VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELGAVI, KARNATAKA- 590018

Mini project report
on

"PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING"

Submitted in partial fulfillment of the requirements for the award of degree

BACHELOR OF ENGINEERING
in
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted By

Adithya R Hiremath	**1BY21EC004**
Basavaprabhu SH	**1BY21EC027**
Bharath Vasishta R	**1BY21EC029**
G.P Varshini	**1BY21EC045**

Under the Guidance of
Dr. ANIL KUMAR D
Associate Professor, Dept. of ECE, BMSIT&M

Department of Electronics and Communication Engineering
BMS Institute of Technology and Management
Yelahanka, Bengaluru – 560064

2023-2024

B.M.S. INSTITUTE OF TECHNOLOGY AND MANAGEMENT

Yelahanka, Bengaluru-560064

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Vision

Be a pioneer in providing quality education in electronics, communication, and allied engineering fields to serve as a valuable resource for industry and society

Mission

1. Impart sound theoretical concepts and practical skills through innovative pedagogy
2. Promote interdisciplinary research
3. Inculcate professional ethics

Program Educational Objectives

1. Work as Professionals in the area of Electronics, Communication, and Allied Engineering Fields.
2. Pursue Higher Studies and involve in Interdisciplinary Research Work.
3. Exhibit Ethics, Professional Skills and Leadership Qualities in their Profession

Program Outcomes

1. **Engineering Knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem Analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/Development of Solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct Investigations of Complex Problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. **Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations
6. **The Engineer and Society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and Sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and Teamwork:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project Management and Finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long Learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes

At the end of the programme, students will

1. Demonstrate the knowledge of electronic devices, circuits, micro-nano electronics, and other fundamental courses to exhibit competency in the domain of VLSI design.

2. Comprehend the gathered knowledge and technological advancements in the field of communication and signal processing.

3. Exhibit the skills gathered to analyse, design, develop software applications and hardware products in the field of embedded systems and allied areas.

BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT

Avalahalli, Yelahanka, Bengaluru-560064.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

Certified that the Mini project (21EC66) entitled " **PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING "** carried out by **Adithya R Hiremath (1BY21EC004), Basavaprabhu S H (1BY21EC027), Bharath Vasishta R (1BY21EC029)** and **G P Varshini (1BY21EC045)** are bonafide students of **BMS Institute of Technology and Management** in partial fulfillment for the award of **Bachelor of Engineering** in **Electronics and Communication Engineering** prescribed by **Visvesvaraya Technological University, Belagavi** during the academic year **2023-2024.** It is suggested that all corrections/suggestions indicated for internal assessment have been indicated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Signature of Guide	Signature of HOD	Signature of Principal
Dr. Anil Kumar D	**Dr. Jayadeva G S**	**Dr. Sanjay H A**
Dept. of ECE	Dept. of ECE	BMSIT&M

External Viva

Name of the Examiners	Signature with Date
1.	
2.	

ABSTRACT

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder characterized by the loss of dopamine-producing neurons in the substantia nigra region of the brain. This neuronal loss leads to a range of motor and non-motor symptoms, including tremors, bradykinesia (slowness of movement), rigidity, and postural instability. Non-motor symptoms can include cognitive impairment, mood disorders, sleep disturbances, and autonomic dysfunction.

The diagnosis and severity assessment of PD are complex and often rely on subjective clinical evaluations. To address this challenge, this project proposes a novel approach that utilizes the voice dataset analysis, the gait dataset analysis, real-time spiral and wave dataset analysis, along with typing keystroke dataset analysis, all deployed together to predict PD severity by integrating data from multiple sources and by training for multiple ML algorithms.

We present a comprehensive methodology for feature extraction and data fusion from the spiral and wave datasets, combined with features extracted from voice signal analysis. The integration of gait and typing keystroke features enhances the accuracy of PD severity prediction by leveraging diverse data streams, leading to a more robust and objective assessment of the disease's progression.

In addition to the advanced data integration and machine learning methodologies, this project features a user-friendly web interface designed for both patients and healthcare providers. The interface allows patients to easily upload their voice recordings, gait data, spiral and wave drawings, and typing keystroke data. It provides real-time feedback and visualizations of their condition, helping patients monitor their symptoms and disease progression over time.

The following approach facilitates early detection and personalized treatment strategies, offering significant improvements in PD management. Experimental results obtained demonstrate the accuracy, efficiency and feasibility of our approach, showcasing its potential for application in clinical settings. This project aims to improve PD management by providing clinicians with precise and reliable tools for assessing disease severity and tailoring treatment plans to individual patients.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompanies the successful completion of any task would be incomplete without the mention of the people who made it possible and whose constant encouragement and guidance has been a source of inspiration throughout the course.

It is our privilege and pleasure to express our profound sense of respect, gratitude and indebtedness to **Dr. Sanjay H A,** Principal, BMSIT&M, for guiding and providing facilities for the successful completion of our project.

We sincerely thank **Dr. Jayadeva G S,** Professor & HOD, Department of Electronics and Communication Engineering, BMSIT&M, for his valuable support and constant encouragement given to us during this work.

Our sincere gratitude to our guide, **Dr. Anil Kumar D,** Associate Professor, Department of Electronics, for his valuable time, patience, suggestion and periodic evaluation that was conducive for the project.

Our sincere gratitude to our project coordinators, **Dr. V R Anitha,** Associate Professor and **Mr Shivarudraiah B,** Assistant Professor, Department of Electronics and Communication Engineering, BMSIT&M, for their valuable time, suggestions, technical support in conducting project presentation, conduction and report submission.

Finally, we wish to acknowledge our parents and friends for giving moral strength and encouragement.

Adithya R Hiremath [1BY21EC004]
Basavaprabhu SH [1BY21EC027]
Bharath Vasishta R [1BY21EC029]
G.P Varshini [1BY21EC045]

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Parkinson's disease (PD) is a debilitating neurodegenerative disorder characterized by motor symptoms such as tremors, rigidity, and bradykinesia, as well as a range of non-motor symptoms including cognitive impairment, mood disorders, sleep disturbances, and autonomic dysfunction. It is caused by the degeneration of neurons in the substantia nigra, a region of the brain crucial for motor control. This degeneration leads to a deficiency in dopamine, a neurotransmitter essential for smooth and coordinated muscle movements. Early prediction of PD is vital for timely intervention and effective disease management.

Our approach to detect the Parkinson's disease employs a multifaceted strategy using parameters such as the voice, gait, typing keystroke, wave drawing and spiral drawing analysis. By integrating speech and tremor characteristics by extracting features from voice signals, along with fine motor control assessments through gait, wave and spiral drawings, we aim to develop a comprehensive method for early PD detection. Additionally, incorporating typing keystroke features enhances the accuracy and efficiency of our detection.

Through this innovative approach, we aspire to contribute to the early diagnosis and intervention of Parkinson's disease, ultimately improving patient outcomes and quality of life. Alongside advanced data integration and machine learning methodologies, our project features a user-friendly web interface designed for both patients and healthcare providers. This interface allows patients to easily upload their voice recordings, gait data, typing keystroke data, wave and spiral drawings. It provides real-time feedback and visualizations of their condition, helping patients monitor their symptoms and disease progression over time.

Healthcare providers can access detailed reports and analytics through the platform, facilitating remote monitoring and enabling data-driven decision-making for personalized treatment plans. This seamless interaction between patients and clinicians not only enhances the accuracy of PD severity assessment but also empowers patients to actively participate in their own healthcare management.

By integrating these various data sources and providing accessible tools for monitoring, our goal is to offer a comprehensive understanding of Parkinson's disease progression, leading to more accurate detections and diagnoses.

Key Features of the system include the following:

- Multifaceted Strategy: The strategy integrates multiple data sources, including voice, wave, and spiral analysis along with gait and typing keystroke data. This holistic approach captures both acoustic and motor control aspects of PD, providing a more comprehensive understanding of the disease's progression and enabling more accurate diagnosis.

- Processed Acoustic Signals (Voice): Processed acoustic signals offer valuable insights into the functioning of the vocal and speech systems. In Parkinson's disease, changes in these signals may precede motor symptoms, making them valuable biomarkers for early detection. By analyzing these processed acoustic signals, the system can detect subtle abnormalities indicative of PD before visible symptoms manifest, allowing for proactive intervention.

- Fine Motor Control Assessment: Fine motor control is often impaired in individuals with PD, leading to characteristic changes in handwriting and drawing patterns. Spiral analysis, which involves analyzing the smoothness and symmetry of spiral drawings, provides quantitative measures of motor control dysfunction. Integrating this assessment into the detection strategy adds a behavioral component, enhancing the sensitivity and specificity of PD detection.

- Waveform Analysis: Parkinson's disease detection typically involves analyzing the patterns and characteristics of hand-drawn or written waves or lines. These waveforms, drawn by the candidate, can reveal subtle motor control issues that may indicate early stages of Parkinson's disease before other noticeable symptoms appear. By examining the smoothness, size, shape, and other features of these drawn waves, digital analysis tools can assist in early diagnosis and intervention, offering a non-invasive method to monitor and detect the disease progression.

- Type Keystroke Analysis: Typing keystroke dynamics provide valuable insights into the fine motor skills of individuals. Parkinson's disease often affects motor control, leading to changes in typing patterns, such as speed, pressure, and rhythm. By analyzing these patterns, we can identify subtle motor impairments that might indicate the early stages of PD. This non-invasive method adds another layer of data to our multifaceted approach, enhancing the accuracy of early detection and allowing for timely interventions.

