A Comprehensive Hybrid Framework for Parkinson's Disease Detection: Integrating Handcraft Features along with Deep Learning-based Feature Extraction with Variational Autoencoder and Traditional Machine Learning Techniques for Classification

by

Md. Iftiajul Alam
20301201
Faria Islam Laiba
22241177
Tahiatun Nazi
20301008
Shirsadip Choudhury
20101028

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2024

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Shirsadip Choudhury 20101028	Tahiatun Nazi 20301008
Faria Islam Laiba 22241177	Md. Iftiajul Alam 20301201

Approval

The thesis titled "A Comprehensive Hybrid Framework for Parkinson's Disease Detection: Integrating Handcraft Features along with Deep Learning-based Feature Extraction with Variational Autoencoder and Traditional Machine Learning Techniques for Classification." submitted by

1. Md. Iftiajul Alam (20301201)
2. Faria Islam Laiba (22241177)
3. Tahiatun Nazi (20301008)
4. Shirsadip Choudhury (20101028)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on October 22, 2024.

Examining Committee:

Supervisor:
(Member)

Md.Golam Rabiul ALam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Rafeed Rahman
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md.Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Neurodegenerative disorders, such as Parkinson's disease, present a significant medical challenge, necessitating innovative approaches for detection. This thesis introduces a comprehensive hybrid framework that combines handcrafted features and deep learning techniques to improve the accuracy of Parkinson's disease detection. The approach leverages pre-trained convolutional neural networks (CNNs) such as VGG16, MobileNet, and EfficientNet, ResNet to extract features of mel-spectrograms generated from the voice samples. A second contribution is the extraction of handcrafted features from the raw audio data. The features extracted are encoded using a Variational Autoencoder (VAE), which further reduces the dimension and integrated them to further train the machine learning algorithms such as Random Forest Classifier (RFC), K-Nearest Neighbour (KNN), Logistic Regression (LR), Support Vector Machine (SVM), and XGBoost to differentiate. To achieve this, we leveraged the combined strengths of these models by integrating both handcrafted and deep learning features to construct a highly optimized and effective classification model using a hybrid approach that highlights the potential of feature extraction techniques and advanced machine learning algorithms for improving the detection and diagnosis of Parkinson's disease and facilitating more progress in computational healthcare and early stage diagnostics.

Key Words: Parkinson's disease detection, Variational Auto Encoder, Hybrid Feature Extraction, Speech analysis, human voice.

Dedication

We truly appreciate the aid and support provided to us in accomplishing our thesis by the individuals and organizations named below.

First and foremost, we want to express our gratitude to our thesis supervisor Md. Golam Rabiul Alam and Rafeed Rahman, our co-advisor, for their unflagging support, astute criticism, and unending patience throughout this research trip. Your knowledge and guidance have been very helpful in determining the course of this study.
We would like to show our appreciation to the teachers and staff at Brac University for providing the tools, spaces, and academic setting that made my research activities possible.

Without the collaboration and assistance of these people and organizations, this thesis would not have been possible. We appreciate your participation in this scholarly project.

Md. Iftiajul Alam

Faria Islam Laiba

Tahiatun Nazi

Shirsadip Choudhury

Table of Contents

List of Figures

Chapter 1

Introduction

A group of diseases known as neurodegenerative diseases (NDD) harm the brain's processes and create a variety of symptoms, culminating in a decrease in cognitive and physical skills (Athisakthi & Rani, 2017)[1].The central nervous system (CNS), notably the brain and spinal cord, is impacted by NDD, which progressively grows worse over time. Parkinson's, Huntington's, and Alzheimer's diseases are three of the most well-known neurodegenerative conditions.These disorders, which are characterized by the progressive degeneration or loss of neurons in the body's central nervous system, primarily in the brain, cause cognitive, motor, and sensory capacities to gradually decline. Parkinson's disease which is the 2nd more probable disease has a prdominance that ranges from 100 to 200 per 100,000 people (Mengarelli et al., 2022)[2]. Hypokinetic movements, tremor, stiffness, stopping of locomotion, and general motor instability come from PD's alteration of the basal ganglia's output. Parkinson's disease (PD) is a major central nervous system degenerative illness that primarily affects the elderly population globally. It can cause tremors, limb rigidity, and balance problems, among other symptoms. Movement-related symptoms and non-motor symptoms are the two fundamental groups of symptoms. It is possible for non-motor symptoms and cognitive decline, to have a greater effect than motor symptoms (Wang et al., 2020)[3].It has been discovered that there is a premotor or prodromal phase to Parkinson's disease (PD). At least five years, and probably up to twenty years, pass between the beginning of neurodegeneration and the onset of the usual clinical motor symptoms (Prashanth et al., 2016)[4]. Early identification, early management, and neuroprotection of Parkinson's disease (PD) are crucial to halting the progression of the disorder and enabling medical professionals in administering suitable treatment (Yang et al., 2021)[5]. This research report uses a gradient boosting technique in conjunction with a number of supervised learning models to predict neurodegenerative disease, with an emphasis on Parkinson's disease. The primary goal was to compare the performance displayed by each model that was employed.The goal of the research was to improve the ability to recognize and classify cases of Parkinson's disease and other neurodegenerative disorders using these machine learning techniques. With relation to the prediction of neurodegenerative illnesses, this strategy made it easier to gain insightful knowledge about the relative efficacy of different predictive models.

1.1 Research Problem

Parkinson's disease (PD) is a neurodegenerative disorder that impacts the patient's movement and speech and the early diagnosis is necessary to slow disease progression. Traditional diagnosis is often based on motor symptoms, which are usually invisible until later stages of the disease. But early vocal impairments can go hand in hand with non motor symptoms and provide an early opportunity for early detection. Machine learning (ML) and deep learning (DL) models have been recently investigated for analysing vocal changes in patients and exploiting these early signs for diagnosis. For example, deep learning models using VGG-16 and VGG-19 have shown the possibility of using image based data (e.g., spiral and wave drawings) to predict PD and it is assumed that this same success might be obtained by applying these models to voice data (Mathkunti et al., 2024)[6].

PD diagnosis is clinically appealing due to the non-invasiveness of data collection and the potential of early detection; voice analysis has developed into an area of interest. In recent studies, various machine learning approaches are compared with deep learning models (CNNs) to see how effectively they can learn from processed voice data. The results show that the deep learning models ought to surpass typical ML models with regard to feature extraction and classification tasks because deep learning models are capable of capturing intricate patterns from mel-spectrograms. However, model overfitting and lack of generalizability are caused by the variability of the voice characteristics across patients (Costantini et al., 2023)[7]. We attempt to address these issues in our study, by combining handcrafted acoustic features together with the features learnt from deep learning to create a highly robust and interpretable model.

One of the problem of using robust the PD diagnostic tool is the tuning of the deep learning models. The improvement of model performance rely on recent advancements on the usage of metaheuristic algorithms, i.e., to choose optimal hyperparamteers and eliminate feature redundancy. For instance, Majhi et al. (2024)[8] have shown that metaheuristic enhanced deep learning could greatly enhance the classification accuracy for PD. All this shows why we need to carefully select and optimize features in order not to add unnecessary complexity in the model. Similarly in our research we employ such techniques as feature selection based on Principal Component Analysis (PCA), and Variance Thresholding to decrease dimensionality at the same time while maintaining high classification accuracy.

Other than feature optimization, philtre based feature selection techniques are also put to use in handling high dimensional data from deep learning models. In PD classification, to improve the speed and accuracy of the model to some extent, Gunduz (2021)[9] proposed a feature philtre based approach for dimensionality reduction. We apply these methods to our research, producing correlation heatmaps to identify and exclude features with high levels of correlation, refines the feature set and thus the model. From this we are able to build a diagnostic model that is both accurate and computationally cheap — suitable for real time clinical applications.

Finally, Variational Autoencoders (VAE) are further used in managing large datasets in PD research. We present these methods which help transform high-dimensional features into a lower dimensional latent space but preserving essential information for classification.Gunduz showed that using VAEs with PD voice data increased model performance while decreasing computational cost[9].In our study we apply deep learning feature extraction techniques in conjunction with VAEs, to avoid losing accuracy on large, complex datasets. We seek to design a scalable solution of early PD diagnosis by combining handcrafted and deep learning features with state-of-the art dimensionality reduction methods.

1.2 Research Contribution

In this work, we design and analyze a hybrid framework for Parkinson's disease detection.In particular, we propose a feature fusion mechanism that leverages the strengths of both types, aiming to improve diagnostic accuracy. The main contributions of the paper are summarized as follows:
1. We design a hybrid feature extraction model to exploit mix of handcrafted acoustic features and deep learning extracted features from mel-spectrograms towards improved diagnosis of Parkinson's Disease (PD).
2. We identified The limitations of the model and the challenges highlighted which include the difference of source characteristics, along with suggestions for additional improvements particularly managing the high dimensionality and complexity of extracted features.
3. We explored techniques for dimensionality reduction and feature selection such as PCA, variance thresholding, and philtre based selection to reduce redundancy and enhance classification accuracy while minimising computational complexity and redundancy in feature sets.
5. We Assess the potential of using pre-trained CNN models (e.g. VGG16, MobileNetV2, EfficientNet, ResNet) for voice based diagnosis in real applications, focusing on their generalizability across multiple patient populations and datasets in the context of practical feasibility and scaling.
6. We evaluate, the broader impact of combining handcrafted and deep learning based features examining, the feasibility of this approach to augment the interpretability and robustness of PD diagnostic models is.

1.3 Thesis Organization

- The research problem, research contribution is discussed in the first chapter.
- Then the literature review in discussed in the second chapter.
- The third chapter describes the dataset, data pre-processing.
- Chapter four presents the methodology, model architecture
- Lastly, chapter 5 discuss about the results and discussion and model-wise comparison.
- Chapter six concludes the research and outlines future direction.

Chapter 2

Literature Review

There are several related works in different research publications. We have studied those related works and summarized those below.

A comprehensive examination of the application of time-dependent spectrum characteristics in the interpretation of gait data for the diagnosis of neurodegenerative diseases (NDDs) is given in the work by[2].Monitoring and evaluation of non-motor diseases (NDDs) are crucial because they offer critical information on motor function and the progression of the disease. The authors examined the prospective uses of time-dependent spectral (PSDTD) and time- dependent (TD) features. In the earlier research, the effectiveness of using TD metrics to find non-diagnostic disorders (NDDs) in gait data has been shown and this study is built on the earlier research. A few fundamental TD traits may enable healthy persons to differentiate between patients with atypical parkinsonism (AS), Parkinson's disease (PD), Huntington's disease (HD), and other NDDs. The "Du's feature set (DUFS)" in particular showed promise, correctly classifying CN-NDD, CN-AS, CN-HD, and CN-PD with an astounding 95.16% accuracy. The study shows that when combined with TD measurements, these spectral characteristics greatly improve classification accuracy. The fact that the k-nearest neighbor (kNN) method achieved 100% accuracy for all binary classifications was a significant finding and demonstrated the method's power. The usefulness of this strategy is demonstrated by the proposed two-step diagnosis pathway (DP1), which makes it possible to quickly identify healthy people and particular NDD types.

When it comes to TBI, Harris et al. (2023)[10] explore the value of early diagnosis, the limitations of present approaches, and the possibility of ocular biomarkers and Raman spectroscopy as fast and precise point-of-care diagnostics techniques.The study demonstrates some of the shortcomings of existing approaches, including the Glasgow Coma Scale, intracranial pressure monitoring and neuroimaging (CT/MRI).The next section of the study describes the eye's reaction to TBI and neurodegeneration, including aberrant eye movements, axonal damage, cerebrospinal fluid (CSF) leaks, blood vessel structural changes and retinal thinning. The report spends a lot of time discussing the biomarkers linked to neuronal injury. Total tau (t-tau),ubiquitin C-terminal hydrolase-L1 (UCHL1), glial fibrillary acidic protein (GFAP), neuron-specific enolase (NSE) and II-spectrin disintregrated products are some of the substances that make up this group. The paper mentions one of the best techniques for detecting the Biomarkers is Raman Spectroscopy. Among the advantages RS can rapidly and non-invasively assess biofluids for neuro-diagnostics it can also classify head injury severity and monitor tissue biochemistry changes after trauma. The study also talks about the In-Vivo and Ex-Vivo studies in ocular RS and their advantages in this field. Lastly, the study mentions Ongoing research in TBI diagnostics, particularly in the context of RS and other techniques, is expected to yield tangible results in the coming years. The integration of RS into handheld devices may become more widespread, enabling non-invasive and real-time monitoring of TBI in critical settings and the development of portable RS systems for ocular diagnostics holds great potential for point-of-care applications.

