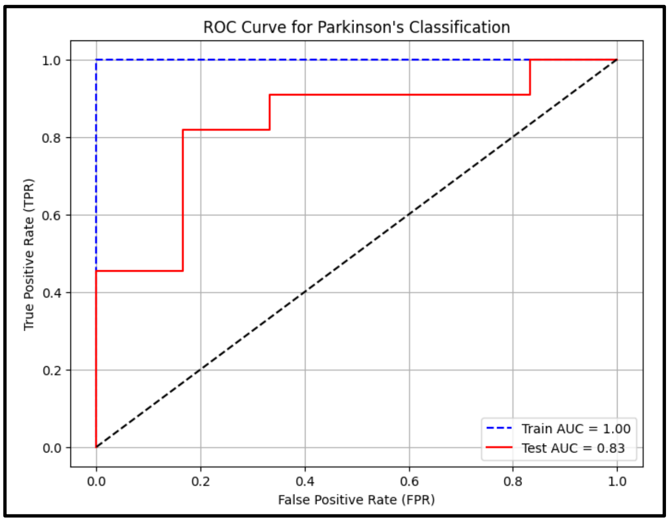
(i) Random Forest





(ii) Feedforward Neural Network (FNN) Classifier



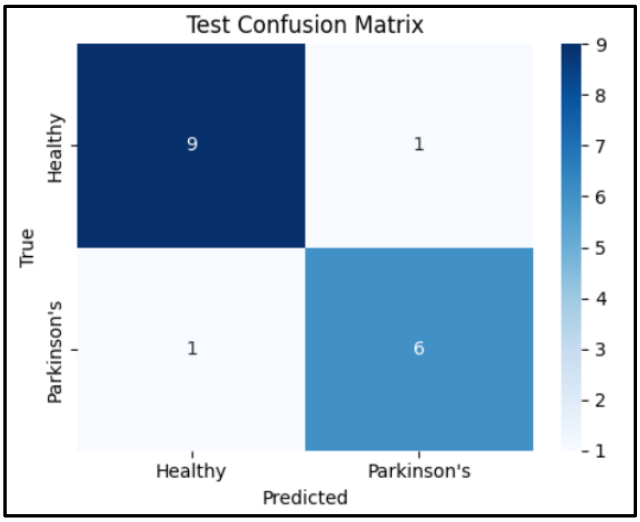


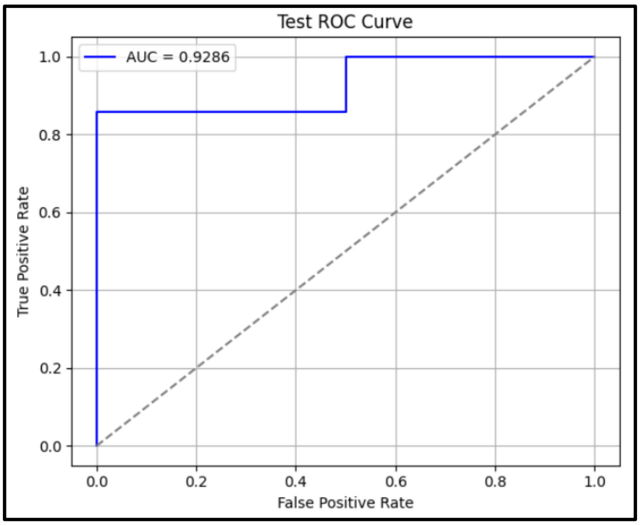
(b) Spectrogram-Based Model

Model	Training accuracy	Testing accuracy	Precision	Recall	F1- Score	AUC-ROC Score
EfficientNet-B0	93.75%	88.24%	0.88	0.88	0.88	0.9286
ResNet -50	79.69%	76.47%	0.76	0.76	0.76	0.7778
Inception -V3	78.12%	70.59%	0.83	0.64	0.62	0.6143

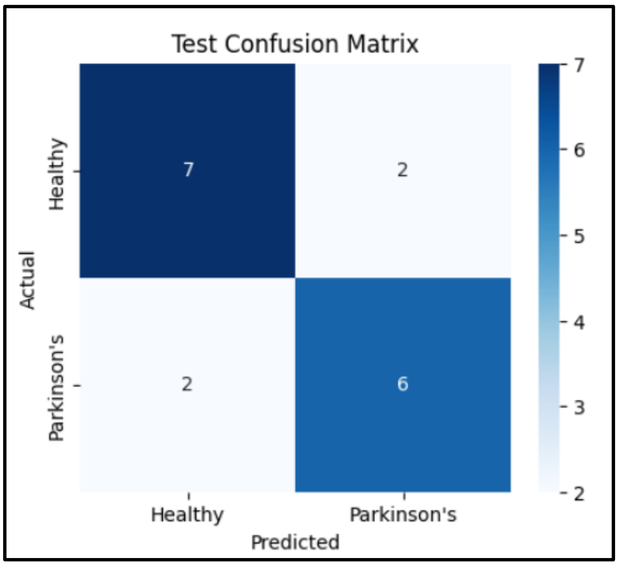
Table -2: Performance of Spectrogram Feature-Based Models