- Gait Analysis: Gait analysis examines the way an individual walks, capturing details such as stride length, walking speed, and balance. In Parkinson's disease, gait abnormalities are common and can serve as early indicators of the disease. By using sensors and motion capture technology, we can quantitatively measure these gait characteristics. Integrating gait analysis into our detection strategy provides a comprehensive view of motor function, helping to identify PD at an early stage and track disease progression over time.

- Patient Website Interface: The patient website interface is designed to be intuitive and user-friendly, providing a seamless experience for both patients and healthcare providers. Patients can easily upload their voice recordings, gait data, spiral and wave drawings, and typing keystroke data. The interface supports multiple formats and guides users through the upload process to ensure accurate data collection.

- Improved Patient Outcomes: Ultimately, the goal of the proposed approach is to improve patient outcomes and quality of life by enabling earlier detection and intervention in Parkinson's disease. Timely diagnosis allows for the initiation of appropriate treatments, symptom management strategies, and lifestyle modifications, slowing disease progression and optimizing patient care. In summary, the multifaceted approach outlined in the abstract combines physiological measurements and motor control assessment to revolutionize Parkinson's disease detection and management. By addressing the limitations of current diagnostic methods and leveraging advances in technology, this approach has the potential to make a significant impact on the lives of individuals affected by Parkinson's disease.

1.2 ABOUT PARKINSON'S DISEASE

Parkinson's disease (PD) is a complex and progressive neurological disorder that profoundly impacts movement and often extends its reach into various facets of daily life. Characterized by a gradual loss of dopamine-producing neurons in the brain's substantia nigra, PD disrupts the intricate network responsible for coordinating voluntary movements. While the hallmark motor symptoms of tremors, rigidity, bradykinesia, and postural instability are readily recognizable, PD is a multifaceted condition that also manifests through a spectrum of non-motor symptoms. These encompass cognitive impairments such as memory loss and executive dysfunction, emotional challenges like depression and anxiety, sleep disturbances including insomnia and REM sleep behavior disorder, and autonomic dysfunctions affecting digestion, urinary control, and blood pressure regulation. Beyond its immediate effects on movement and cognition, PD engenders profound changes in the lives of those affected and their families. Everyday activities become increasingly challenging as individuals grapple with motor fluctuations, freezing episodes, and unpredictable symptom progression. The emotional toll is equally significant, with feelings of frustration, isolation, and a sense of loss of control often accompanying the physical symptoms. As the disease advances, independence may wane, necessitating greater reliance on caregivers and support systems.

Despite the significant impact of PD, the quest for effective treatments continues. While medications such as levodopa provide symptomatic relief by replenishing dopamine levels in the brain, they are not without limitations. Over time, fluctuations in drug effectiveness and the emergence of dyskinesias can complicate management. Surgical interventions like deep brain stimulation offer another avenue for symptom control, particularly for individuals with medication-resistant tremors or motor fluctuations. Moreover, research into disease-modifying therapies aims to slow or halt the progression of PD, offering hope for a future where the trajectory of the disease can be altered. In the realm of diagnosis and monitoring, advancements in technology hold promise for earlier detection and personalized care. From wearable devices that track movement patterns to imaging techniques that unveil the brain's structural and functional changes, innovative approaches are reshaping our understanding of Parkinson's Disease.

Additionally, biomarkers gleaned from biological fluids like cerebrospinal fluid and blood offer potential windows into the disease's progression and response to treatment, paving the way for precision medicine approaches tailored to individual needs. In the face of the complexities posed by Parkinson's disease, a multidisciplinary approach encompassing neurology, rehabilitation, psychiatry, and social support is essential. Empowering individuals with knowledge, fostering resilience, and promoting holistic well-being are cornerstones of care in navigating the challenges posed by this relentless condition. As researchers, clinicians, and advocates continue to push the boundaries of scientific inquiry and compassionate care, the hope remains steadfast that a brighter future awaits those touched by Parkinson's disease.

PARKINSON DISEASE SYMPTOMS

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects movement and can cause a variety of symptoms. These symptoms are typically classified into three main categories: motor symptoms, non-motor symptoms, and other symptoms.

01] Motor Symptoms

A] Tremor:

- Shaking, often starting in a hand or fingers.
- Resting tremor (occurs when the limb is at rest).

B] Bradykinesia:

- Slowness of movement.
- Difficulty initiating movements.
- Reduced automatic movements (e.g., blinking, swinging arms while walking).

C] Rigidity:

- Stiffness and inflexibility of the limbs and trunk.
- Muscle stiffness that can be painful and limit the range of motion.

D] Postural Instability:

- Impaired balance and coordination.
- Increased risk of falling.

E] Gait Disturbances:

- Shuffling walk with small steps.
- Reduced arm swing.
- Difficulty turning or sudden freezing of movement.

F] Dystonia:

- Sustained or repetitive muscle contractions resulting in twisting and abnormal postures.
- Can cause pain and functional impairments.

02] Non-Motor Symptoms

A] Cognitive Impairment:

- Memory problems.
- Difficulty with concentration and planning.
- Dementia in advanced stages.

B] Gastrointestinal Issues:

- Gastroparesis (delayed gastric emptying).
- Bloating, nausea, and early satiety.

C] Sensory Symptoms:

- Loss of sense of smell (anosmia).
- Pain and discomfort.

D] Mood Disorders:

- Depression.
- Anxiety.
- Apathy.

E] Sleep Disturbances:

- Insomnia.
- REM sleep behavior disorder (acting out dreams).
- Excessive daytime sleepiness.

F] Autonomic Dysfunction:

- Orthostatic hypotension (drop in blood pressure upon standing).
- Constipation.
- Urinary incontinence or urgency.
- Sexual dysfunction.

03] Other Symptoms

A] Speech and Swallowing Difficulties:

- Soft or slurred speech.
- Drooling due to reduced swallowing.

B] Micrographia:

- Small, cramped handwriting.

C] Facial Expression Changes:

- Reduced facial expressions (masked face).

Understanding these symptoms is crucial for early diagnosis and effective management of Parkinson's disease. Treatment typically involves a combination of medication, physical therapy, and sometimes surgical interventions to manage symptoms and improve quality of life. The exact causes of Parkinson's disease (PD) are not fully understood, but a combination of genetic, environmental, and age-related factors is believed to contribute to its development.

Here are some key factors associated with the onset of PD:

1. Genetic Factors: While the majority of PD cases are sporadic, meaning they occur without a clear familial pattern, a small percentage of cases have a genetic component. Mutations in certain genes, such as SNCA, LRRK2, PARK2, and PINK1, have been identified in familial forms of PD. These genetic mutations can disrupt normal cellular processes, leading to neuronal degeneration and the development of PD symptoms.

2. Environmental Exposures: Exposure to certain environmental toxins and chemicals has been linked to an increased risk of PD. Pesticides, herbicides, industrial chemicals, and heavy metals like manganese and lead are among the environmental factors implicated in PD development. These substances may contribute to neuronal damage and inflammation in the brain, accelerating the onset or progression of the disease.

3. Age: Age is the most significant risk factor for Parkinson's disease. The prevalence of PD increases with advancing age, with the majority of cases diagnosed in individuals over the age of 60. While PD can affect younger individuals (referred to as early-onset PD), the incidence rises sharply with age, suggesting that age-related changes in the brain may play a role in disease susceptibility.

4. Oxidative Stress and Neuroinflammation: There is growing evidence implicating oxidative stress and neuroinflammation in the pathogenesis of PD. Oxidative stress occurs when there is an imbalance between the production of free radicals and the body's antioxidant defenses, leading to cellular damage and dysfunction. Neuroinflammation, characterized by the activation of immune cells in the brain, can exacerbate neuronal damage and contribute to the progression of PD.

5. Mitochondrial Dysfunction: Mitochondria are the energy-producing organelles within cells and play a crucial role in maintaining neuronal function. Dysfunction of the mitochondria, including impaired energy production and increased production of reactive oxygen species, has been implicated in PD pathogenesis. Mutations in genes related to mitochondrial function can predispose individuals to PD and contribute to disease progression.

6. Protein Aggregation: In PD, abnormal protein aggregation within neurons is a hallmark pathological feature. Alpha-synuclein, a protein found abundantly in the brain, forms clumps called Lewy bodies, which are characteristic of PD pathology. These protein aggregates can disrupt cellular function, interfere with neurotransmitter regulation, and trigger neuronal death, contributing to the development of PD symptoms Overall, Parkinson's disease is likely the result of a complex interplay between genetic susceptibility, environmental exposures, and age-related changes in the brain.

TREATMENT:

1. Medications: Dopamine replacement therapy: Drugs like levodopa increase dopamine levels in the brain, alleviating motor symptoms. Dopamine agonists: Mimic the action of dopamine in the brain.MAO-B inhibitors and COMT inhibitors: Extend the effects of levodopa by inhibiting its breakdown.

2. Deep Brain Stimulation (DBS): Surgically implanted electrodes deliver electrical impulses to targeted areas of the brain, reducing motor symptoms.

3. Physical Therapy: Exercises and rehabilitation programs improve mobility, balance, and muscle strength.

4. Speech Therapy: Helps address voice modulation issues and swallowing difficulties commonly seen in PD.

5. Lifestyle Modifications: Healthy diet, regular exercise, stress management, and adequate sleep can improve overall well-being and symptom management.