This pilot study (Dinesh et al.,2016) [11] presents a unique method for tracking and analyzing motor symptoms in Parkinson's and Huntington's illnesses utilizing small, discrete sensors attached to the body. The paper elucidates specific signals linked to the motor that underscores potential advantages associated with this methodology, with consequential implications for the development of non-invasive, continuous monitoring systems for individuals afflicted by the (NDD) neurodegenerative disorders.These findings from the paper ultimately contributes to the enhancement of diagnostic accuracy in the research field.

Parkinson's disease (PD) affects motor functions and speech, making early diagnosis critical. Machine learning techniques, particularly those using vocal features have gained prominence in PD detection. Early studies used vocal features like shimmer and minimum redundancy maximum relevance (mRMR) for features selection, achieving high sensitivity. Deep learning approaches have also emerged. Karan et al. (2020) [12] used auto encoders to achieve 87% accuracy. Variational auto encoders which help capture a latent features pace have been particularly used for dimensionality reduct noisy data [9]. Hybrid model combining feature based feature selection (Relief, Fisher Score) and VAEs have improved accuracy and reduced dimensionality. Gunduz demonstrated that combining these methods with multi-kernel support vector machines (SVM) achieved 91.6% accuracy, outperforming models without dimensionality reduction [9].

The paper "Contrastive Machine Learning Reveal Parkinson's Disease Specific Features Associated with Disease Severity and Progression" employs a deep learning technique Contrastive Variational Autoencoders to identify neuroanatomical changes specific to Parkinson's disease (PD) (Zheng et al., 2024)[13]. Analyzing MRI data from 932 PD patients and 366 controls, the study uncovers disease specific alterations in the subcortical and temporal brain regions. These alterations correlate with clinical severity, dopamine transporter deficits, neurodegenerative biomarkers, and cerebrospinal fluid proteins, particularly those related to immune function. The study's use of CVAE allows for the isolation of PD-specific neuroanatomical features, offering a more individualized analysis of disease progression. PD specific features so significant correlations with motor and cognitive deterioration as measured by clinical assessment a semen assessment such as the unified Parkinson Disease rating scale. The findings suggest potential early diagnostic markers and support the development of treatments aimed at slowing neurodegeneration.

Recent advancements in Parkinson's disease have increasingly focused on leveraging deep learning models to improve diagnostic accuracy. This study by Manimegalai et al. (2022)[14] employs DenseNet, ResNet, and VGG16 models to analyze hand-drawn images and DaTscan brain images, achieving accuracies of 91% and 95%, respectively. By combining motor symptom data with neuroimaging techniques, their approach addresses the challenge of early PD diagnosis, particularly in identifying subtle degenerative changes in the brain. The integration of convolutional neural networks (CNNs) has proven effective in processing complex image data, offering promising improvements in PD detection.

Chapter 3

Dataset

3.1 Dataset Description

The dataset utilized in this study consists of voice recordings collected from participants during a series of vocal exercises specifically designed to elicit distinct phonetic characteristics. These characteristics are essential for the analysis of acoustic features relevant to the detection of Parkinson's Disease. These tasks encompass a wide range of speech characteristics essential for categorizing PD, such as prolonged phonations of particular vowels, syllable articulation, and readings of texts that are phonetically balanced. The recordings include voice data from 28 Parkinson's disease patients and 22 elderly healthy controls. This diversified dataset provides a baseline for normal qualities that are unaltered by age, allowing for an in depth investigation of how Parkinson's disease affect speech patterns. with the young healthy controls serving as a baseline for normal vocal characteristics unaffected by age or neurological conditions. The dataset contains recordings from two groups: individuals diagnosed with Parkinson's Disease (PD) and healthy control subjects. For each recording, the duration of the audio samples is measured to analyze their temporal characteristics. The summary statistics of the dataset are as follows:

Class	**Num of Samples**	**Mean Duration (s)**	**Median Duration (s)**	**Min Duration (s)**	**Max Duration (s)**
Parkinson	448	22.945338	11.121688	5.50275	149.44325
Healthy	352	25.034026	11.75000	3.37500	250.31250

Table 3.1: Summary of voice sample durations for Parkinson's and healthy subjects.

Filecode encapsulates the distinctive identifiers or labels given to particular voice activities that participants accomplish. Each audio file's code is utilized to identify the kind of voice task that is done. Voice samples from both healthy people and patients with Parkinson's disease are used in this research study. These voice samples are arranged in accordance with the particular voice tasks they were used for. Each process generates a distinct kind of audio sample that can display various vocal traits. File code helps to categorize and comprehend the context of each audio sample. Diverse vocal task can highlight distinct aspects of the voice. For example, reading tasks ("B1", "B2") emphasizes speech rhythm and articulation, ("VA1", "VA2") emphasizes on the phonation of the vowel "a", whereas phonation tests highlight pitch and frequency variation. A more comprehensive understanding of the disease's impact on voice can be obtained by comparing information from various tasks. The dataset contains 16 voice samples for each patient. The FileCode here represents different types of phonation tasks for each voice sample of a patient.

Code	**Task Description**
B1	First reading of a phonetically balanced text
B2	Second reading of a phonetically balanced text
D1	Execution of the syllable "pa" for 5 seconds
D2	Execution of the syllable "ta" for 5 seconds
FB1	Reading of phonetically balanced phrases
FB2	Reading of phonetically balanced words
VA1	2 phonations of the vowel "a"
VA2	Additional phonation of the vowel "a"
VE1	2 phonations of the vowel "e"
VE2	Additional phonation of the vowel "e"
VI1	2 phonations of the vowel "i"
VI2	Additional phonation of the vowel "i"
VO1	2 phonations of the vowel "o"
VO2	Additional phonation of the vowel "o"
VU1	2 phonations of the vowel "u"
VU2	Additional phonation of the vowel "u"
PR1	Reading of another long phonetically balanced text

Table 3.2: Vocal tasks performed by each participant

3.1.1 Audio Visualization

Audio visualization translates sound into visual formats like wave forms , allowing detailed analysis of speech patterns. The figures given presents one of several examples of waveform representations comparing audio samples from both Parkinson's and healthy subjects, each consisting of 16 voice recordings. The upper plot, representing a Parkinson's patient , reveals significant irregularities in amplitude with frequent, erratic fluctuations, indicative of the vocal instability commonly associated with Parkinson's disease. This contrasts with the lower plot, which visualizes the waveform of a healthy subject . Here, the amplitude is more consistently distributed, reflecting stable vocal control and clearer articulation. The wave-form generated by the healthy participants has a balanced structure without the wild changes found in the Parkinson's sample, emphasizing the different acoustic characteristics of the two groups. Among Other examples, this one sheds important light on the possibility of using waveform analysis to identify vocal abnormalities.

The audio file "B1ABNINSAC46F240120171753.wav" has a waveform that corresponds to a patient suffering from parkinson's disease. The waveform shows a succession of peaks and troughs that fluctuate in amplitude over time signifying variations in the intensity of the audio stream. The y-axis indicates the amplitude, while the x-axis measures time in seconds, roughly ranging from 0 to 1 minute and 10 seconds. The fluctuating signal may be caused by the vocal or speech characteristics that are commonly linked with Parkinson's disease, such as tremors or irregular pauses, as the pattern indicates frequent alterations in sound.

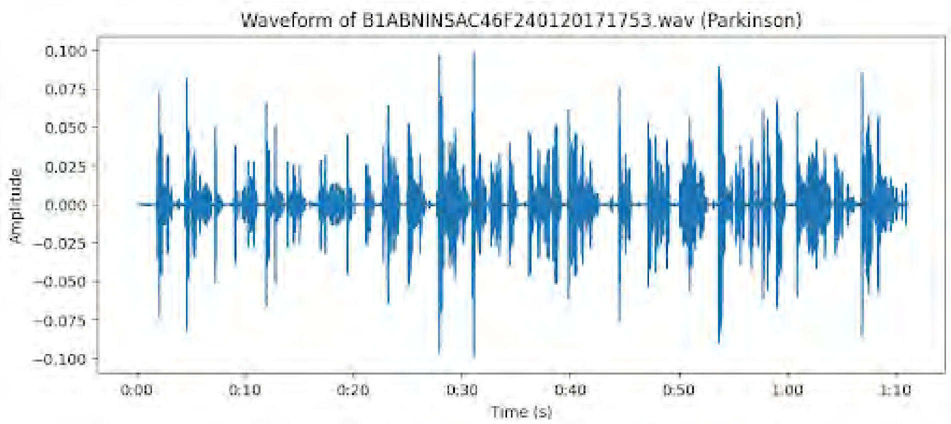

Figure 3.1: Audio Visualization(Parkinson)

The waveform of the file labeled The file named "B1ACNAGRERA49F210320170916.wav" has a waveform that indicates a healthy person. The waveform exhibits a more even amplitude across time, with distinct and largely constant peaks and troughs. The audio lasts for around 50 seconds, and at some moments the amplitude appears somewhat greater than the Parkinson's waveform, indicating stronger vocal strength. A smoother and possibly more rhythmic speech pattern with steady vocal modulations is suggested by this image, which is characteristic of people without speech-related problems.

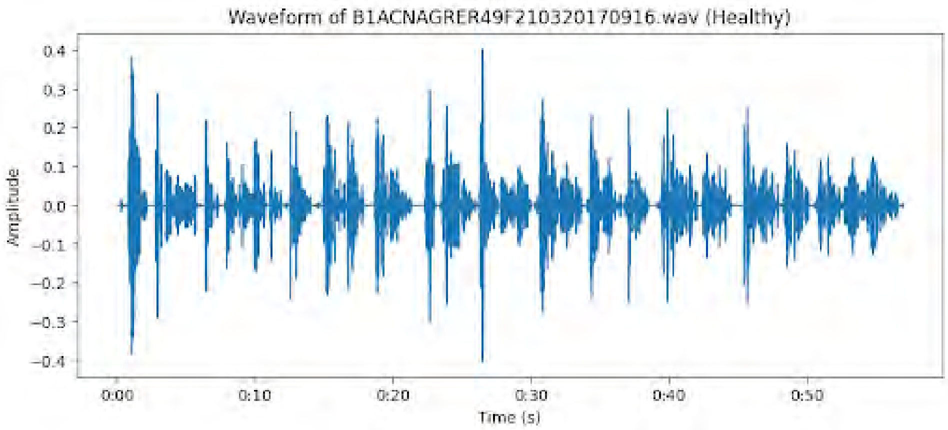

Figure 3.2: Audio Visualization(Healthy)

3.1.2 Mel-spectrogram Visualization

A Mel-spectrogram visualizes sound by mapping the frequency spectrum of an audio signal over time using the Mel scale, which mirrors human auditory perception. Frequencies are extracted via a Fourier Transform and then converted to the Mel scale, highlighting frequencies humans perceive most naturally. This technique captures subtle variations in speech patterns, making it particularly useful for analyzing vocal characteristics, such as those altered by Parkinson's disease.

The Mel-spectrogram of the audio file represents a healthy individual's vocal sample. The mel-spectrogram visualizes the distribution of sound frequencies over time, with the y-axis representing the Mel frequency (in Hz) and the x-axis showing time (in seconds). Higher frequency components, visible above 2048 Hz, are prominently present, with bright areas indicating stronger intensity, as represented by the color scale on the right. The consistent, repeated vertical patterns suggest a regular, rhythmic vocal sound, with amplitude concentrated around the mid-range frequencies. The range of intensity varies between -80 dB and 0 dB, showing dynamic shifts in energy, possibly reflecting normal, unimpeded speech patterns.

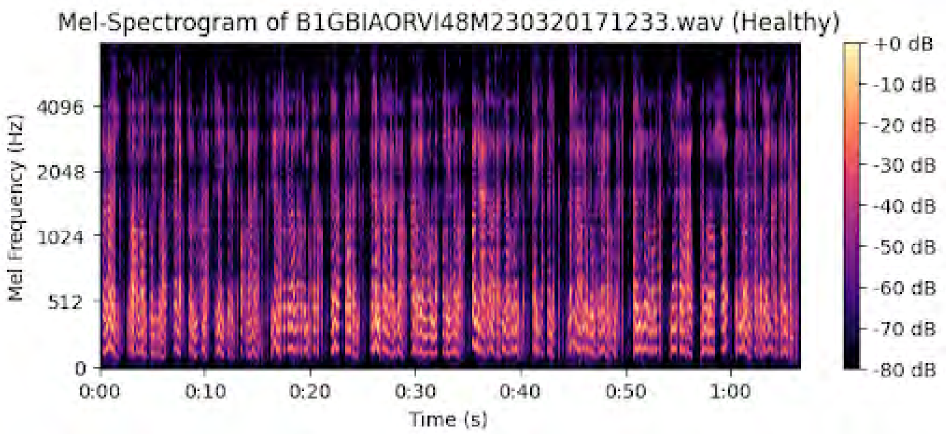

Figure 3.3: Mel spectrogram Visualization (Healthy)

The Mel-spectrogram for corresponding to a Parkinson's patient, shows a different pattern over a longer time span (about 2 minutes and 30 seconds). Although the structure is somewhat similar in terms of frequency distribution, with frequent vertical bands indicating vocalization, the intensity appears more subdued in comparison to the healthy sample. The lower intensities and less bright areas suggest that the vocal signal may be weaker, reflecting common symptoms in Parkinson's such as reduced vocal strength or monotonic speech. The higher frequencies above 2048 Hz also appear less pronounced, indicating possible attenuation or loss of higher-pitched vocal components.