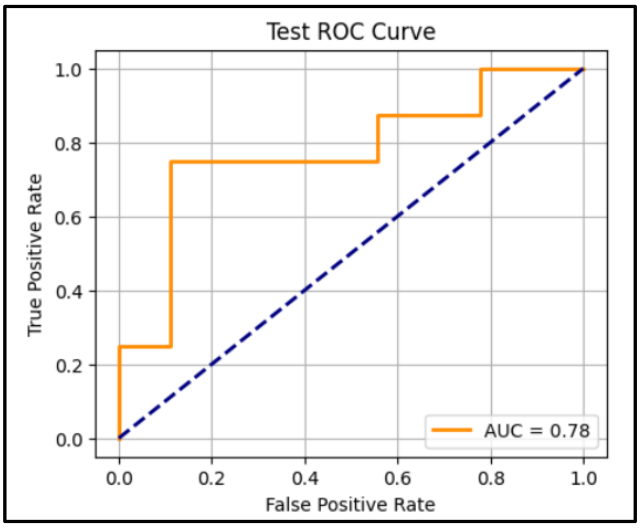
(i) EfficientNet-B0



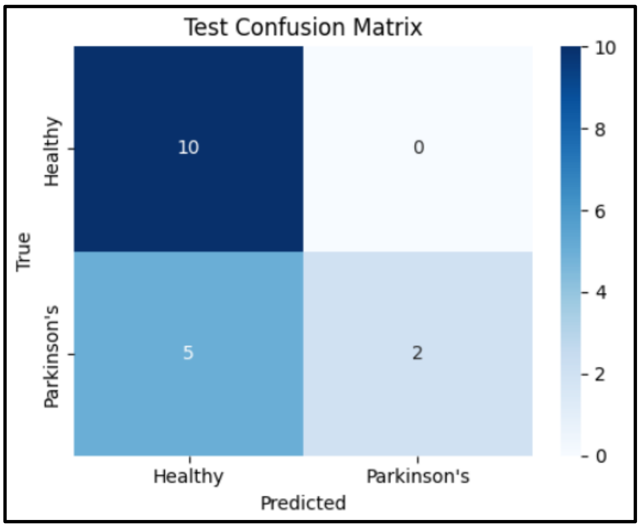


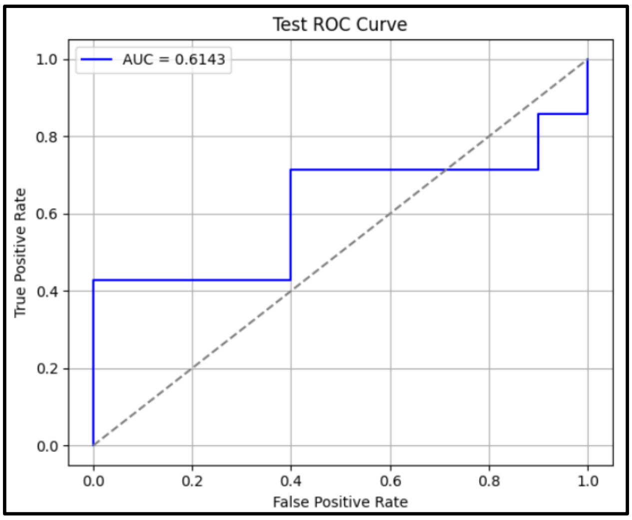
(ii) ResNet-50





(iii) Inception -V3



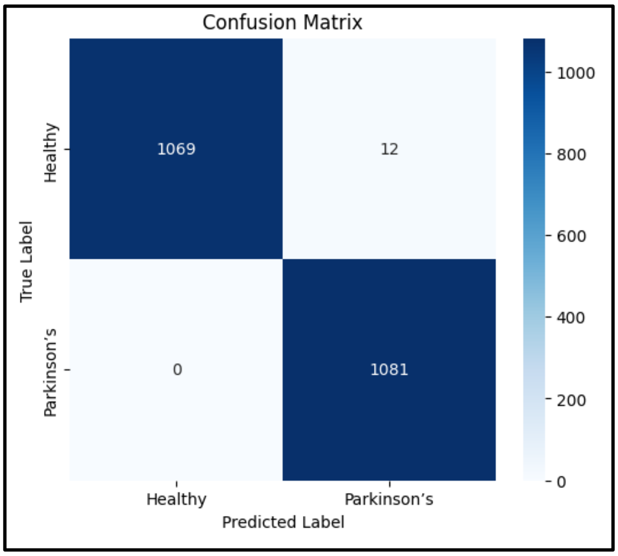


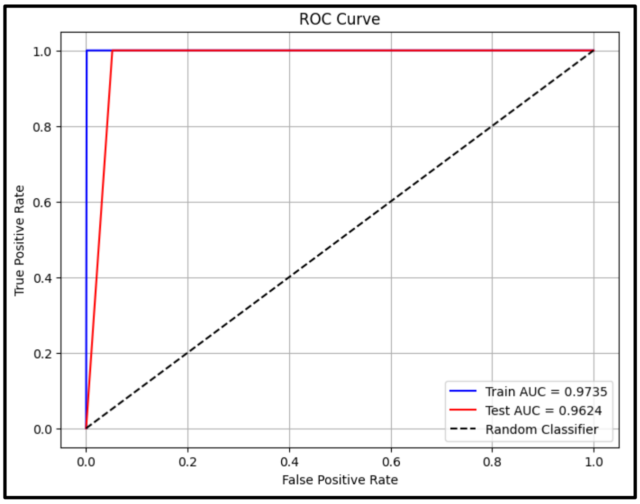
(c) Speech-to-Text + Textual BERT