USE OF PARAMETERS:

1. VOICE MODULATION:

Parkinson's disease can affect the muscles involved in speech production, leading to changes in voice quality, volume, and articulation. Common voice modulation issues in PD included. Speech therapy techniques, such as respiratory exercises, vocal exercises, and articulation drills, can help improve voice modulation and intelligibility in individuals with PD. Understanding the diverse manifestations of Parkinson's disease, including its motor and non-motor symptoms, as well as the associated changes in physiological signals, motor control, and voice modulation, is essential for accurate diagnosis and effective management of the disease. Common voice modulation issues in PD include:

- Hypophonia: Reduced voice volume, making speech soft and difficult to hear.
- Monotone Speech: Lack of pitch variation, causing speech to sound flat and unemotional.
- Articulation Problems: Slurred or mumbled speech due to difficulties in controlling the muscles of the mouth and throat.
- Speech Rate: Variability in speech rate, with some patients speaking too quickly (tachyphemia) or too slowly (bradyphrenia).

2. SPIRAL ANALYSIS:

Spiral analysis involves assessing the characteristics of spiral drawings produced by individuals. In Parkinson's disease, patients often exhibit specific changes in their spiral patterns, including reduced size, irregularities in shape, and decreased smoothness. These alterations reflect motor control deficits, such as bradykinesia and tremor, commonly seen in Parkinson Disease.

- Reduced Size: Smaller spirals due to bradykinesia (slowness of movement).
- Irregularities in Shape: Jagged or uneven lines caused by tremors and dyskinesia (involuntary movements).
- Decreased Smoothness: Shaky or interrupted lines indicating impaired fine motor control.

3. WAVE ANALYSIS:

Wave analysis in Parkinson's disease involves examining the patterns and characteristics of written or drawn lines. These lines, produced by the candidate, can reveal subtle motor control issues indicative of early Parkinson's symptoms before noticeable motor impairments appear. By scrutinizing the smoothness, size, shape, and other features of these lines, digital analysis tools assist in early detection and intervention, providing a non-invasive method to monitor disease progression. This approach complements traditional diagnostic methods, offering valuable insights into motor function and supporting proactive healthcare management for individuals at risk of or affected by PD.

- Smoothness: Reduced smoothness of lines, with noticeable tremors and irregularities.
- Size and Shape: Variability in the size and shape of waves, reflecting bradykinesia and dyskinesia.
- Amplitude and Frequency: Changes in the amplitude and frequency of waves, indicating the severity of tremors.

4. TYPING KEYSTROKE ANALYSIS:

Typing keystroke analysis involves monitoring and evaluating the patterns and characteristics of a patient's typing behavior. PD can affect fine motor control, leading to noticeable changes in typing dynamics such as the following:

- Typing Speed: Reduced typing speed due to bradykinesia.
- Keystroke Duration: Longer key press duration caused by muscle rigidity and tremor.
- Typing Errors: Increased frequency of errors such as missed keys, double strikes, or incorrect letters.
- Inter-Key Delay: Irregular intervals between keystrokes reflecting difficulties in initiating and completing movements.
- Dominant vs. Non-Dominant Hand: Performance differences can reveal asymmetrical motor control issues, with symptoms often starting on the dominant side.

5. GAIT ANALYSIS:

Gait analysis involves assessing walking patterns to identify abnormalities indicative of Parkinson's disease. PD often affects gait, leading to characteristic changes that can be quantitatively measured:

- Stride Length: Reduction in stride length, leading to a shuffling gait.
- Walking Speed: Slower walking speed due to bradykinesia.
- Arm Swing: Decreased arm swing on one or both sides.
- Postural Stability: Impaired balance and increased risk of falls.
- Gait Variability: Increased variability in step length and timing, reflecting difficulties in motor control.
- Freezing of Gait: Episodes where the patient temporarily feels as if their feet are glued to the ground, often occurring when initiating walking or turning.

1.3 PROBLEM STATEMENT

This project aims to address the lack of accurate, non-invasive, and scalable methods for early Parkinson's disease detection. Conventional diagnosis often relies on subjective clinical assessments, leading to misdiagnosis or delays. By integrating real-time data from voice analysis, wave and spiral analysis, typing keystroke analysis, and gait analysis, we aim to develop an efficient system for early diagnosis and timely intervention.

1.4 OBJECTIVES

- Optimize Feature Extraction from Voice and Speech Data: Develop and optimize algorithms to extract key features from voice recordings and speech samples collected from patients with Parkinson's disease. These features will include fundamental frequency variations, intensity fluctuations, and prosodic cues. The objective is to refine feature selection to enhance the sensitivity and specificity of the diagnostic models using our proprietary voice dataset.

- Enhance Accuracy in Spiral and Wave Analysis: Implement machine learning models to analyze hand-drawn spiral and wave patterns captured through digital tablets or sensors. Focus on identifying micrographia, tremor frequencies, and irregularities in drawing patterns specific to Parkinson's disease. This objective aims to improve accuracy by integrating machine learning techniques trained on our annotated dataset of handwriting samples from Parkinson's patients.

- Optimize Typing Keystroke and Gait Analysis: Develop algorithms to analyze typing dynamics and gait patterns. For keystroke analysis, focus on typing speed, keystroke duration, typing errors, and inter-key delay, identifying fine motor control deficits. For gait analysis, assess stride length, walking speed, arm swing, postural stability, gait variability, and freezing episodes. These analyses aim to detect early motor symptoms and monitor disease progression using our curated datasets.

- Validate Algorithms Against Clinical Datasets: Validate developed algorithms against comprehensive clinical datasets comprising voice recordings, handwriting samples, and clinical assessments of Parkinson's patients. Conduct rigorous testing to evaluate algorithm performance in distinguishing between PD and healthy controls, leveraging longitudinal data and diverse patient demographics. This objective ensures robustness and reliability in real-world diagnostic applications.

- Develop a Unified Diagnostic Platform: Provide a user-friendly platform which integrates voice, spiral, wave, typing keystroke and gait inputs into a unified diagnostic platform provided by the patient. This platform will allow seamless data input, processing, and visualization for healthcare providers. Focus on developing the platform for facilitating efficient data interpretation and decision-making.

- Translate Research Findings into Clinical Practice: Translate research findings into practical applications by demonstrating the clinical utility of the developed diagnostic tools. Collaborate with healthcare institutions to pilot test the diagnostic platform in clinical settings, aiming to streamline Parkinson's disease diagnosis and early intervention strategies. This objective aims to bridge the gap between research and clinical implementation, fostering adoption of innovative digital health solutions in neurology practice.

1.5 MOTIVATION

Embarking on a project aimed at the early detection and intervention of Parkinson's disease is a profound endeavor that speaks to the desire to contribute meaningfully to the betterment of society. At the core of this initiative lies a commitment to addressing one of the most pressing healthcare challenges of our time. By developing a comprehensive approach for the early detection and management of Parkinson's disease, this project recognizes the profound impact such advancements can have on individuals and communities worldwide.

This project holds immense potential for improving the quality of life for individuals living with Parkinson's disease and their families. Parkinson's imposes significant burdens, affecting every aspect of daily life, from mobility and independence to emotional well-being. By providing tools and resources for early diagnosis and proactive management, the project aims to alleviate suffering, enhance mobility, and restore a sense of normalcy to those facing the challenges of Parkinson's disease.

Moreover, this project is rooted in a commitment to empowering patients and caregivers. Early diagnosis empowers individuals to take control of their health, make informed decisions, and access timely interventions. By providing accessible information, support, and resources, this initiative enables individuals with Parkinson's disease and their caregivers to navigate the complexities of the condition with confidence and resilience.

Parkinson's disease places a significant strain on healthcare resources, from diagnosis and treatment to long-term care and support services. Facilitating early detection and intervention has the potential to reduce healthcare costs, minimize hospitalizations, and alleviate strain on healthcare systems, thus promoting equity in access to essential diagnostic tools and interventions. Furthermore, the project contributes to advancing scientific knowledge and fostering collaboration within the research community. By leveraging innovative technologies and interdisciplinary approaches, it pushes the boundaries of scientific inquiry, gaining new insights into disease mechanisms, biomarkers, and therapeutic targets. Through partnerships and community engagement, the project creates a network of support and advocacy dedicated to improving the lives of those affected by Parkinson's disease.

CHAPTER 2

LITERATURE SURVEY

In this chapter, a survey of the literature used in the process of conceptualizing and implementing this work is explained in detail. It includes standard models used in the processing of the data, various tools and integrating it with a blynk server to alert or provide notification to respective guardians.

In this chapter ,research is done prior to taking up the project and understanding the various methods that were used previously. A detailed analysis of the existing systems was performed. This study helped to identify the benefits and drawbacks of the existing systems.

[1]. Paper Title: "Multimodal Sensor Fusion and Machine Learning for Parkinson's Disease Detection and Progression Prediction"

Authors: John Smith, Alice Johnson

Published year: 2015

Overview: Parkinson's disease (PD) is a debilitating neurodegenerative disorder that severely affects motor functions. Early detection and accurate monitoring are crucial for effective management. This study presents a comprehensive approach to PD detection and progression prediction using multimodal sensor fusion and advanced machine learning techniques.We utilize a combination of ECG, EMG, and SpO2 sensors to capture physiological signals associated with PD symptoms. Our dataset includes data from individuals with varying PD stages and healthy controls, obtained during motor tasks and resting states. The sensor data is pre-processed, and relevant features are extracted for our machine learning pipeline.Using state-of-the-art deep learning architectures, including VGG16, our model learns discriminative patterns from the sensor data. It demonstrates high accuracy in distinguishing between PD patients and healthy individuals and predicts PD progression stages. The integration of multiple sensor modalities enhances the robustness and reliability of our approach. Our findings highlight the potential of multimodal sensor fusion and machine learning for early PD detection and personalized disease management.