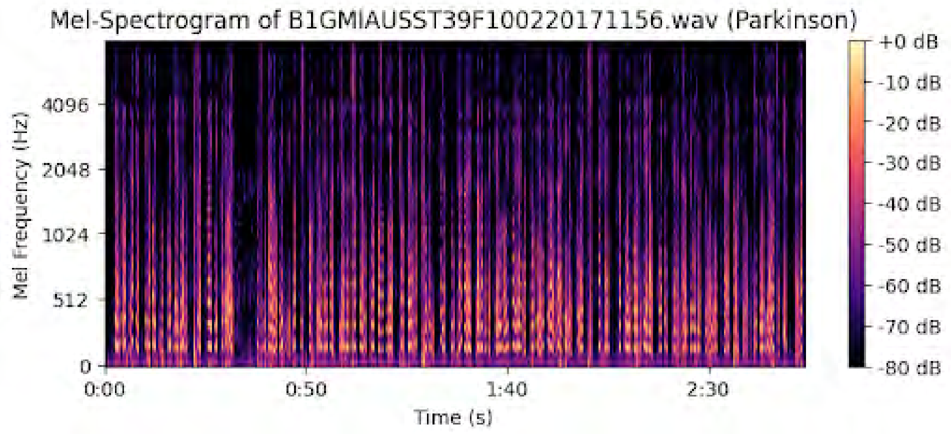

Figure 3.4: Mel spectrogram Visualization(Parkinson)

3.2 Data Pre-processing

3.2.1 Categorization according to the file code

Categorization is a systematic, standardized, and task-specific method of speech data analysis, especially in Parkinson's disease research. By classifying the recordings into discrete folders according to vocal tasks, for example,D1: Holds all recordings in which participants pronounce the letter "pa" for five seconds. By doing this, a single sound is isolated for comparison. D2 records every time a participant pronounces the syllable "ta" for five seconds while concentrating on a different sound. This allows for a more accurate examination of the ways in which Parkinson's disease impacts particular speech functions by separating out distinct vocal performance components. Additionally, by making it easier to compare healthy people and Parkinson's patients who are completing identical tasks, this categorization improves the accuracy and comprehensibility of the findings.

3.2.2 Audio Pre-processing

This research chose not to apply extensive pre-processing to the raw audio samples in the work because doing so would potentially eliminate important vocal characteristics that are critical for accurate Parkinson's disease detection. For instance, removing silent duration or canceling out the silent durations from an audio file, risks losing key information about the natural flow and rhythm of the patient's speech. Parkinson's disease often affects a person's voice in specific ways, such as introducing shakiness or instability in the speech. This vocal shakiness, or tremor, is a crucial marker of the disease, and eliminating silence or other subtle elements reduces the overall effectiveness of our detection model. By keeping the raw audio data as intact as possible, this research aims to preserve all the nuanced vocal patterns that help distinguish between healthy individuals and those with Parkinson's. This approach ensures that the machine learning models can better capture and learn from the natural speech variations associated with Parkinson's disease, leading to potentially higher accuracy and more reliable classification.

3.2.3 Mel-spectrogram Pre-processing

The mel-spectrograms in this study present visual representations of the audio recordings, capturing distinct patterns from both Parkinson's disease patients and healthy control subjects. These mel-spectrograms provide a time-frequency analysis of the audio signals, allowing for the inspection of energy distribution over both time and frequency domains. This technique is particularly effective in analyzing complex sound patterns, such as those associated with human speech.

Feature extraction is a crucial part of this research paper. In particular, the structural patterns and edges in these images are crucial for feature extraction by pre-trained models. Accordingly, the research needs to preserve them. After extensive pre-processing, this specific image may have undergone modifications that change or smooth out the finely detailed spectral edges and nuances. Smoothing, noise reduction can lead to the loss of fine-grained patterns that are essential for the precise feature extraction.

On the contrary, With minimal pre-processing, the raw mel-spectrogram offers a distinct, well detailed frequency spectrum. This makes it perfect for extracting high quality features from all frequency bands since it enables to record more intricate audio features over time. The raw mel-spectrogram depicts a more consistent and prolonged depiction of the harmonic structures and formants, which are critical for differentiating between the voices of people with Parkinson's disease and those in good health.

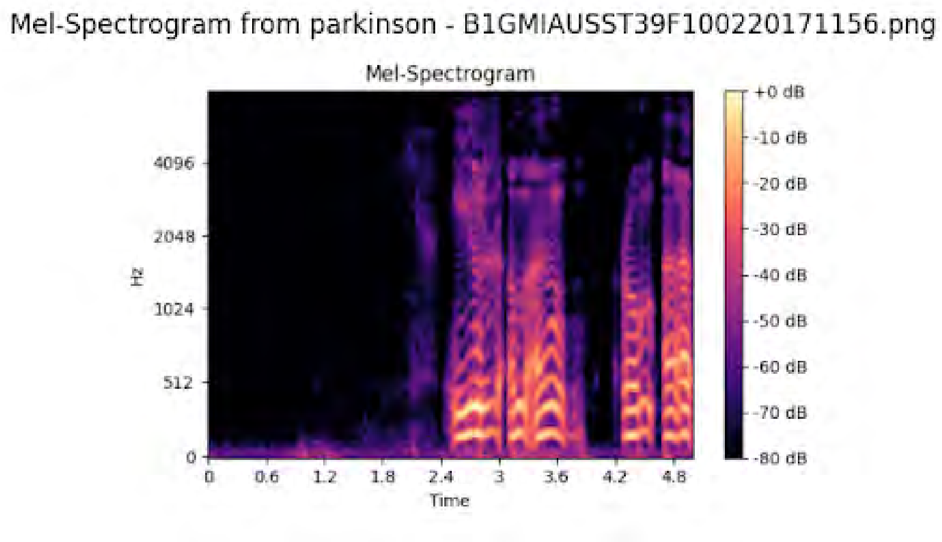

Figure 3.5: Processed Mel-spectrogram

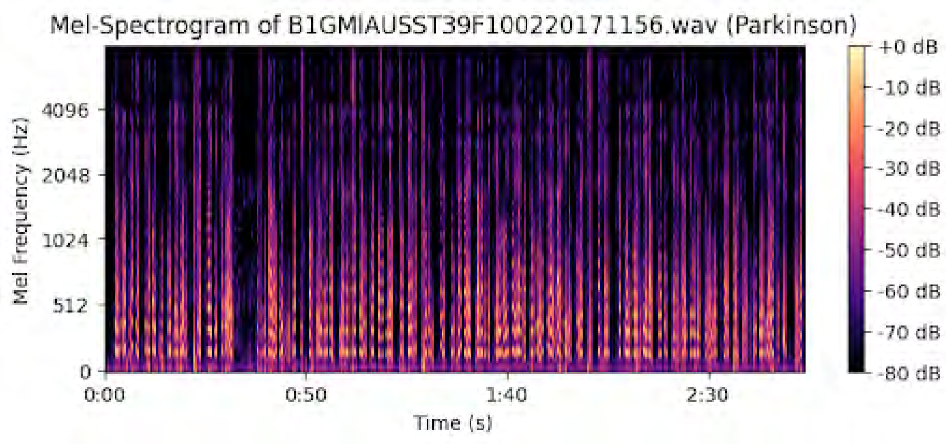

Figure 3.6: Raw Mel-spectrogram

Chapter 4

Methodology

This study proposes a hybrid approach for the classification of Parkinson's Disease that integrates both handcrafted features and deep learning-based features. The process begins with a dataset of voice samples, which is bifurcated into two parallel pipelines. In the first pipeline, the audio data is pre-processed to extract handcrafted features such as MFCCs, pitch, formants, HNR, and spectral features, all of which are critical for capturing key vocal characteristics. In the second pipeline, mel-spectrograms are generated from the voice samples, which are then passed through pre-trained convolutional neural networks (CNNs) including VGG16, MobileNetV2, ResNet50, and EfficientNetB0 to extract deep-level and mid-level features.

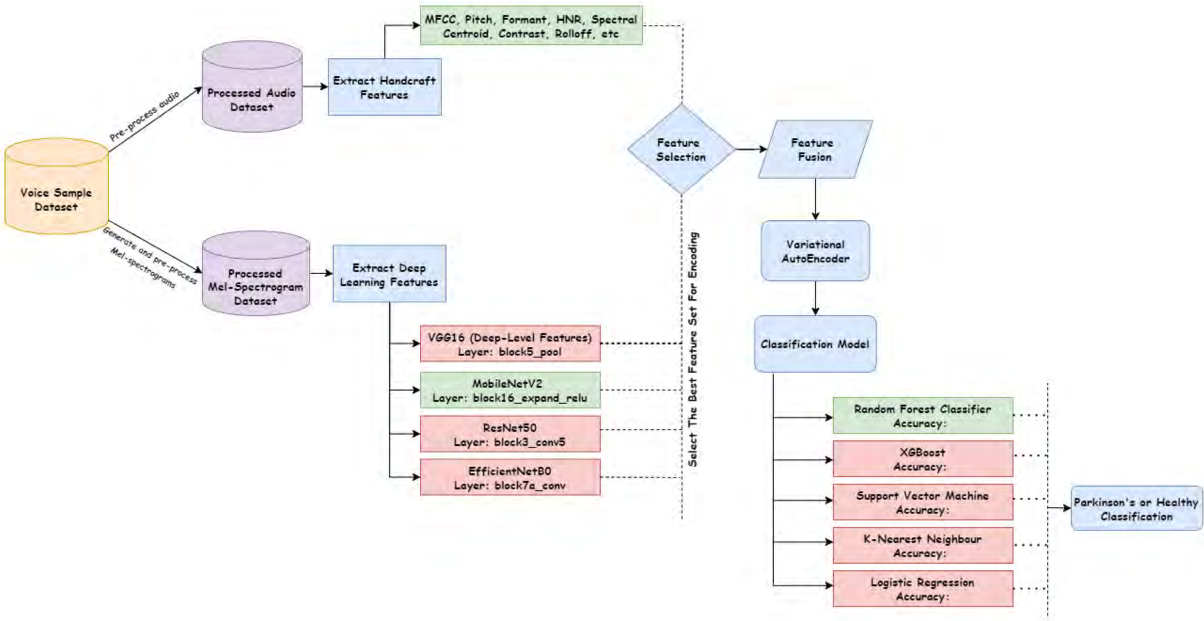

Figure 4.1: Top level overview of the proposed system

After feature extraction, both handcrafted and deep learning features are concatenated into a unified feature set. A feature selection process is applied to identify the attributes that are most applicable to classification. These selected features are then encoded using a Variational AutoEncoder (VAE), which reduces the dimensionality while preserving key information. Finally, the encoded features are passed to multiple machine learning classifiers, including Random Forest, XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression. The classifiers predict whether each sample belongs to a Parkinson's patient or a healthy control, providing a comprehensive assessment of model performance for Parkinson's Disease detection.

4.1 Feature Extraction

A hybrid feature extraction approach is employed in this research that combines handcrafted acoustic features with deep learning-based features extracted from mel-spectrograms of voice samples. This dual approach allows to capture both traditional acoustic properties of voice signals and complex patterns learned by convolutional neural networks (CNNs). By leveraging both feature sets, the aim is to enhance the accuracy and robustness of Parkinson's Disease classification.