Model	Training accuracy	Testing accuracy	Precision	Recall	F1- Score	AUC-ROC Score
BERT	99%	97.4%	0.98	1.0	0.99	0.9624
Roberta	99.19%	97.23%	0.97	0.97	0.97	0.9723
Electra	99.65%	98.85%	0.99	0.99	0.99	0.9970
Alberta	99.13%	97.46%	0.97	0.98	0.97	0.9992

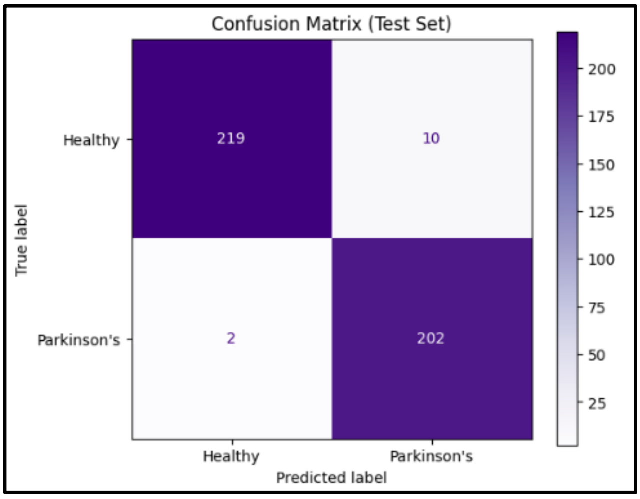
Table -3: Performance of Text-Based Models