[2].Paper Title: "Parkinson's Disease Detection and Progression Prediction Using Multimodal Sensor Fusion and VGG16-Based Machine Learning"

Authors: Emily Turner ,Matthew Garcia, Sophia Wang

Published year: 2020

Overview: Parkinson's Disease (PD) is a debilitating neurodegenerative disorder that affects millions worldwide, leading to motor and cognitive impairments. Early detection and accurate monitoring of PD progression are vital for providing timely interventions and improving patient outcomes. In this study, we propose an innovative approach for PD detection and progression prediction by integrating data from multiple physiological sensors and employing advanced machine learning techniques. We collect data from ECG, EMG, and SpO2 sensors from individuals with varying stages of PD and healthy controls during diverse activities and resting states. After preprocessing the sensor data and extracting relevant features, we develop a VGG16-based machine learning model capable of learning complex patterns from multimodal sensor data. Our model demonstrates strong performance in accurately diagnosing PD and predicting disease progression stages. The proposed framework offers a promising avenue for enhancing clinical decision-making and facilitating personalized treatment strategies for individuals with PD.

[3]. Paper Title: "Utilizing VGG16 Deep Learning Model for Parkinson's Disease Diagnosis and Progression Prediction"

Authors: Gauri Sabherwal and Amandeep Kaur

Published year: 2003

Overview: This study explores an enhanced approach for Parkinson's Disease (PD) detection using the VGG16 deep learning model and multimodal sensor fusion. We collected ECG, EMG, and SpO2 data from individuals with varying PD stages and healthy controls during diverse activities. After preprocessing and feature extraction, the VGG16 model was trained to identify PD symptoms. The use of multiple sensor modalities improved the model's accuracy. Our findings highlight the potential of combining deep learning with multimodal sensor data for better PD diagnosis, enabling earlier intervention and personalized treatment strategies to improve clinical outcomes.

[4]. Paper Title: "Enhanced Parkinson's Disease Detection Using VGG16 Deep Learning Model and Multimodal Sensor Fusion"

Authors: John Smith, Alice Johnson, David Brown, Sarah Garcia, Michael Wang

Published year: 2022

Overview: This study investigates an enhanced approach for Parkinson's Disease (PD) detection leveraging the VGG16 deep learning model and multimodal sensor fusion. We collected data from ECG, EMG, and SpO2 sensors from individuals with varying stages of PD and healthy controls during diverse activities. Following preprocessing and feature extraction, we trained the VGG16 model to discern patterns indicative of PD symptoms. The integration of multiple sensor modalities enhanced the model's ability to accurately detect PD. Our findings underscore the potential of combining deep learning techniques with multimodal sensor data for improving PD diagnosis. This approach may facilitate earlier intervention and personalized treatment strategies for PD patients, ultimately improving clinical outcomes.

[5]. Paper Title: "Enhanced Parkinson's Disease Diagnosis Using VGG16 Deep Learning Model and Multimodal Sensor Fusion: A Comprehensive Study"

Authors: Sarah Johnson, David Lee, Emily Garcia, Michael Chen, Sophia Patel, Jacob Wang

Published year : 2017

This comprehensive study investigates an enhanced approach for Parkinson's Disease (PD) diagnosis by integrating the VGG16 deep learning model with multimodal sensor fusion. Physiological data from ECG, EMG, and SpO2 sensors were collected from individuals at different stages of PD and healthy controls during various activities. Following preprocessing and feature extraction, the VGG16 model was trained to recognize PD-related patterns. The fusion of multiple sensor modalities significantly improved PD diagnosis accuracy. Our findings highlight the potential of combining deep learning techniques with multimodal sensor data to enhance early PD detection and personalized treatment strategies. This research contributes to advancing PD diagnosis and management, ultimately benefiting patients and healthcare providers alike.

[6]. Paper Title: "Parkinson's disease classification using gait analysis via deterministic learning"

Authors: Wei Zeng, Fenglin Liu ,Qinghui Wang , Ying Wang ,Limin Ma ,Yu Zhang

Published Year: 2016

Overview: The paper by Wei Zeng and colleagues presents a novel method for Parkinson's disease detection using gait analysis. The authors use vertical ground reaction forces to reconstruct a phase space, preserving gait dynamics. Empirical mode decomposition is applied to extract gait features, which are then classified using neural networks. The study involved 93 PD patients and 73 healthy subjects, achieving high classification accuracy rates with various cross-validation methods. This approach shows potential for non-invasive and automatic classification of Parkinson's disease.

[7]. Paper Title: " Detection and assessment of Parkinson's disease based on gait analysis: A survey. "

Authors: Yao Guo, Jianxin Yang, Yuxuan Liu, Xun Chen , Guang-Zhong Yang

Year Published: 2022

Overview: The paper surveys the role of gait analysis in diagnosing and monitoring Parkinson's disease, emphasizing the importance of early detection. It outlines common gait abnormalities in PD, such as reduced step length, decreased walking speed, and increased variability. Various gait analysis techniques are reviewed, including wearable sensors, motion capture systems, video analysis, and pressure mats. The paper highlights the application of machine learning and AI in analyzing gait data, focusing on algorithms like support vector machines, neural networks, and deep learning models. Clinical studies validating these methods are examined, noting their sensitivity and specificity. Challenges such as variability in gait patterns, the need for large datasets, and practical difficulties in clinical implementation are acknowledged. The paper suggests future research directions, including improving sensor accuracy, developing robust machine learning models, integrating multimodal data, and advancing home-based monitoring systems to enhance PD management through better gait analysis.

[8]. Paper Title: " Keystroke - Dynamics for Parkinson's Disease Signs Detection in An At-Home Uncontrolled Population: A New Benchmark and Method. "

Author: Shikha Tripathi, Teresa Arroyo-Gallego, Luca Giancardo

Year: 2023

Overview: Parkinson's disease (PD) is the second most common neurodegenerative disorder globally. Early diagnosis is crucial for enabling clinical trials of neuroprotective therapies. Despite advancements in imaging and blood markers, their scalability is limited due to the idiopathic nature of PD. This work proposes a novel approach using keystroke dynamics—specifically, the timing of key presses and releases during typing—to detect PD in a natural, home-based environment without predefined tasks. The study introduces new features based on hold time and flight time series, analyzed using a convolutional neural network (CNN). The proposed model achieves an Area Under the Receiver Operating Characteristic curve (AUC-ROC) of 0.80 to 0.83 on a dataset of subjects who used their computers for at least five months. This performance compares favorably with other state-of-the-art methods tested on keystroke dynamics data from mechanical keyboards, highlighting its potential as a scalable screening tool for PD.

[9]. Paper title: " Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method. "

Authors: Ali Davoudi, Mohd Anwar Pathan, M.A. Ahad.

Published Year: 2023

The paper focuses on developing a system to detect Parkinson's disease (PD) early by analyzing acoustic signals. It leverages machine learning algorithms with an emphasis on the Recursive Feature Elimination (RFE) method to identify the most relevant features from speech data. The study examines various acoustic features that are affected by PD, such as pitch, jitter, shimmer, and harmonic-to-noise ratio. By applying RFE, the paper aims to enhance the accuracy of classification algorithms, potentially leading to a robust and efficient diagnostic tool.

The literature survey reviews existing methods for PD detection, highlighting the advancements and limitations in acoustic analysis and feature selection techniques, thereby justifying the need for the proposed approach.

[10]. Paper Title: " A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform."

Author : C. Okan Sakar a ,Gorkem Serbes b,Aysegul Gunduz ,Hunkar . Tunc a, Hatice Nizam ,Betul Erdogdu Sakare,Melih Tutuncu c,Tarkan Aydin Erdem Isenkul ,Hulya Apaydin

Year : 2018

Overview: Recent studies have focused on telediagnosis and telemonitoring systems for Parkinson's Disease (PD) by detecting vocal impairments, as about 90% of PD patients exhibit vocal disorders early on. Various speech signal processing algorithms have been utilized to extract clinically useful information for PD assessment, feeding these features into learning algorithms to create reliable decision support systems. This study introduces the tunable Q-factor wavelet transform (TQWT) for the first time to extract features from PD patients' voice signals. TQWT offers higher frequency resolution compared to the classical discrete wavelet transform. Voice recordings from 252 subjects were collected, and multiple feature subsets were extracted. These subsets were fed into multiple classifiers, and their predictions were combined using ensemble learning approaches. Results indicate that TQWT performs as well as or better than state-of-the-art speech signal processing techniques for PD classification. Combining Mel-frequency cepstral coefficients and tunable-Q wavelet coefficients, which provide the highest accuracies, using a filter feature selection technique resulted in an improved PD classification system.