4.1.1 Handcraft Features

State of the art audio processing libraries such as librosa and Praat were used to extract hand crafted features based on speech analysis. In particular, the focus was on features that are known to be helpful in problems of voice pathology, particularly when associated with neurological conditions, like Parkinson's disease, whereby motor control of the voice is affected. The set of handcrafted features included 26 key acoustic properties, each designed to capture different aspects of the voice.

i. MFCCs (Mel-Frequency Cepstral Coefficients 1-14)
The short term power spectrum of the voice samples were represented by first extracting the first 14 MFCCs. Power spectrum of the signal is mapped to Mel scale of human auditory perception such that MFCCs are derived. These coefficients are critical ones in phonetic information and are used otherwise in speech processing tasks. Specifically, the lower order MFCCs contain information from the lower frequency region (broad spectrum), and higher order MFCCs are finer spectral details which allow the model to distinguish between different behaviours of speech which arise from different stages of Parkinson's Disease (e.g., onset and progression of symptoms, silent periods).

ii. Pitch Mean
The average value across each of the voice sample was computed as the fundamental frequency, or pitch. Consistent pitch variability is a known hall mark of Parkinson's Disease, in which patients may have difficulty in modulating pitch (monotonicity) as a result of impaired vocal fold control. We can quantify overall vocal stability and range by measuring the pitch mean.

iii. Formant Frequencies (Formant 1 and Formant 2)
Resonant or formants are frequencies audible qualities of the vocal tract during phonation. Vowel differentiation is greatly dependent on the first and second formants (F1 and F2) and is known to be affected by articulatory changes in people with Parkinson's. Because the ability to capture changes in vowel production and resonance, for example, is common in speech disorders, these features are important for measuring these effects.

iv. Intensity Mean
Each voice sample is calculated as mean of its sample intensity (or loudness). Weakness of respiratory and laryngeal control results in diminished vocal intensity in Parkinson's Disease. It permits listening to voice variations in feeding effort and thus contributes to the description of voice performance.

v. Jitter (local)
Cycle to cycle variation of pitch (fundamental frequency) reflects the stability of vocal fold vibration and is the measurement of jitter. Increase in jitter implies irregular vocal fold motion and constitutes a marker of dysphonia. Because higher jitter values occur in Parkinson's Disease due to vocal tremors and in inconsistent vocal fold movements, this is a critical feature for indicating voice instability.

vi. Shimmer (local)
The amplitude variation between consecutive vocal cycles, as measured by shimmer, is quantified as a probe into loudness stability. Shimmer values indicate higher instability of vocal, where such as patients with Parkinson Disease having difficulty in the amplitude of their voice across cycles.

vii. Harmonics-to-Noise Ratio (HNR)
The harmonic to noise ratio or HNR, is a measurement of the proportion of the voice signal's periodic (noise) and periodic (harmonic) components. A lower HNR suggests that there is more noise in the signal thus you are more likely to have irregular vocal fold vibration. Patients with Parkinson's have reduced HNR because of increased breathiness or hoarseness in their voice.

viii. Spectral Centroid
The spectral centroid is often referred to as the 'center of mass' of the frequency spectrum and is directly related to the perceived brightness of a sound. A lower spectral centroid corresponds to darker, more muted tones, while a higher centroid is associated with brighter, sharper sounds. This feature is crucial in distinguishing between different timbral qualities of speech or audio signals.

ix. Spectral Bandwidth
The frequency band's width is measured by spectral bandwidth, providing insights into the spread of frequencies around the spectral centroid. A wider bandwidth indicates a more complex spectrum, often suggesting a richer and more dispersed frequency distribution, while a narrower bandwidth implies a more focused or simpler frequency range.

x. Spectral Contrast
Spectral contrast highlights the differences between spectral peaks and valleys, making it useful for distinguishing between voiced and unvoiced sounds. This feature captures the variation in amplitude across different frequency regions, which is important for identifying the texture and tonal quality of a sound. It is particularly useful in separating harmonic structures from noise-like components in a signal.

xi. Spectral Flatness
Spectral flatness quantifies how flat or peaky the sound spectrum is. A high spectral flatness value indicates that the spectrum is relatively flat, resembling white noise, while lower values suggest a more tonal quality with distinct peaks. This feature is important for classifying sounds as either harmonic (tonal) or noisy (non-harmonic).

xii. Spectral Rolloff
The frequency below which a given portion (typically 85%) of the total spectrum is present is known as spectral roll off. It effectively differentiates between high-energy and low-energy frequency bands, offering insight into the distribution of spectral energy. A higher rolloff frequency suggests a greater concentration of energy in the upper frequency range, while a lower rolloff indicates lower frequencies contain most of the concentrated energy.

To explore the relationships among the handcrafted features, a feature correlation matrix was computed, offering insight into the linear associations between feature pairs and revealing potential redundancy or collinearity within the feature set. Highly correlated features may suggest overlapping information, while less correlated features highlight distinct vocal attributes. Additionally, the dimensionality of the feature space was reduced by apllying Principal Component Analysis (PCA) was applied to reduce the dimensionality of the feature space, allowing for a two-component visualization. The PCA plot demonstrates how effectively the handcrafted features capture the variance in the data and reveals the degree of separation between the Parkinson's and Healthy classes in this reduced space.

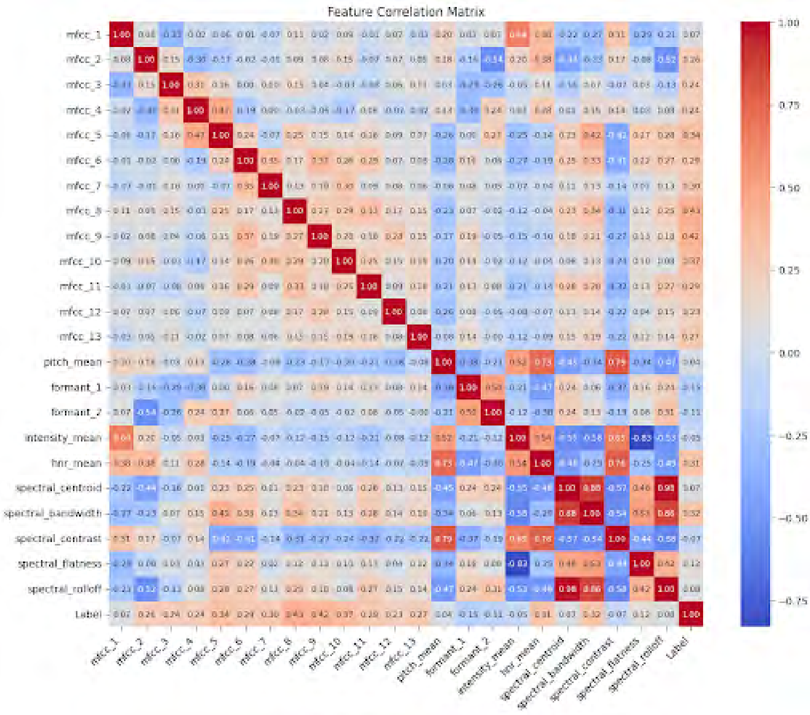

Figure 4.2: Feature Correlation Matrix

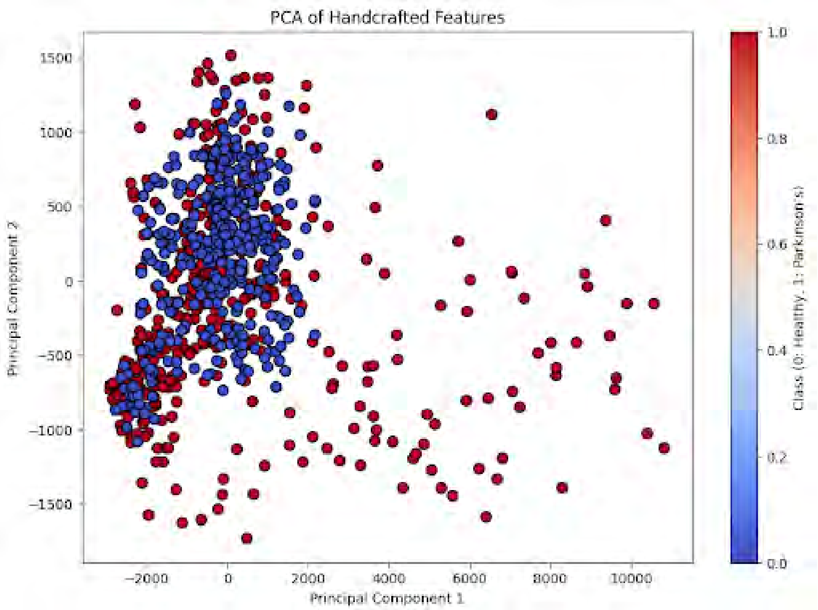

Figure 4.3: PCA of Handcraft Features

Figure 4.2 presents the correlation matrix of the handcrafted features, where color intensity reflects the strength of correlation between feature pairs. Positive correlations indicate that features increase together, while negative correlations show that one decreases as the other increases.
Figure 4.3 displays the PCA scatter plot of the two principal components, illustrating the separation between Parkinson's and Healthy classes. Each point represents a voice sample, providing a clear visualization of how the handcrafted features contribute to distinguishing the two groups.

To conclude, the analysis of the handcrafted features, through both the correlation matrix and PCA plot, reveals that while these features capture important vocal characteristics relevant to Parkinson's Disease, they may not be sufficient on their own for optimal classification. The correlation matrix helps identify independent and overlapping features, and the PCA plot shows some degree of class separation between Healthy and Parkinson's samples. However, the limited separability suggests that these handcrafted features may not fully capture the complex patterns necessary to distinguish between the two groups. This indicates that while handcrafted features provide valuable insights, they likely need to be combined with other features to better capture non-linear patterns and improve classification accuracy.

4.1.2 Deep Learning Features

To complement the handcrafted features, deep learning-based feature extraction was employed utilizing pre-trained Convolutional Neural Networks (CNNs). CNNs are well-suited for automatically learning hierarchical, abstract representations from complex data, such as mel-spectrograms derived from voice samples. For this purpose, we leveraged four state-of-the-art models: VGG16, MobileNetV2, ResNet50, and EfficientNetB0. These architectures were selected for their proven effectiveness in capturing multi-level feature representations, ranging from mid-level to deep-level abstractions, which are critical for identifying subtle patterns in vocal signals indicative of Parkinson's Disease. The features extracted from specific layers of these networks provide a robust and discriminative feature set that enhances the classification performance.

a. VGG16

VGG16 is known for its simple yet effective architecture, with consistently sized convolutional layers. For mid-level features, we selected the block4_conv3 layer. This layer focuses on detecting edges, vertices, and texture patterns in the mel-spectrograms—key components that can capture the fine structure of voice signals affected by Parkinson's Disease (Yang et al., 2021)[15]. By stopping at this intermediate layer, we ensure that we capture important vocal traits without overly

abstracting the information.
For deep-level features, we extracted from the block5_pool layer. At this deeper stage, the network captures more abstract and global patterns in the data, such as higher-level acoustic structures that distinguish Parkinson's from Healthy voices. This combination of mid- and deep-level features ensures that we capture both local acoustic properties and higher-order abstractions.

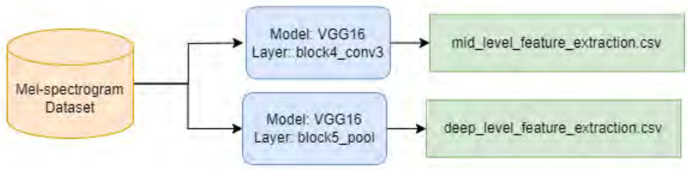

Figure 4.4: Mid-level and deep-level feature extraction

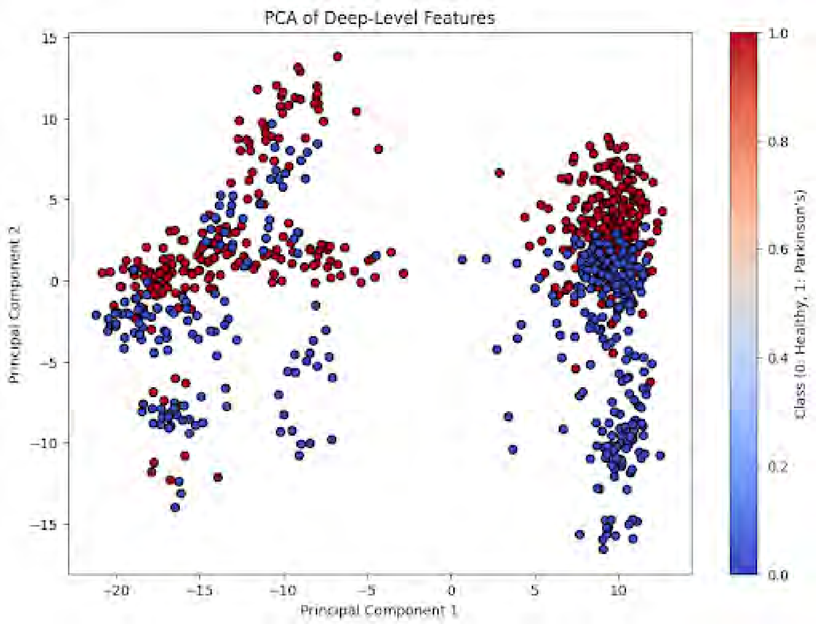

Figure 4.5: PCA of VGG16

By seeing the clusters that the PCA projection created, it can be seen how efficiently the features separate the two classes. Better classification performance results from features that are more discriminative, meaning red and blue clusters can be distinguished more easily.

b. MobileNetV2

MobileNetV2, designed for efficient computation, uses depthwise separable convolutions, which makes it well-suited for capturing features from spectrograms with fewer parameters (Akay et al., 2021)[16]. We extracted deep-level features from the block_16_expand_relu layer. This layer is particularly effective at preserving critical high-level information while maintaining a balance between feature expressiveness and model efficiency. The extracted features are highly compressed, yet they retain enough detail to differentiate between the subtle variations in voice patterns caused by Parkinson's disease.