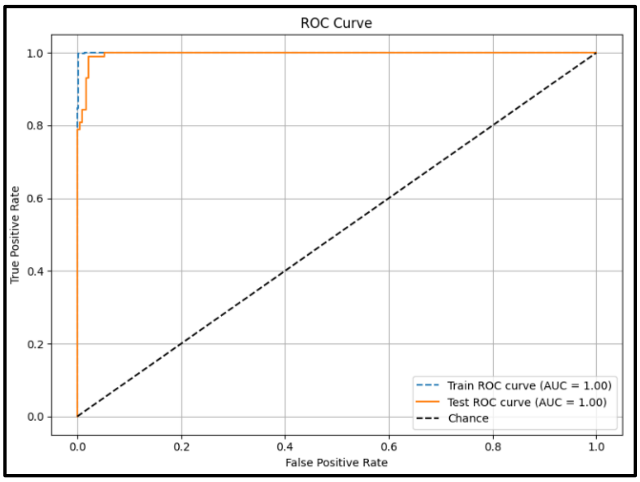
(i) BERT



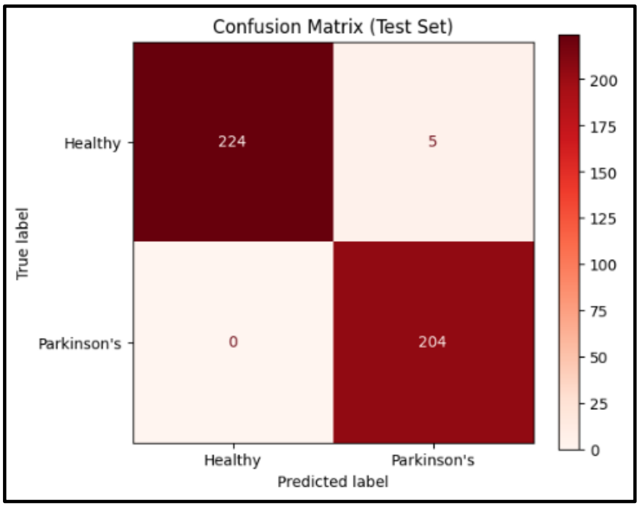


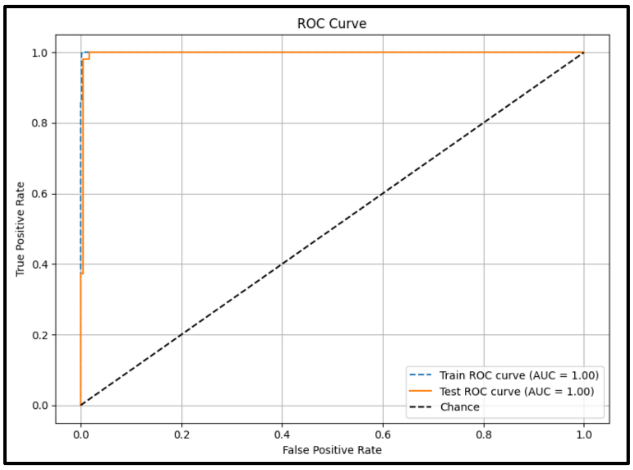
(ii) Roberta



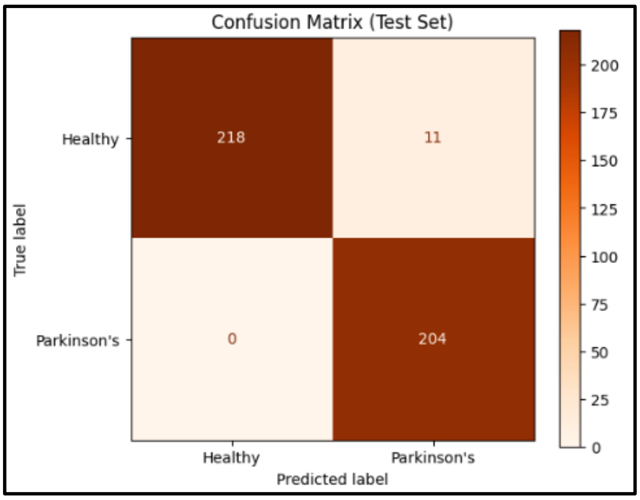


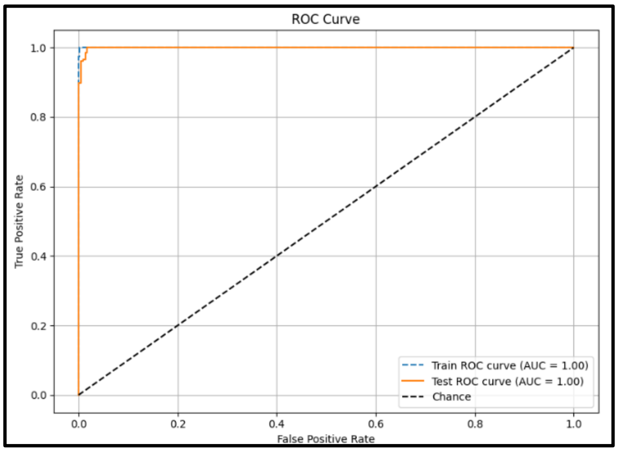
(iii)Electra





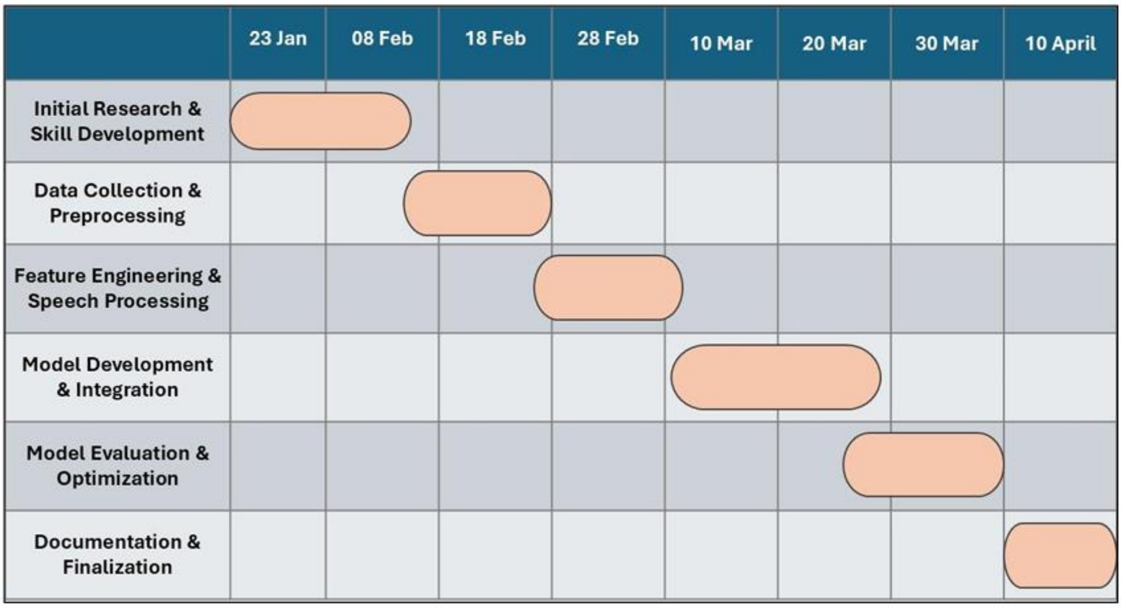
(*iv)Albert*





4.4 MONTH WISE PLAN OF WORK (PROGRESS CHART/TIMELINE CHART)

NeuroSpeechBERT Project Timeline



5. CONCLUSION AND FUTURE PLAN

5.1 CONCLUSION

This research developed an integrated framework for Parkinson's Disease detection via speech analysis, combining numerical features, spectrogram-based methods, and transformer-based NLP models. Notably, the text-based approach using BERT and its related models demonstrated strong diagnostic potential, highlighting how linguistic patterns can complement traditional acoustic features in identifying neurological changes. The study reinforces the value of speech as a non-invasive biomarker, offering a practical and scalable foundation for early PD detection and accessible screening.

5.2 FUTURE DIRECTIONS

Future work will focus on:

- Longitudinal Monitoring: Tracking disease progression through speech changes.
- Multimodal Integration: Combining speech with other digital biomarkers (e.g., gait, typing).
- Cross-Linguistic Adaptation: Expanding support beyond English.
- Clinical Integration: Embedding the framework into real-world diagnostic workflows.

These steps aim to enhance early diagnosis, improve patient outcomes, and broaden the clinical relevance of speech-based Parkinson's assessment.

REFERENCES