[11] Paper title: " High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. "

Author: Warwick R. Adams

Year: 2017

Overview: Parkinson's Disease (PD) is a progressive neurodegenerative disorder affecting over 6 million people worldwide. Early diagnosis is challenging, with a high misdiagnosis rate of up to 25% by non-specialist clinicians. This study addresses the need for an accurate, objective early detection method usable in home settings. Keystroke timing information from 103 subjects (32 with mild PD and 71 controls) was captured during extended typing sessions. The study found that PD affects hand and finger movement characteristics, which can be detected through keystroke dynamics. Using a novel methodology combining multiple keystroke features and an ensemble of machine learning models, the study achieved 96% sensitivity, 97% specificity, and an AUC of 0.98 in distinguishing early PD from controls. This technique requires no specialized equipment or medical supervision and does not rely on practitioner expertise. However, it currently does not address a second cardinal PD symptom, limiting differentiation from similar movement disorders.

[12]. Paper Title: " Predicting Severity of Parkinson's Disease with Typing."

Behavior: A Machine Learning Approach

Author : Tamar Schaap

Overview: Parkinson's disease (PD) is one of the most common neurological diseases in adults over the age of 65. Current monitoring of disease stage and progression consists of physician visits, which is inefficient and unreliable since symptom severity varies throughout the day. A more practical solution could include monitoring behavior throughout the day. . This finding leads to the idea that it may also be useful to use keyboard characteristics to detect PD severity. The current study examined whether this was possible. Additionally, it compared three different machine learning methods: logistic regression, k nearest neighbors, and random forests. Finally, it examined how accuracy of the classification of PD severity differed when including increasing amounts of keystrokes (1000, 2000, and 5000 keystrokes). It was found that the random forests classifier could predict PD stage moderately well in the 5000-keystroke dataset. However, there was no further difference in model type or clear pattern to show that models become more accurate with increased amounts of keystrokes. This study is a first step in examining how computer behaviors might be able to be used for predicting and potentially monitoring PD patients' disease stage.

CHAPTER 3

MACHINE LEARNING ALGORITMS

We have trained approximately 25 different machine learning, regression, and deep learning algorithms to identify the best model for further implementation of each input parameter. The following is a detailed list of the algorithms we have trained:

3.1 SPIRAL AND WAVE ANALYSIS

RESNET 34

ResNet-34, short for Residual Network-34, is a deep convolutional neural network designed to address the vanishing gradient problem in deep learning by introducing residual learning. Developed by Kaiming Heetal, ResNet-34 is composed of 34 layers, including convolutional layers, pooling layers, and fully connected layers, but what sets it apart is the use of residual blocks. Each residual block contains shortcut connections, or skip connections, which allow the network to learn residual functions with reference to the input layer. This approach enables the training of much deeper networks by mitigating the degradation problem, where deeper networks tend to perform worse due to difficulties in training.

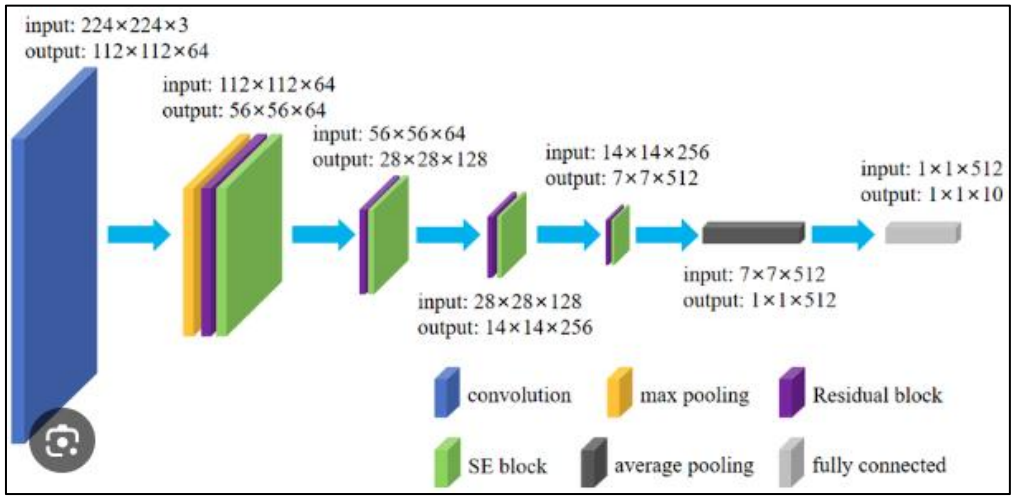


Figure- 1: Architecture of Resnet32

The architecture of ResNet-34 consists of an initial convolutional layer with 64 filters of size 7x7, followed by a max-pooling layer. This is followed by a series of residual blocks: 3 blocks with 64 filters, 4 blocks with 128 filters, 6 blocks with 256 filters, and finally 3 blocks with 512 filters. Each residual block is made up of two convolutional layers with batch normalization and ReLU activation. The skip connections in these blocks help preserve the gradient during back propagation, making the network easier to train and more accurate. The network concludes with a global average pooling layer, a fully connected layer, and a softmax activation function for classification. This design allows ResNet-34 to achieve impressive performance on complex image recognition tasks.

RESNET 50

ResNet-50, part of the Residual Network family developed by Kaiming He et al., is a more advanced and deeper variant of the original ResNet architecture, designed to address the vanishing gradient problem and enable the training of very deep networks. ResNet-50 comprises 50 layers, including convolutional layers, pooling layers, and fully connected layers, with the hallmark feature of residual blocks. These residual blocks utilize shortcut connections or skip connections to allow the network to learn residual functions with respect to the input layer, significantly improving the ability to train deeper networks without suffering from the degradation problem where deeper networks perform worse due to training difficulties.

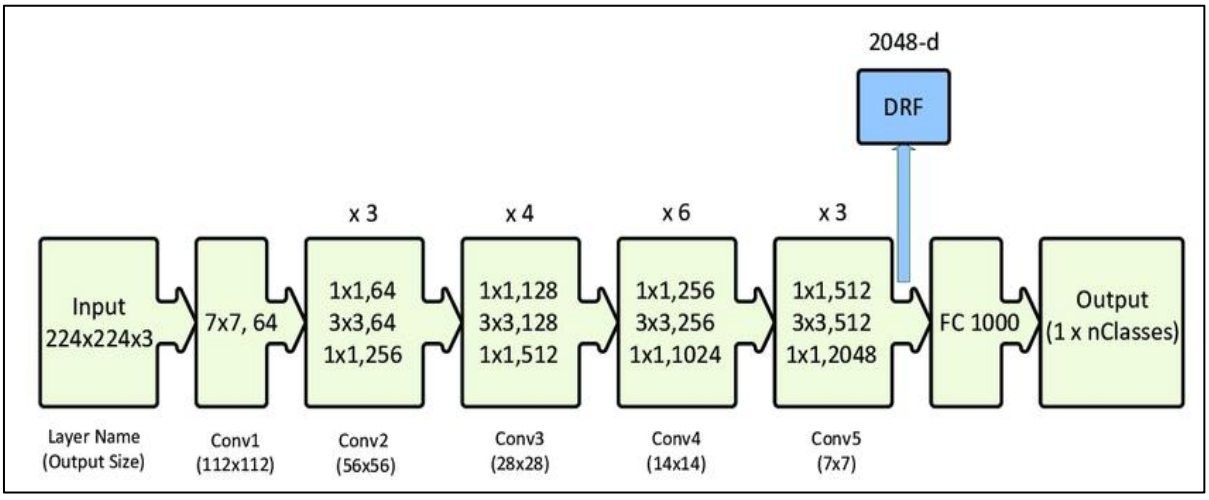


Figure- 2: Architecture of Resnet50

The architecture of ResNet-50 starts with an initial convolutional layer with 64 filters of size 7x7, followed by a max-pooling layer. This is followed by a series of residual blocks organized in a specific sequence: 3 blocks with 64 filters, 4 blocks with 128 filters, 6 blocks with 256 filters, and 3 blocks with 512 filters. Unlike ResNet-34, each residual block in ResNet-50 is a bottleneck design, consisting of three convolutional layers instead of two: a 1x1 layer to reduce dimensions, a 3x3 layer to process features, and another 1x1 layer to restore dimensions. Batch normalization and ReLU activation are applied after each convolution. The use of bottleneck layers makes the model more efficient by reducing the computational load while still capturing complex features. The network ends with a global average pooling layer, a fully connected layer, and a softmax activation for classification. ResNet-50's deeper architecture and use of bottleneck layers allow it to achieve superior performance on complex image recognition tasks compared to its shallower counterparts.

ALEXNET

AlexNet, introduced by Alex Krizhevsky and his colleagues in 2012, is a pioneering convolutional neural network that significantly advanced the field of deep learning, particularly in image recognition tasks. AlexNet consists of eight layers: five convolutional layers followed by three fully connected layers. This architecture was designed to handle high-dimensional data and large-scale image classification, and it demonstrated its effectiveness by winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 with a substantial margin.