Figure 4.6: MobileNet feature extraction

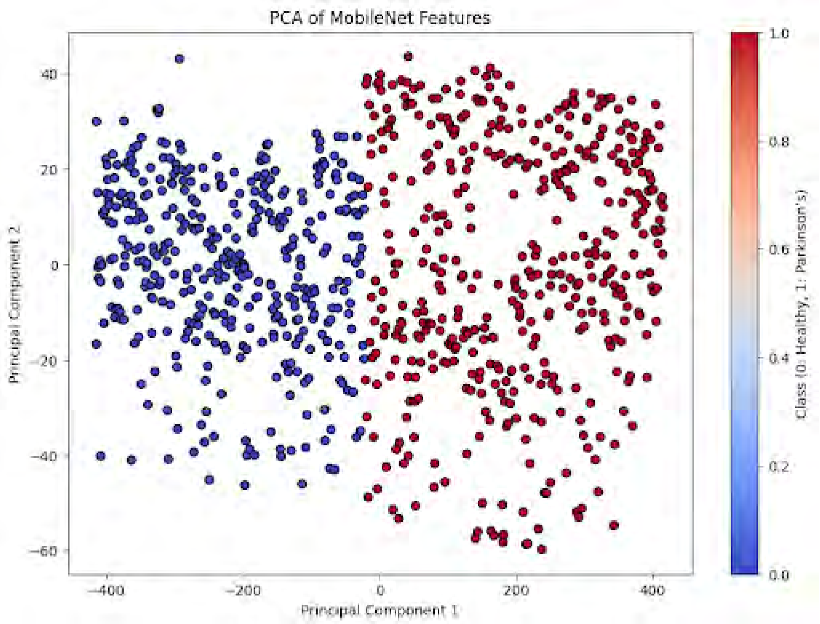

Figure 4.7: PCA of MobileNetB0

The PCA of MobileNetB0 show much clearer separation between the two classes. This clear division suggests that classifiers perform better.

c. EfficientNetB0

Network's width,depth, and resolution are balanced by a compound scaling method that EfficientNetB0 uses (Goutham et al., 2022)[17] . We extracted deep-level features from the block7a_conv layer. This deep convolutional layer is optimized to capture global, abstract features from the spectrograms, representing high-level time-frequency patterns. The model's efficient architecture allows it to capture these features with a relatively low number of parameters, making it suitable for extracting nuanced representations from the voice data, which can aid in the accurate classification of Parkinson's Disease.

Figure 4.8: EfficientNet feature extraction

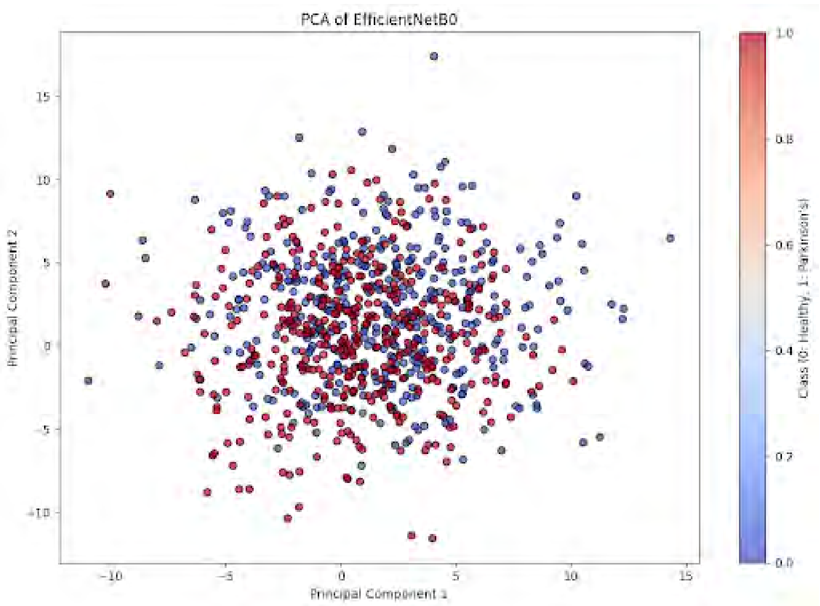

Figure 4.9: PCA of EfficientNet

This PCA plot here represents the projection of deep features extracted from EfficientNetB0 model. The PCA shows the overlap between the two classes, leading to lower classification accuracy as it becomes difficult for the classifiers to compare between healthy controls and Parkinson's patients.

d. ResNet50

ResNet50 introduces skip connections (residual learning) to overcome the degradation problem in deep networks (Liang, 2020)[18]. We extracted deep-level features from the block3_conv5 layer. By selecting this layer, we leverage the residual connections that allow the network to learn deep representations without losing critical information about the input. The features extracted at this stage represent highly abstract patterns in the mel-spectrograms, capturing complex temporal and spectral variations in the voice signals. However, for our work, ResNet50 may not be efficient due to the relatively small size of our dataset. Deep networks like ResNet50 require large amounts of data to fully exploit their capacity and avoid overfitting. In smaller datasets, such complex models tend to overfit, capturing noise instead of meaningful patterns, leading to suboptimal performance. Therefore, models with fewer parameters or more efficient architectures, like MobileNetV2 or EfficientNet, are more suitable for our classification task.

Figure 4.10: ResNet feature extraction

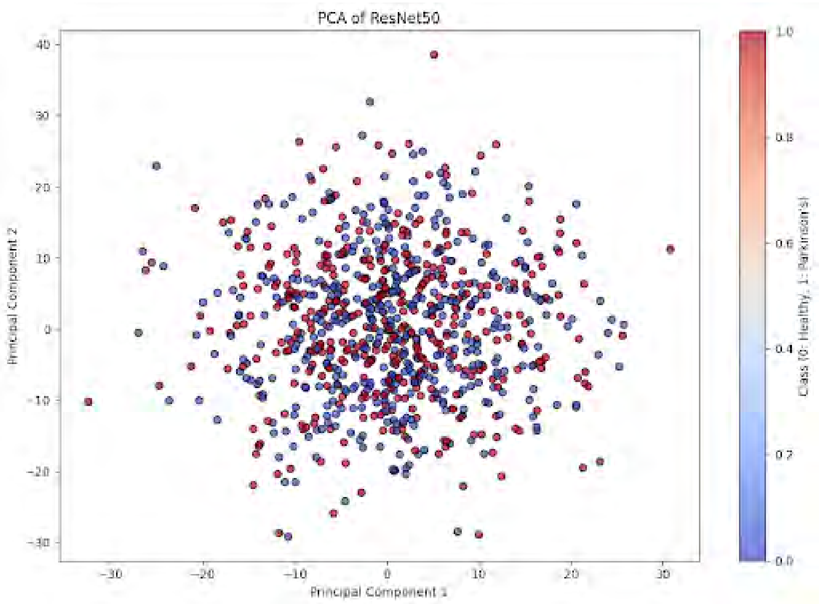

Figure 4.11: PCA of ResNet50

This PCA shows notable overlap between the two classes, making it difficult for classifiers to distinguish between healthy control and Parkinson's patients. Due to this overlap of classes, the classifiers provide less accuracy.

By employing these diverse CNN architectures, we have constructed a comprehensive feature extraction pipeline that not only captures the nuanced spectral and temporal relationships within the voice samples but also optimizes the feature space also known as machine learning-based classification.

4.1.3 Feature Selection and Fusion

During the feature extraction process, researchers identified numerous columns that were entirely filled with zeros. These zero valued features were effectively non-informative, providing no variance or useful data for the classification models. Hence, these features were removed. Retaining such features would have diluted the model's learning capacity, leading to overfitting or incorrect conclusions about feature importance. In addition to completely zero valued features, there were other features with sparse, yet small non-zero values, such as (0.00230, 0.010349). These low-variance features contributed minimally to the overall feature space. To ensure that the model was not biased by such insignificant variations, a variance-based approach was applied to either remove or normalize these values. In some cases, these values were replaced based on the overall variance, mean or median of the dataset, ensuring consistency across features while preserving the integrity of the more informative data points. This feature refinement process—removing zero-valued features and normalizing low-variance ones—was systematically applied across all the models that were used, including ResNet (conv5_block3_out), MobileNetV2 (block_16_expand_relu), VGG16 (block5_pool), and EfficientNet-B0 (block7a_layer). The objective was to reduce noise, prevent skewed learning, and focus the classification models on features that provided substantive insight into the differences between Parkinson's patients and healthy controls. This careful curation of feature spaces allowed models to perform more efficiently, ultimately leading to more accurate predictions.

4.2 Latent Feature Extraction Using Variational Autoencoder(VAE)

To address the high dimensionality and potential redundancy introduced by the concatenation of handcrafted and deep learning features, we employed a Variational AutoEncoder (VAE) for feature selection and dimensionality reduction. The VAE, a generative model, is particularly suited for compressing high-dimensional data into a lower-dimensional latent space while preserving the most critical information necessary for classification.

The VAE comprises two main components: an encoder, which maps the input features into a compact latent space, and a decoder, which uses the latent representation to try and recreate the original input (Dai Wipf, 2019)[19]. By minimizing the reconstruction error and a regularization term (Kullback-Leibler divergence), the VAE ensures that the encoded latent features are both informative and compact.For this task, after implementing handcraft and deep learning (VGG16, MobileNetV2, ResNet50, and EfficientNetB0) feature fusion, the VAE was used to reduce the combined feature set into a lower-dimensional space. This dimensionality reduction step is crucial for mitigating overfitting, improving computational efficiency, and enhancing the discriminative power of the features used in the subsequent classification models.

The latent features extracted by the VAE were then used as input for the classification models, providing a more compact and informative feature representation, which contributed to improved classification accuracy.

4.3 Model Specification

For the classification of Parkinson's Disease and Healthy control voice samples, we employed several machine learning algorithms, each known for its ability to handle diverse feature sets and complex decision boundaries. The models utilized in this study include Random Forest Classifier (RFC), Logistic Regression (LR), K-Nearest Neighbors (KNN), XGBoost, and Support Vector Machine (SVM). These models were chosen for their distinct approaches to classification, allowing us to compare performance across different learning paradigms.

4.3.1 CNN Architectures

A specific kind of deep neural network called a convolutional neural network (CNN) is mainly used to handle structured, grid-like data, like pictures. Using convolutional, pooling, and fully connected layers, CNNs are made to automatically and adaptively learn spatial hierarchies of characteristics from input data (Alzubaidi et al., 2021)[20].

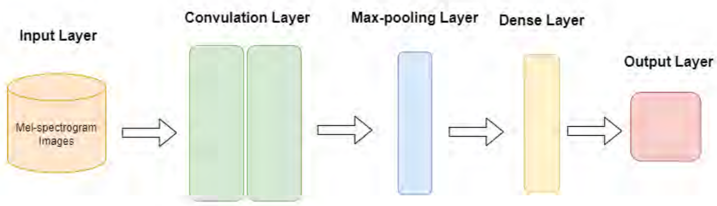

Figure 4.12: CNN Architecture

Convolutional Layers: Convolutional Layers identify local patterns such as edges and textures by applying filters (kernels) to input data.
Pooling Layers: By downsampling the data, pooling layers reduce computation and dimensionality while maintaining significant features.
Fully Connected Layers: These layers use feature interpretation to provide predictions following feature extraction.

a. VGG16

VGG16 is widely recognized for its architectural simplicity and uniformity, which lies in its consistent use of design elements throughout its layers. A defining characteristic of VGG16 is the repeated use of 3x3 convolutional filters across its convolutional layers [15]. This deliberate design choice not only simplifies the architecture but also enhances its ability to maintain uniformity across different stages of the network. VGG16 excels in extracting hierarchical features, where "hierarchical" refers to the ability to capture features at multiple levels of abstraction, from low-level edges to high-level complex structures. This multi-scale feature extraction proves highly effective for comprehensive data representation, enabling the model to capture patterns of varying complexity. The study of VGG16's architecture is motivated by its proven capability to efficiently extract and reveal hierarchical patterns within diverse and complex image datasets, thus demonstrating its robustness in feature extraction across a wide range of applications.

b. MobileNetV2

MobileNet is an efficient convolutional neural network (CNN) architecture designed for resource-constrained environments, such as mobile devices and embedded systems. Its key innovation lies in the use of depthwise separable convolutions, which split standard convolutions into two operations: depthwise convolution, where each filter is applied to a single input channel, and pointwise convolution, where 1x1 filters combine these outputs. This significantly reduces the number of parameters and computations, making the model faster and more lightweight while maintaining strong performance. MobileNet also includes two key hyperparameters: the width multiplier, which controls the number of filters in each layer, and the resolution multiplier, which adjusts the input resolution, offering flexibility to scale the model for different tasks and resource constraints. This architecture is highly effective for capturing features at different levels of abstraction, making it suitable for a wide range of tasks, including the processing of voice samples, as it can efficiently extract patterns from complex datasets with reduced computational cost[16].

c. ResNet50

ResNet50 introduces skip connections (residual learning) to overcome the degradation problem in deep networks[18]. We extracted deep-level features from the block3 conv5 layer. By selecting this layer, we leverage the residual connections that allow the network to learn deep representations without losing critical information about the input. The features extracted at this stage represent highly abstract patterns in the mel-spectrograms, capturing complex temporal and spectral variations in the voice signals. These high-level abstractions are crucial for distinguishing between Parkinson's patients and healthy controls.

d. EfficientNetB0

Compound scaling is used by EfficientNetB0 to balance network, width, depth and resolution. We extracted deep-level features from the block7a conv layer. This deep convolutional layer is optimized to capture global, abstract features from the spectrograms, representing high-level time-frequency patterns[17]. The model's efficient architecture allows it to capture these features with a relatively low number of parameters, making it suitable for extracting nuanced representations from the voice data, which can aid in the accurate classification of Parkinson's Disease.