- Alalayah KM, Senan EM, Atlam HF, Ahmed IA, Shatnawi HSA. Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method. Diagnostics (Basel). 2023 May 31;13(11):1924. doi: 10.3390/diagnostics13111924. PMID: 37296776; PMCID: PMC10253064.
- A. Nogales, Á. J. García-Tejedor, A. M. Maitín, A. Pérez-Morales, M. D. D. Castillo and J. P. Romero, "BERT Learns From Electroencephalograms About Parkinson's Disease: Transformer-Based Models for Aid Diagnosis," in IEEE Access, vol. 10, pp. 101672-101682, 2022, doi: 10.1109/ACCESS.2022.3201843.
- Escobar-Grisales D, Ríos-Urrego CD, Orozco-Arroyave JR. Deep Learning and Artificial Intelligence Applied to Model Speech and Language in Parkinson's Disease. Diagnostics (Basel). 2023 Jun 25;13(13):2163. doi: 10.3390/diagnostics13132163. PMID: 37443557; PMCID: PMC10340628.
- C. Quan, K. Ren, Z. Luo, Z. Chen, and Y. Ling, "End-to-end deep learning approach for Parkinson's disease detection from speech signals," Biocybernetics and Biomedical Engineering, vol. 42, no. 2, pp. 556–574, 2022. doi: 10.1016/j.bbe.2022.04.002.
- M. R. Salmanpour et al., "Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 206, p. 106131, 2021.
- A. Favaro et al., "Interpretable speech features vs. DNN embeddings: What to use in the automatic assessment of Parkinson's disease in multi-lingual scenarios," *Computers in Biology and Medicine*, vol. 166, p. 107559, 2023
- K. Wu et al., "Learning acoustic features to detect Parkinson's disease," *Neurocomputing*, vol. 318, pp. 102–108, 2018.
- https://github.com/SJTU-YONGFU-RESEARCH-GRP/Parkinson-Patient-Speech-Dataset/
- https://figshare.com/articles/dataset/Voice_Samples_for_Patients_with_Parkinson_s_Disease_and_Healthy_Controls/23849127