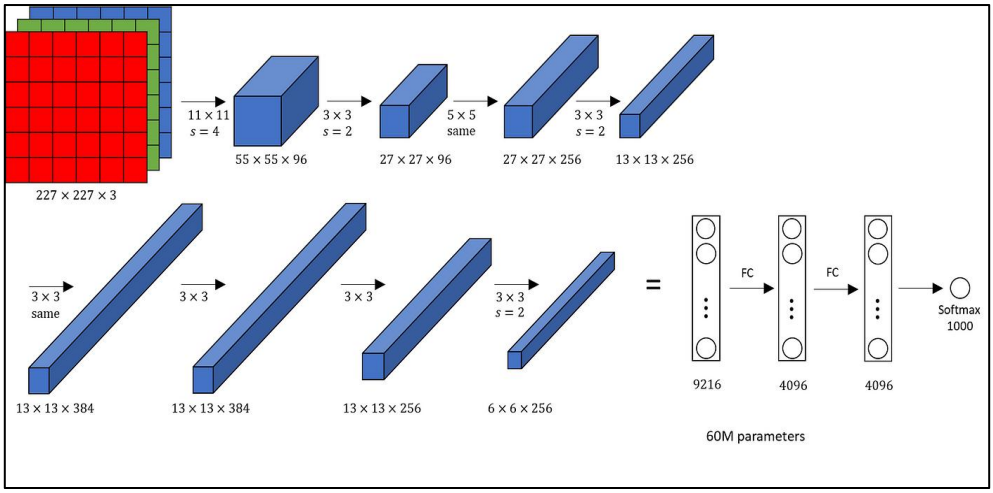


Figure- 3: Architecture of AlexNet

The first layer of AlexNet applies 96 filters of size 11x11 with a stride of 4, followed by max-pooling. The second layer uses 256 filters of size 5x5, also followed by max-pooling. The third, fourth, and fifth convolutional layers are sequential, with 384 filters of size 3x3 in the third layer, 384 filters in the fourth layer, and 256 filters in the fifth layer. The final three layers are fully connected, consisting of 4096 neurons each, with the last layer being a softmax classifier for 1000 classes. AlexNet introduced several key innovations, such as the use of Rectified Linear Units (ReLU) for non-linearity, overlapping max-pooling, and dropout to prevent over-fitting. It also utilized data augmentation techniques and trained on network architectures.

VGG16

VGG-16, developed by the Visual Geometry Group (VGG) at the University of Oxford, is a deep convolutional neural network known for its simplicity and uniform architecture. Introduced in 2014 by Karen Simonyan and Andrew Zisserman, VGG-16 consists of 16 layers, including 13 convolutional layers and 3 fully connected layers, along with max-pooling layers interspersed throughout. The network is characterized by its use of small 3x3 convolutional filters, which are applied repeatedly, allowing the network to capture fine-grained features in the images while maintaining computational efficiency.

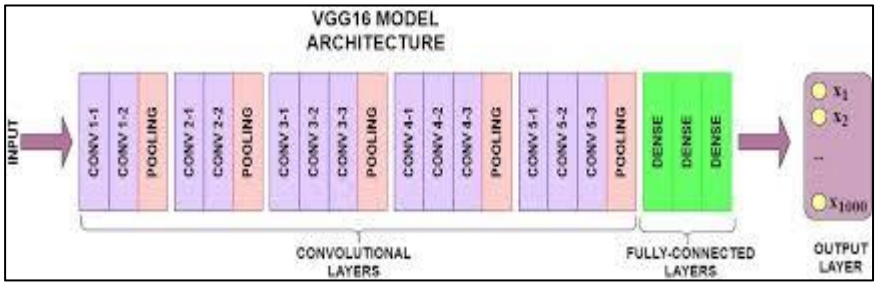


Figure- 4: Architecture of VGG16

The architecture of VGG-16 starts with two convolutional layers with 64 filters each, followed by a max-pooling layer. This pattern is repeated with increasing filter depths: two convolutional layers with 128 filters, two convolutional layers with 256 filters, and then three convolutional layers with 512 filters, each followed by max-pooling. The final set of layers includes three convolutional layers with 512 filters, followed by max-pooling. After the convolutional and pooling layers, the network has three fully connected layers, each with

4096 neurons, and ends with a softmax layer for classification. The use of small 3x3 filters throughout the network allows VGG-16 to capture intricate patterns and details in the images while keeping the model architecture straightforward and manageable. VGG-16's deep yet uniform structure contributed to its success in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014, setting a new benchmark for image classification tasks.

DENSENET 121

DenseNet-121, short for Densely Connected Convolutional Network-121, is a deep learning architecture introduced by Gao Huang and colleagues in 2017. It is designed to improve the flow of information and gradients through the network, which facilitates training and improves performance. DenseNet-121 consists of 121 layers, including convolutional layers, pooling layers, and fully connected layers, but its distinctive feature is the dense connectivity pattern. In this pattern, each layer receives input from all preceding layers, ensuring maximum information flow between layers and reducing the vanishing gradient problem.

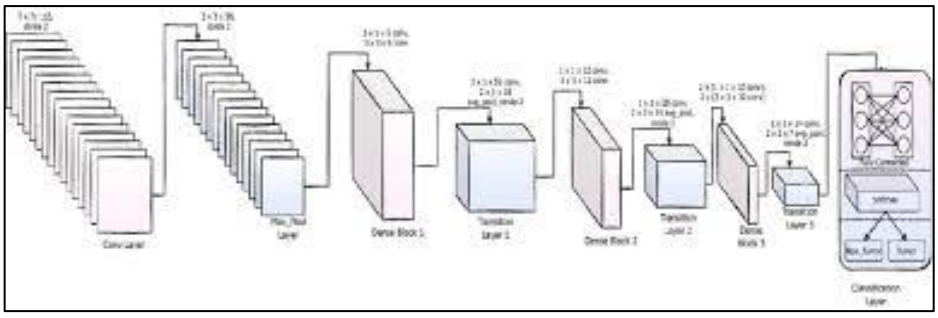


Figure- 5: Architecture of Densenet121

The architecture of DenseNet-121 begins with an initial convolution layer followed by a pooling layer. The network is then structured into four dense blocks, interspersed with transition layers that include convolution and pooling operations to reduce the spatial dimensions of the feature maps. Each dense block consists of multiple convolutional layers, with each layer connected to every other layer in a feed-forward fashion. This means that the input to each layer is the concatenation of the outputs from all previous layers within the same dense block. DenseNet-121 ends with a global average pooling layer followed by a fully connected layer and a softmax activation function for classification. The use of dense connectivity allows the network to use fewer parameters while maintaining high performance, making DenseNet-121 both efficient and powerful for complex image recognition tasks.

DENSENET 169

DenseNet-169, a part of the DenseNet family introduced by Gao Huang et al. in 2017, is designed to enhance the flow of information and gradients across the network layers. DenseNet-169 features a total of 169 layers, comprising convolutional layers, pooling layers, and fully connected layers. The hallmark of DenseNet architectures is their dense connectivity pattern, where each layer receives inputs from all preceding layers and passes its own feature maps to all subsequent layers within the same dense block. This dense connectivity encourages feature reuse and alleviates the vanishing gradient problem, enabling effective training of very deep networks.

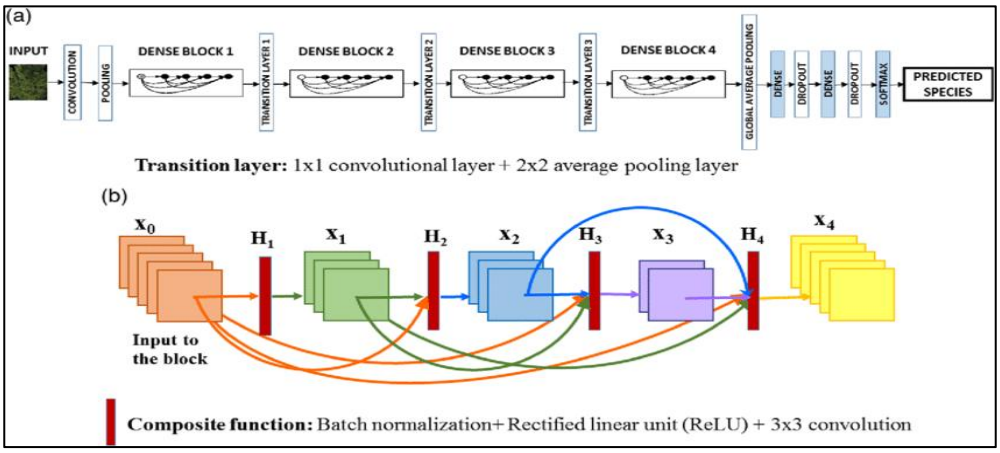


Figure- 6: Architecture of Densenet169

The architecture of DenseNet-169 starts with an initial convolution layer followed by a pooling layer. It then consists of four dense blocks, separated by transition layers that include 1x1 convolution and 2x2 average pooling to reduce the dimensionality of the feature maps. Each dense block comprises multiple layers, with each layer taking as input the concatenated feature maps of all previous layers in that block. Specifically, DenseNet-169 includes 6, 12, 32, and 32 layers in its four dense blocks, respectively. This dense connectivity means each layer has direct access to the gradients from the loss function and the original input signal, promoting efficient learning and feature propagation. The network concludes with a global average pooling layer, a fully connected layer, and a softmax activation for classification. DenseNet-169's structure leads to high performance on image recognition tasks while using fewer parameters than traditional deep networks, making it both effective and computationally efficient.

CUSTOM CONVOLUTION NEURAL NETWORK

A Custom Convolutional Neural Network (Custom CNN) is a deep learning model tailored to process visual data for specific tasks like image classification, object detection, and segmentation. Unlike predefined architectures, a Custom CNN offers flexibility in layer configuration, filter sizes, and connectivity patterns, making it highly adaptable.The architecture begins with an initial convolutional layer that captures low-level features such as edges and textures, followed by a pooling layer, usually MaxPooling, to reduce spatial dimensions and control overfitting. The core comprises several convolutional blocks, each containing multiple convolutional layers with varying filters to learn complex features, interspersed with pooling layers for efficiency.