4.3.2 Variational Autoencoder(VAE) Architecture

A Variational Autoencoder (VAE) is a particular type of generative model: it learns a probability distribution over its latent space (typically Gaussian) by encoding input data into a latent space [19]. Unlike standard autoencoders, the VAEs use a probabilistic approach where for any input a distribution is encoded with its mean and variance instead of a single fixed point. The VAE optimizes two key objectives: KL divergence, which basically helps make the output match the input, and re- construction loss. With this balance, VAEs can produce new data from latents sampled from learned latent space. For tasks like data generation, compression and feature extraction, VAEs are proven to be effective and, especially for much more complex data like voice samples, they are capable of capturing structured, latent representations.

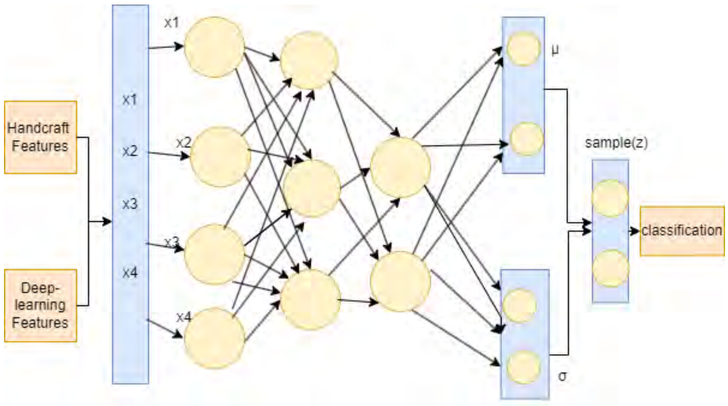

Figure 4.13: VAE Architecture

4.3.3 Traditional Machine Learning Architecture

a. Random Forest Classifier(RFC)

Random Forest Classifier (RFC) is an ensemble learning algorithm used for classification tasks. It works by building multiple decision trees during training and aggregating their results to make a final prediction (Belgiu Drăguţ, 2016)[21]. Each tree is trained on a random subset of the data, and a random subset of features is considered for each split, which helps reduce overfitting and improves generalization. By averaging the predictions from numerous trees, RFC provides more robust and accurate predictions compared to a single decision tree. RFC is highly versatile, handling large datasets with high dimensionality effectively, and is known for its resilience to noise and ability to handle missing data.

b. K-Nearest neighbour (KNN)

K-Nearest Neighbors (KNN) is a simple, non-parametric algorithm used for classification and regression tasks. In KNN, a data point is classified based on the majority class of its K closest neighbors in the feature space, where distance is typically measured using Euclidean distance (Zhang Han, 2007)[22]. The algorithm does not assume any underlying distribution of the data and makes predictions by comparing new data points to existing labeled instances. KNN is highly intuitive and easy to implement but can become computationally expensive with large datasets, as it requires storing all the data points and calculating distances for each prediction. It works well for smaller datasets and is particularly effective when the decision boundary is not linear.

c. Logistic Regression (LR)

Logistic Regression (LR) is a statistical model commonly used for binary classification tasks. It predicts the probability that a given input belongs to a particular class by modeling the relationship between the input features and the class labels using the logistic function (Yadav Singh, 2021)[23]. The logistic function outputs a value between 0 and 1, which can be interpreted as a probability. LR estimates the parameters (weights) by maximizing the likelihood of the observed data, typically using techniques like gradient descent. Although linear in nature, it is powerful in distinguishing between two classes and is widely applied in classification tasks due to its simplicity, efficiency, and interpretability.

d. Support Vector Machine (SVM)

SVM are employed for tasks involving regression and classification. For tasks involving regression and classification, supervised learning models called support vector machines (SVM) are employed. One of the main goal of SVM is to accurately identify the hyperplane that divides data points into the appropriate classes. The support vector machine (SVM) looks for the hyperplane that maximizes the margin, or the distance between the hyperplane and the closest data points from each class. SVM minimizes classification errors and enhances generalization by optimizing this margin. SVM may map data into higher dimensions when a linear separation is feasible by using kernel functions for non-linearly separable data. SVM is well renowned for its accuracy and robustness, especially when dealing with high-dimensional datasets, and is particularly useful in binary classification applications(Soumaya et al., 2020)[24].

e. XGBoost

XGBoost is a powerful gradient boosting algorithm commonly used for classification and regression tasks. It builds an ensemble of decision trees sequentially, with each new tree correcting the errors of the previous ones. XG Boost stands out for its speed and performance due to its advanced features like regularization to prevent overfitting, handling missing values, and parallel processing capabilities. It optimizes both accuracy and computational efficiency, making it suitable for large-scale datasets. XG Boost is widely favored in machine learning competitions due to its ability to handle complex datasets and deliver highly accurate predictions (Surya et al., 2021)[25].

Chapter 5

Results and Discussion

5.1 Performance Metrics

The performance metrics utilized to evaluate the classifiers are outlined as follows:

- **Accuracy**: Represents the overall proportion of correct predictions relative to the total number of predictions made. It provides a basic measure of model performance but may be insufficient in the presence of class imbalance.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$$

- **Precision**: Reflects the proportion of true positive predictions among all instances predicted as positive. Precision is crucial in scenarios where minimizing false positives is a priority.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall** (Sensitivity or True Positive Rate): Denotes the proportion of actual positive instances that were correctly identified by the model. This metric is essential when the goal is to maximize the identification of positive cases, especially when false negatives carry a higher cost.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1-Score**: Represents the harmonic mean of precision and recall, offering a balanced measure that is particularly useful in cases where the dataset is imbalanced, or both false positives and false negatives are important.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Confusion Matrix**: A tabular representation that provides a detailed summary of the model's classification performance by displaying the counts of true positives, true negatives, false positives, and false negatives. The confusion matrix offers deeper insights into misclassifications and model behavior across different classes.

These metrics collectively offer a comprehensive evaluation of classifier performance, guiding both model selection and refinement.

5.2 Result Analysis

A hybrid model is developed and evaluated using four distinct approaches, each combining handcrafted acoustic features with deep learning-based features extracted from different CNN architectures for improved Parkinson's Disease detection.

The first approach integrates handcraft features with deep-level features from VGG16, followed by classification using Random Forest Classifier (RFC), Support Vector Machine (SVM), KNN, LR, and XGBoost. Similarly, the second approach combines handcrafted features with deep features extracted from MobileNetV2, applying the same classifiers (RFC, SVM, KNN, LR, and XGBoost) for evaluation. In the third approach, EfficientNet was used in conjunction with the handcrafted features and again tested with the five classifiers. Finally, the fourth approach incorporated features from ResNet alongside the handcrafted features, followed by classification using RFC, SVM, KNN, LR, and XGBoost. Through this systematic hybrid approach, the integration of traditional and deep learning-based features is explored to enhance the accuracy and robustness of Parkinson's Disease classification.

a. Feature Fusion of Handcraft and MobileNetV2

These models have run through 5 classifiers;Random Forest(RF), XG Boost, SVM, Knn and logistic Regression(LR) and their classification table is given below:

Model	Accuracy	Precision	Recall	F1-Score	Support
Random Forest	0.94	0.95	0.94	0.94	250
XGBoost	0.94	0.95	0.94	0.94	250
SVM	0.93	0.93	0.93	0.93	250
K-Nearest Neighbors	0.93	0.93	0.93	0.93	250
Logistic Regression	0.99	0.99	0.99	0.99	250

Table 5.1: Classification Table for extracted features (Handcraft and MobileNetV2)

Model Performance Visualization (MobileNetV2)

A comparison of machine learning classifiers based on features taken from MobileNetV2 is shown in the graph in Figure 5.1. Accuracy, Recall, Precision, and F1-Score are just a few of the performance measures that demonstrate how well Random Forest and XGBoost function consistently, receiving high scores in each category. SVM and K-Nearest Neighbors perform somewhat worse, especially in F1-Score and Recall. Though it lags behind the best-performing models, logistic regression nonetheless performs satisfactorily. Overall, the findings show that Random Forest and XGBoost perform well when paired with characteristics that were taken from MobileNetV2.

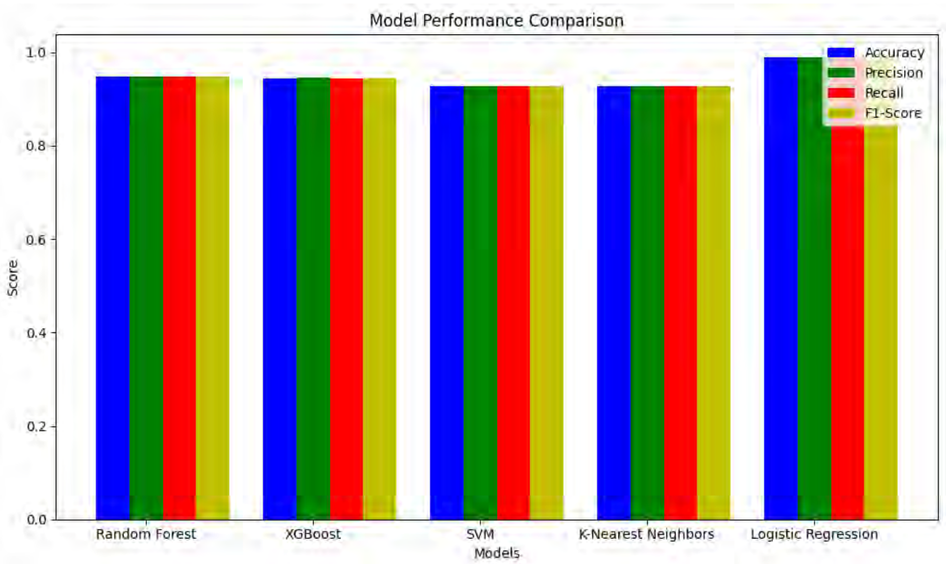

Figure 5.1: Model Performance of MobileNetV2

b. Feature Fusion of Handcraft and ResNet50

These models have run through 5 classifiers;Random Forest(RF),XG Boost,SVM,Knn and logistic Regression(LR) and their classification table is given below

Model	Accuracy	Precision	Recall	F1-Score	Support
Random Forest	0.85	0.85	0.85	0.85	249
XGBoost	0.82	0.82	0.82	0.82	249
SVM	0.71	0.76	0.69	0.68	249
K-Nearest Neighbors	0.79	0.79	0.78	0.78	249
Logistic Regression	0.89	0.89	0.89	0.89	249

Table 5.2: Classification Table for extracted features (Handcraft andResNet50)

Model Performance Visualization (ResNet50)

The graph in Figure 5.2 exhibits the performance of machine learning classifiers using features extracted from ResNet50. Strong performance is exhibited by Random Forest and XGBoost ,achieving high Accuracy, Precision, Recall, and F1-Score values. While SVM shows a noticeable drop, particularly in Recall and F1-Score, K-Nearest Neighbors performs similarly but with slightly lower scores. Logistic Regression, though consistent, lags behind the top-performing models in a few metrics. Therefore, Random Forest and XGBoost appears as the most effective classifiers when combined with ResNet50-extracted features.

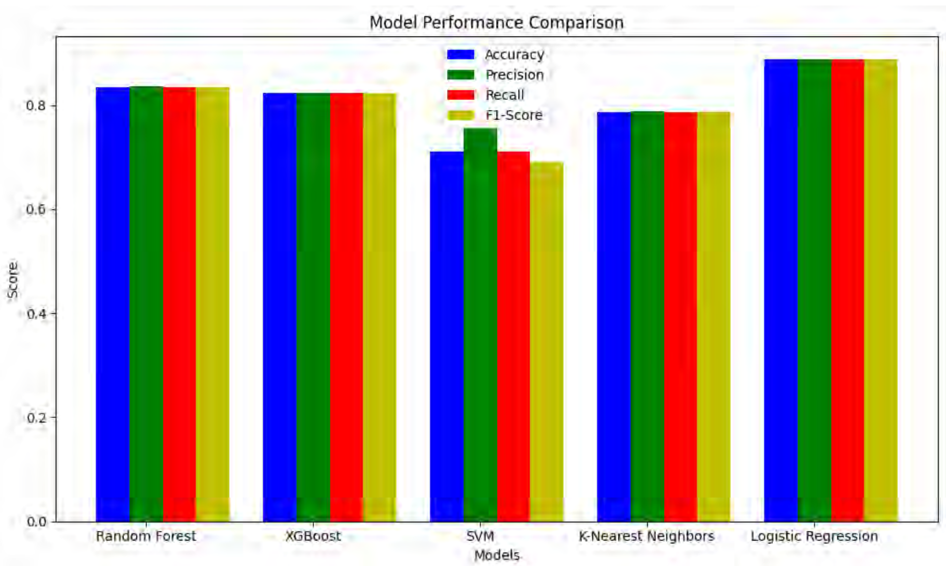

Figure 5.2: Model Performance of ResNet50

c. Feature Fusion of Handcraft and EfficientNetB0

These models were evaluated using five classifiers: Random Forest (RF), XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR), with their classification performance summarized in the table below.

Model	Accuracy	Precision	Recall	F1-Score	Support
Random Forest	0.83	0.83	0.83	0.83	250
XGBoost	0.82	0.82	0.82	0.82	250
SVM	0.69	0.70	0.68	0.68	250
K-Nearest Neighbors	0.74	0.74	0.73	0.73	250
Logistic Regression	0.76	0.76	0.76	0.76	250

Table 5.3: Classification Table for extracted features (Handcraft and EfficientNetB0)

Model Performance Visualization (EfficientNetB0)
Based on features taken from the graph in Figure 5.3 evaluates the performance of various classifiers using features extracted from EfficientNetB0. Random Forest and XGBoost maintain high performance across all metrics, including Accuracy, Precision, Recall, and F1-Score. SVM, on the other hand, exhibits a discernible decline in both Recall and F1-Score, highlighting its weaker performance. K-Nearest Neighbors and Logistic Regression perform consistently well but still lag slightly behind the top-performing models. This analysis underscores the robust classification ability of Random Forest and XGBoost when utilizing EfficientNetB0 features.

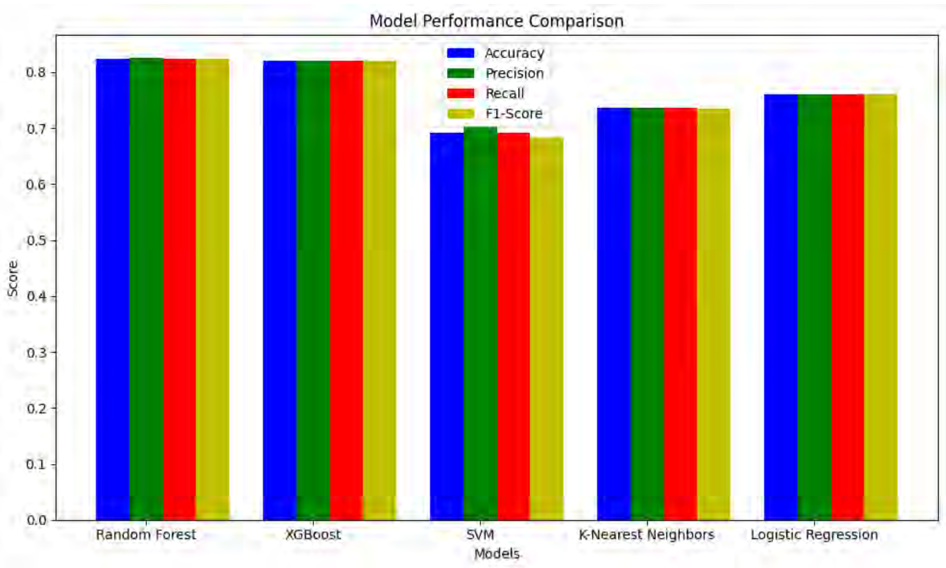

Figure 5.3: Model Performance of EfficientNetB0

d. Feature Fusion of Handcraft and VGG16

These models were evaluated using five classifiers: Random Forest (RF), XGBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR), with their classification performance summarized in the table below.

Model	Accuracy	Precision	Recall	F1-Score	Support
Random Forest	0.83	0.84	0.84	0.84	250
XGBoost	0.84	0.84	0.84	0.84	250
SVM	0.70	0.75	0.68	0.67	250
K-Nearest Neighbors	0.77	0.77	0.77	0.77	250
Logistic Regression	0.78	0.78	0.78	0.78	250

Table 5.4: Classification Table for extracted features (Handcraft and VGG16)

Model Performance Visualization (VGG16)

The graph in Figure 5.4 presents a performance comparison of machine learning classifiers using features extracted from VGG16. Random Forest and XGBoost perform admirably and consistently across all metrics, including Accuracy, Precision, Recall, and F1-Score. SVM, on the other hand, exhibits lower scores, particularly in Recall and F1-Score. K-Nearest Neighbors and Logistic Regression perform fairly well, though they do not quite match the top classifiers. This evaluation highlights the superior performance of Random Forest and XGBoost when leveraging VGG16 features for classification.

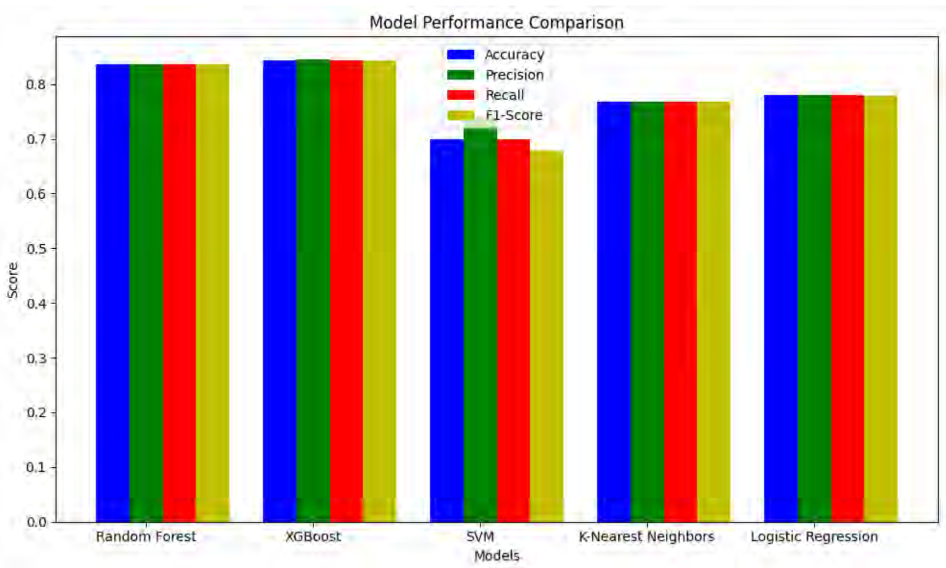

Figure 5.4: Model Performance of VGG16

5.2.1 Model Performance Evaluation

The performance of five different classifiers on four distinctive combined data frames is compared in this study in order to select the best model for each frame. The combined feature set of Handcraft and MobileNETV2 feature shows phenomenal accuracy in this study. Analysis shows that the combined data frame of Handcraft and MobileNETV2 features achieves 99% accuracy with the Logistic Regression Model. However, Logistic Regression is slightly under-performing with an accuracy of 85% using the combined feature set of Handcraft and ResNet50. The classification accuracy of the feature set HandCraft + EfficientNETB0 is 83% achieved by random Forest Classifier. Lastly, the classification model XGBoost shows better performance compared to the remaining models used for the feature set HandCraft + VGG16. Classification accuracy for this model is 84%. Taken together, the results underscore the importance of choosing an appropriate classifier optimized to the specific characteristics and features of a given data frame, suggesting more generally how such model selection may be optimized in other data analysis contexts.

Data Frame	**Classifier (best)**	**Accuracy**
HandCraft + MobileNETV2	Logistic Regression	99%
HandCraft + ResNET50	Logistic Regression	85%
HandCraft + EfficientNETB0	Random Forest Classifier	83%
HandCraft + VGG16	XGBoost	84%

Table 5.5: Classification Performance for Various Data Frames

The confusion matrix is a great tool for the evaluation of classification models since it provides a fair level of detail over the classification's performance that is not extracted from the overall accuracy. In general, this gives a more granular understanding of the classifier's behavior, especially in terms of true positives, false positives, true negatives, and false negatives — or at least provides insights for these parameters, given imbalanced datasets or critical error cases. In this study, we construct confusion matrices based on the best-performing classifiers which achieved the highest accuracies shown in Table 5.5. These confusion matrices give us an overview of confusion and correct classes made by the models. For example, as applied by the HandCraft + MobileNetV2 combine feature set, the highest accuracy of 99% was obtained through the Logistic Regression classifier and the accompanying confusion matrix shows the model's prediction performance in detail. We also include confusion matrices for some of the other high-performing models like HandCraft + ResNet50 that reached 85% accuracy to show how the classification patterns look and how precise were the predictions in the dataset. The effectiveness of these matrices lies in using them to evaluate the robustness and reliability of the models used to predict Parkinson's disease and to account for why the models, in total, might prove effective.

a. The confusion matrix for the best-performing classifier (Handcraft and MobileNetV2 Feature Fusion)

The confusion matrix in Figure 5.5 shows the performance of Logistic Regression with handcrafted and MobileNetV2 feature fusion. The model accurately classified 113 negative cases and 134 positive cases, with only three false positives and no false negatives. This highlights the strong accuracy of Logistic Regression, particularly in correctly identifying both classes with minimal errors.

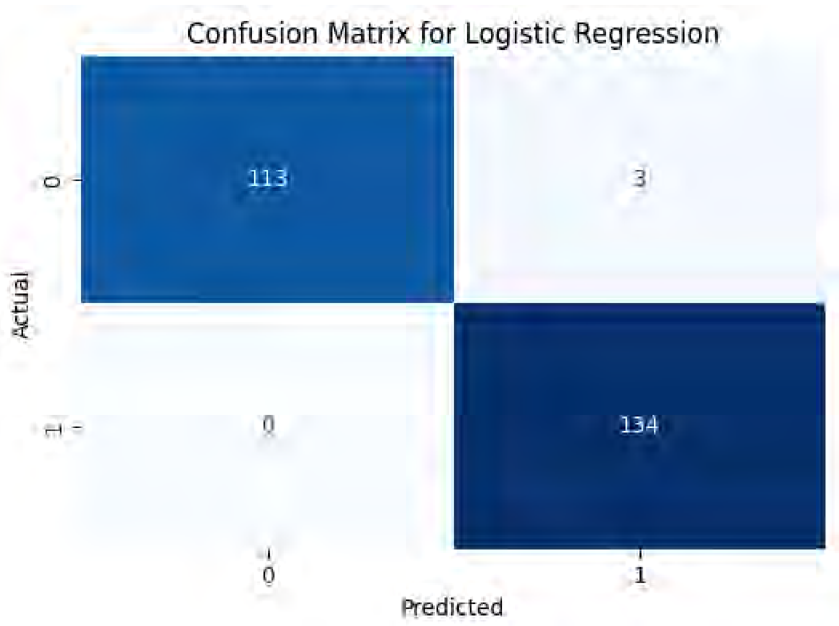

Figure 5.5: Confusion Matrix of MobileNetV2

b. The confusion matrix for the best-performing classifier (Handcraft and ResNet Feature Fusion)

The confusion matrix in Figure 5.6 displays the performance of Logistic Regression with handcrafted and ResNet feature fusion. 102 negative cases and 119 positive cases are accurately classified, with 14 false positives and 14 false negatives. While the overall accuracy is strong, the presence of false classifications in both classes suggests room for improvement in handling the fused feature set.