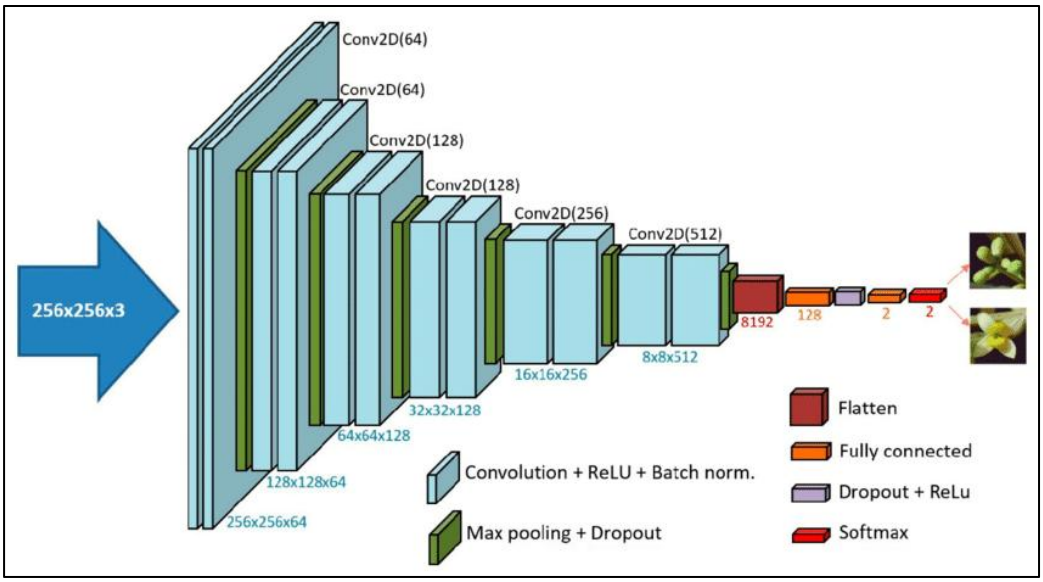


Figure- 7: Architecture of a Neural Network

Towards the end, feature maps are flattened into a vector for fully connected layers that perform high-level reasoning and predictions. Dropout layers may be used to prevent overfitting, and the final layer often employs a softmax activation function for classification. Custom CNNs incorporate advanced techniques like data augmentation, batch normalization, and custom loss functions, ensuring a powerful and efficient network optimized for specific tasks. This tailored approach ensures that the network is not only powerful but also efficient, making the most of available data and computational resources.

3.2. VOICE SIGNAL ANALYSIS

LOGISTIC REGRESSION

Logistic Regression is a statistical model used for binary classification tasks. It estimates the probability that a given input belongs to a particular class. The model applies the logistic function to a linear combination of input features, producing an output between 0 and 1, which is then threshold to determine the class label. Logistic Regression is simple and effective, particularly when the relationship between features and the target is approximately linear.

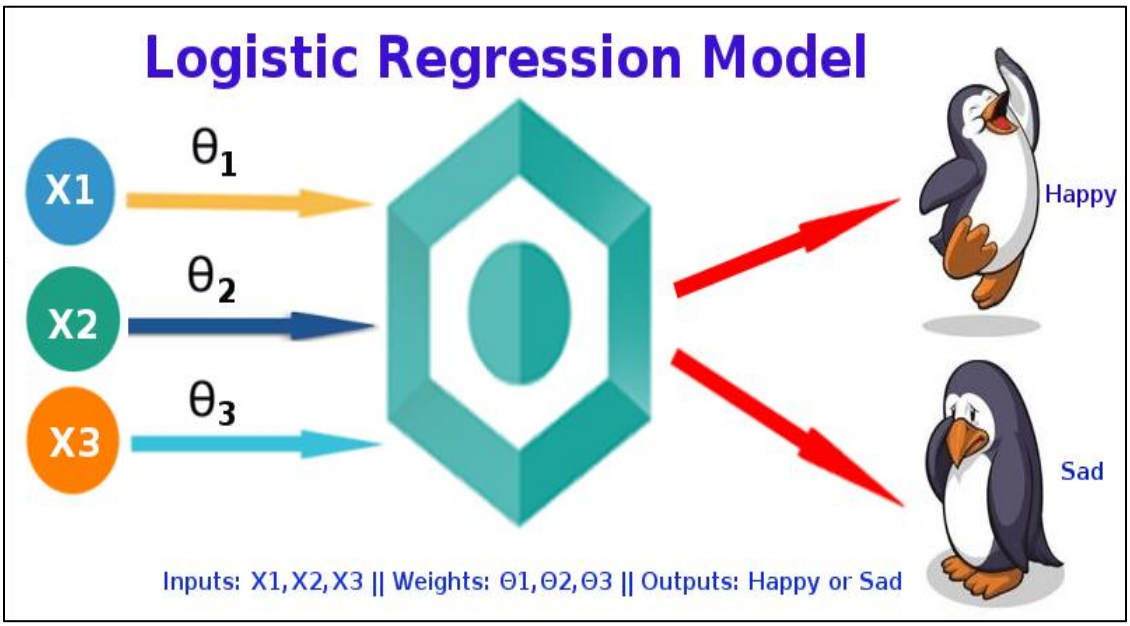


Figure- 8: Working of Logistic Regression

K-NEIGHBORS CLASSIFIER

K-Neighbors Classifier (KNN) is a non-parametric, instance-based learning algorithm used for classification tasks. It classifies a data point based on the majority class of its k nearest neighbors in the feature space. The distance between points is typically measured using Euclidean distance. KNN is simple to implement and effective for small datasets but can be computationally intensive for large datasets due to the need to calculate distances for all points.

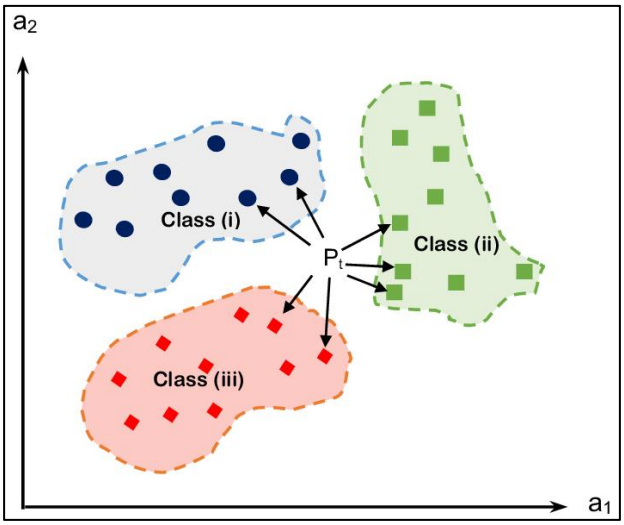


Figure- 9: Working of K-Neighbor's Classification

SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a powerful supervised learning model used for classification and regression tasks. It works by finding the hyperplane that best separates the classes in the feature space, maximizing the margin between the closest points of each class, known as support vectors. SVM is effective in high-dimensional spaces and is robust to overfitting, particularly with the use of kernel functions that allow it to handle non-linear relationships.

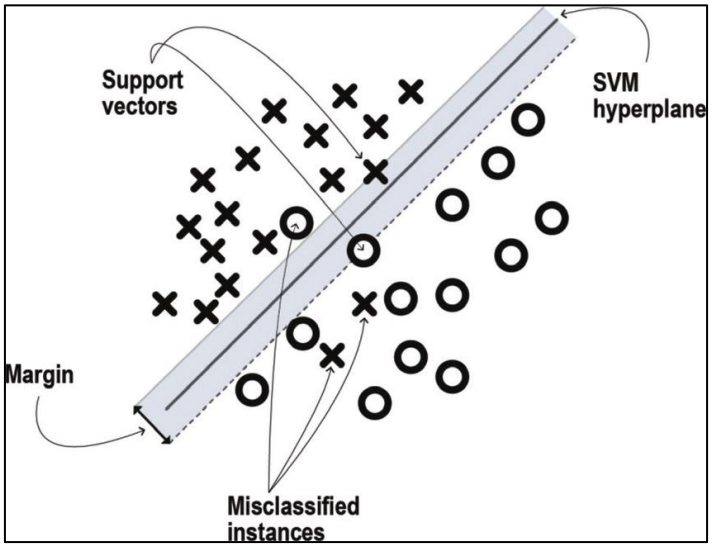


Figure- 10: Working of Support Vector Machine

GAUSSIAN NAIVE BAYES

Gaussian Naive Bayes is a probabilistic classifier based on Bayes' theorem, assuming independence between features. For continuous data, it assumes a Gaussian (normal) distribution. The model calculates the probability of each class given the input features and selects the class with the highest probability. Gaussian Naive Bayes is simple and fast, often used as a baseline in classification tasks, and performs well with large datasets despite its naive assumptions.