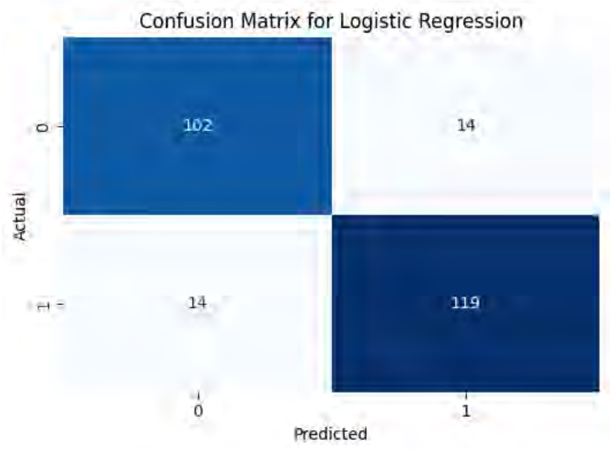

Figure 5.6: Confusion Matrix of ResNet

c. The confusion matrix for the best-performing classifier (Handcraft and EfficientNet Feature Fusion)

The confusion matrix in Figure 5.7 shows the performance of the Random Forest classifier with handcrafted and EfficientNet feature fusion. The model correctly classified 92 negative cases and 115 positive cases, with 24 false positives and 19 false negatives. Although the overall accuracy is reasonable, the false positives and false negatives indicate some challenges in distinguishing between the two classes using the fused feature set.

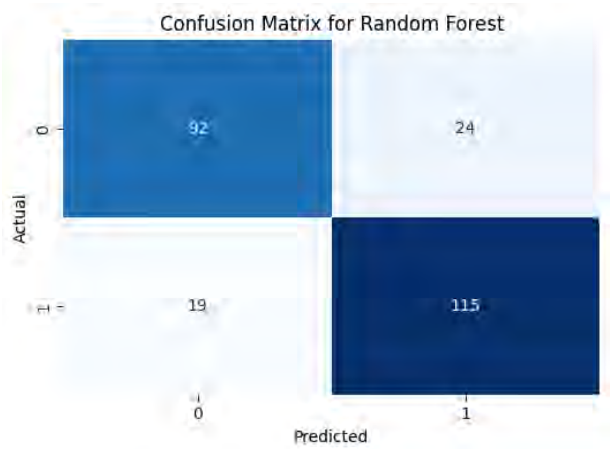

Figure 5.7: Confusion Matrix of EfficientNet

d. The confusion matrix for the best-performing classifier (Handcraft and VGG16 Feature Fusion)

The confusion matrix in Figure 5.8 demonstrates the performance of the XGBoost classifier with handcrafted and VGG16 feature fusion. The model accurately classified 112 negative cases and 124 positive cases, with only 4 false positives and 10 false negatives. This result highlights the strong performance of XGBoost in distinguishing between the two classes, with minimal misclassifications, making it one of the most effective classifiers in this context.

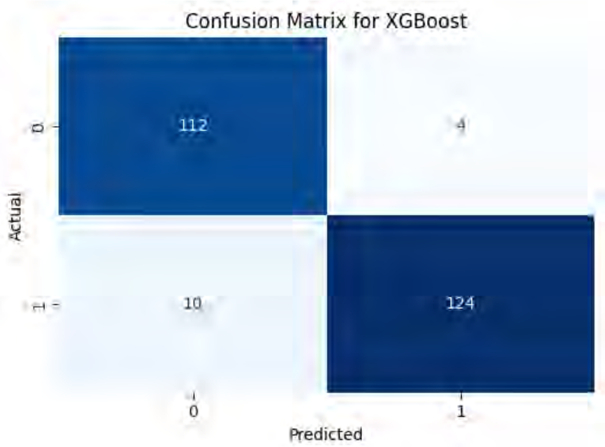

Figure 5.8: Confusion Matrix of VGG16

5.3 Discussion

The comparative analysis of various classifier and feature extraction combinations is performed to show which is the most effective classifier used in Parkinson's disease (PD) versus Healthy controls. A model of the HandCraft and MobileNetV2 feature set shows the highest accuracy of 99% and outperformed other models, demonstrating that MobileNetV2 features provide superior discriminative power when concatenated with handcrafted acoustic features. Handcraft + ResNet50 achieved an accuracy of 85% when run through the Logistic Regression classifier. Following closely, EfficientNETB0 with an 83% accuracy, shows the low robustness of such deep learning architectures in capturing general patterns over mel-spectrogram data. The synergy between feature extraction and the chosen classifier is explored and the performance of the model using the Logistic Regression classifier notably shows strong performance. By contrast, an accuracy of 84% was obtained by the HandCraft + VGG16 (deep layer) feature set and XGBoost classifier, indicating that some deeper layers of VGG16 might not be capturing the most salient features for this particular classification task. Overall, the results show the importance of the selection of appropriate feature extraction methods and classifiers for maximum classification accuracy in this domain.

Chapter 6

Conclusion And Future Directions

6.1 Conclusion

In conclusion, the performance of five classifiers, that is, Random Forest, XGBoost, a Support Vector Machine (SVM), K-Nearest Neighbours, and Logistic Regression, on four hybrid data frames built with handcrafted acoustic features and deep learning-based features obtained from pre-trained convolutional neural networks (CNNs) namely, MobileNetV2, ResNET, EfficientNetB0, and VGG16 is evaluated in this research. The first one was to find the best classifier feature extraction combination to distinguish Parkinson's disease from healthy controls. Although all four machine learning techniques had reasonably high classification accuracy, for example, Logistic Regression performed the best classification accuracy of 99% across the data frames while using the combined feature set of HandCraft and MobileNetV2 features. The findings in these examples highlight that it is crucial to pick the right classifiers and feature extraction techniques to achieve the best result for biomedical data analysis. Overall, these results indicate the value of hybrid feature sets and classifier selection in improving disease classification models.contexts.

6.2 Future Direction

The research in these works could be extended in the future with the inclusion of more advanced attention mechanisms, which would further improve the performance of the model. Selecting which features to attend to has shown to be an effective attention mechanism that could help with the detection of subtle patterns in the hybrid features as extracted from handcrafted and deep learning-based approaches. Attention-based techniques could facilitate more efficient, more precise weighting of features, and thus improve the model's ability to differentiate Parkinson's disease patients from healthy controls. This exploration can also fine-tune attention layers in combination with feature extraction models to make them more interpretable and better classification results for complex biomedical datasets. Further expanding the framework in this way promises to improve diagnostic accuracy and reliability.

Bibliography

[1] Athisakthi, A., & Rani, M. P. (2017, February). Statistical Energy Values and Peak Analysis (SEP) Approach for the Detection of Neurodegenerative Diseases. 2017 World Congress on Computing and Communication Technologies (WCCCT). https://doi.org/10.1109/wccct.2016.66

[2] Mengarelli, A., Tigrini, A., Fioretti, S., & Verdini, F. (2022, December). Identification of Neurodegenerative Diseases From Gait Rhythm Through Time Domain and Time-Dependent Spectral Descriptors. IEEE Journal of Biomedical and Health Informatics, 26(12), 5974–5982. https://doi.org/10.1109/jbhi.2022.3205058

[3] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning. IEEE Access, 8, 147635–147646. https://doi.org/10.1109/access.2020.3016062

[4]]Prashanth, R., Dutta Roy, S., Mandal, P. K., & Ghosh, S. (2016, June). High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning. International Journal of Medical Informatics, 90, 13–21. https://doi.org/10.1016/j.ijmedinf.2016.03.001

[5] Yang, Y., Wei, L., Hu, Y., Wu, Y., Hu, L., & Nie, S. (2021, February). Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning. Journal of Neuroscience Methods, 350, 109019. https://doi.org/10.1016/j.jneumeth.2020.109019

[6] Mathkunti, N. M., Ananthanagu, U., M, E. P. (2024). Brain Disease Parkinson's Diagnosis using VGG-16 and VGG-19 with Spiral and Waves drawings as Input. In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024. https://doi.org/10.1109/i2ct61223.2024.10543635

[7] Costantini, G., Cesarini, V., Di Leo, P., Amato, F., Suppa, A., Asci, F., Pisani, A., Calculli, A., Saggio, G. (2023). Artificial Intelligence-Based Voice Assessment of Patients with Parkinson's Disease Off and On Treatment: Machine vs. Deep-Learning Comparison. Sensors, 23(4), 2293. https://doi.org/10.3390/s23042293

[8] Majhi, B., Kashyap, A., Mohanty, S. S., Dash, S., Mallik, S., Li, A., Zhao, Z. (2024). An improved method for diagnosis of Parkinson's disease using deep learning models enhanced with metaheuristic algorithm. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01335-z

[9] Gunduz, H. (2021). An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification. Biomedical Signal Processing and Control, 66, 102452. https://doi.org/10.1016/j.bspc.2021.102452

[10] Harris, G., Rickard, J. J. S., Butt, G., Kelleher, L., Blanch, R. J., Cooper, J., & Oppenheimer, P. G. (2023). Review: Emerging Eye-Based Diagnostic Technologies for Traumatic Brain Injury. IEEE Reviews in Biomedical Engineering, 16, 530–559. https://doi.org/10.1109/rbme.2022.3161352

[11] Dinesh, K., Xiong, M., Adams, J., Dorsey, R., & Sharma, G. (2016). Signal analysis for detecting motor symptoms in Parkinson's and Huntington's disease using multiple body-affixed sensors: A pilot study. 2016 IEEE Western New York Image and Signal Processing Workshop (WNYISPW). https://doi.org/10.1109/wnyipw.2016.7904834

[12] Karan, B., Sahu, S. S., Orozco-Arroyave, J. R., Mahto, K. (2020). Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech. Biomedical Signal Processing and Control, 61, 102050. https://doi.org/10.1016/j.bspc.2020.102050

[13] Zheng, L., Zhou, C., Mao, C., Xie, C., You, J., Cheng, W., Liu, C., Huang, P., Guan, X., Guo, T., Wu, J., Luo, Y., Xu, X., Zhang, B., Zhang, M., Wang, L., Feng, J. (2024). Contrastive machine learning reveals Parkinson's disease specific features associated with disease severity and progression. Communications Biology, 7(1). https://doi.org/10.1038/s42003-024-06648-x

[14] Manimegalai, R., Ramya, A., Swedha, S. U., Pruthvi, M., Lokesh, S. (2022). Deep Learning Based Approach for Identification of Parkinson's Syndrome. IEEE Xplore Logo. https://doi.org/10.1109/iciiet55458.2022.9967676 https://doi.org/10.1147/jrd.2017.2648700

[15] Yang, H., Ni, J., Gao, J., Han, Z., Luan, T. (2021). A novel method for peanut variety identification and classification by Improved VGG16. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-95240-y

[16] Akay, M., Du, Y., Sershen, C. L., Wu, M., Chen, T. Y., Assassi, S., Mohan, C., Akay, Y. M. (2021). Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model. IEEE Open Journal of Engineering in Medicine and Biology, 2, 104–110. https://doi.org/10.1109/ojemb.2021.3066097

[17] Goutham, V., Sameerunnisa, A., Babu, S., Prakash, T. B. (2022). Brain Tumor Classification using EfficientNet-B0 Model. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). https://doi.org/10.1109/icacite53722.2022.9823526

[18] Liang, J. (2020). Image classification based on RESNET. Journal of Physics Conference Series, 1634(1), 012110. https://doi.org/10.1088/1742-6596/1634/1/012110

[19] Dai, B., Wipf, D. (2019). Diagnosing and enhancing VAE models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1903.05789

[20] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8(1). https://doi.org/10.1186/s40537-021-00444-8

[21] Belgiu, M., Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[22] Zhang, G., Han, K. (2007). Investigating the Performance of Naive-Bayes Classifiers and K- Nearest Neighbor Classifiers. 2007 International Conference on Convergence Information Technology (ICCIT 2007). https://doi.org/10.1109/iccit.2007.148

[23] Yadav, S., Singh, M. K. (2021). Hybrid Machine Learning Classifier and Ensemble Techniques to Detect Parkinson's Disease Patients. SN Computer Science, 2(3). https://doi.org/10.1007/s42979-021-00587-8

[24] Soumaya, Z., Taoufiq, B. D., Benayad, N., Yunus, K., Abdelkrim, A. (2020). The detection of Parkinson disease using the genetic algorithm and SVM classifier. Applied Acoustics, 171, 107528. https://doi.org/10.1016/j.apacoust.2020.107528

[25] Surya, A. T. R., Yaswanthram, P., Nair, P. R., Prasath, S. S. R., Akella, S. V. V. S. (2021). Prediction of Parkinson's Disease Using Machine Learning Models—A Classifier Analysis. In Lecture notes in networks and systems (pp. 453–460). https://doi.org/10.1007/978-981-16-2164-2_35