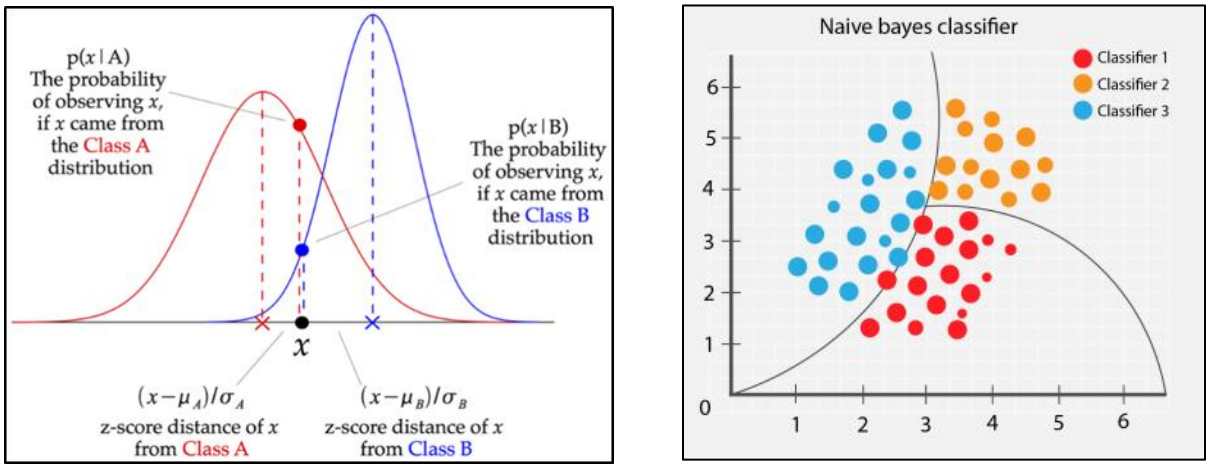


Figure- 11: Working of Gaussian Naïve Bayes

DECISION TREE CLASSIFIER

Decision Tree Classifier is a non-parametric, supervised learning algorithm used for classification tasks. It builds a tree-like model of decisions based on the features of the input data. Each internal node represents a feature, each branch represents a decision rule, and each leaf node represents an output class. Decision Trees are easy to interpret and visualize but can be prone to overfitting, especially with complex datasets.

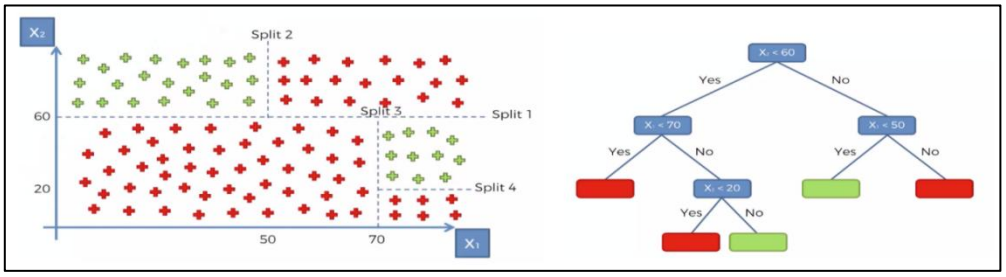


Figure- 12: Working of Decision tree Classifiers

RANDOM FOREST CLASSIFIER

Random Forest Classifier is an ensemble learning method that combines multiple decision trees to improve classification performance. It builds several decision trees during training and outputs the mode of the classes predicted by individual trees. This approach reduces overfitting and increases accuracy by leveraging the diversity of the trees. Random Forests are robust and handle large datasets and high-dimensional data well.

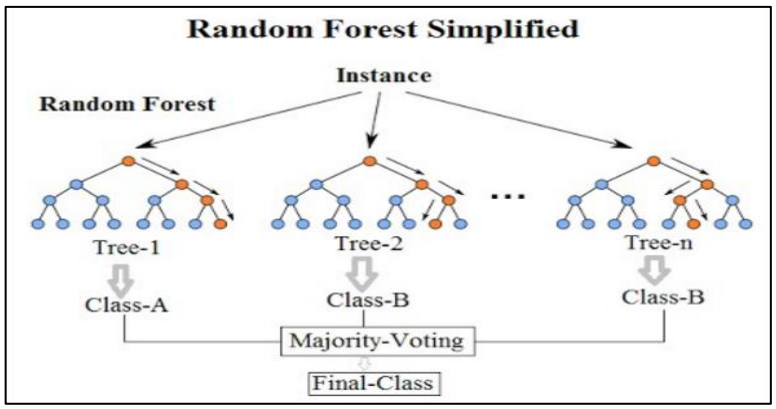


Figure- 13: Working of Random Forest Classifier

GRADIENT BOOSTING CLASSIFIER

Gradient Boosting Classifier is an ensemble technique that builds models sequentially, with each new model correcting errors made by the previous ones. It optimizes the model by minimizing a loss function, typically using gradient descent. Gradient Boosting combines the strengths of weak learners, usually decision trees, to create a strong overall model. It is highly effective for complex datasets but can be computationally expensive and prone to overfitting if not properly tuned.

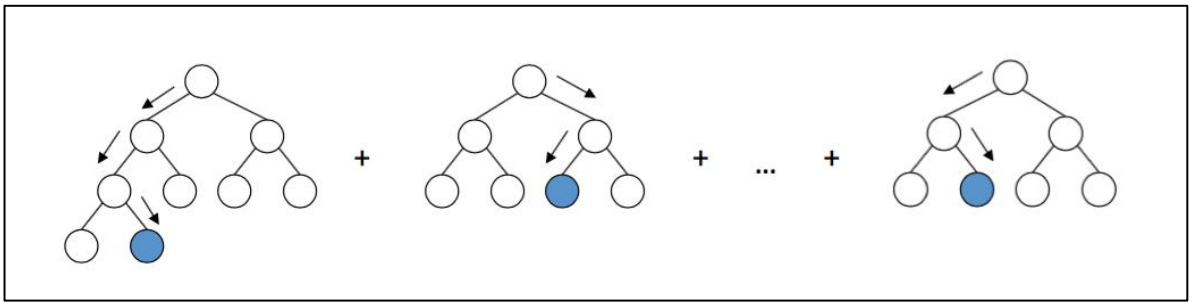


Figure- 14: Working of Gradient Booster Classifier

MULTI-LAYER PERCEPTRON NEURAL NETWORK CLASSIFIER

Multi-layer Perceptron (MLP) Classifier is a type of artificial neural network used for classification tasks. It consists of an input layer, one or more hidden layers, and an output layer. Each layer contains neurons that apply a weighted sum and activation function to the inputs. MLPs can model complex, non-linear relationships in data and are trained using back propagation. They are highly versatile but require careful tuning of hyper parameters and sufficient computational resources.

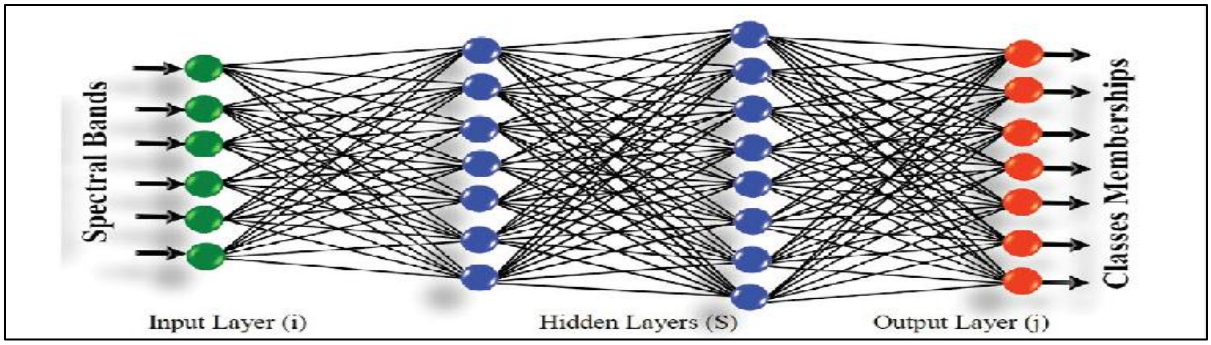


Figure- 15: Architecture of Multi-layer Perceptron

3.3. TYPE KEYSTROKE ANALYSIS

GRADIENT BOOSTING REGRESSOR

Gradient Boosting Regressor is an ensemble learning method used for regression tasks. It builds models sequentially, where each new model corrects the errors of the previous ones. By optimizing a loss function, typically using gradient descent, it combines the strengths of weak learners, usually decision trees, to create a strong overall model. Gradient Boosting Regressor is highly effective for complex datasets but can be computationally intensive and prone to overfitting if not properly tuned.

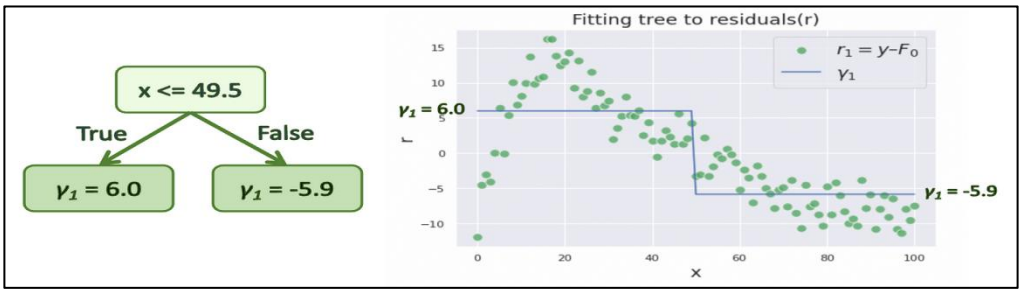


Figure- 16: Working of Gradient Booster Regressor

VOTING CLASSIFIER

Voting Classifier is an ensemble model that combines the predictions from multiple different models to improve classification accuracy. It can be configured in two ways: hard voting, where the majority class is chosen, or soft voting, where the class probabilities are averaged, and the class with the highest probability is selected. The Voting Classifier leverages the strengths of its individual models, often leading to better performance than any single model alone.

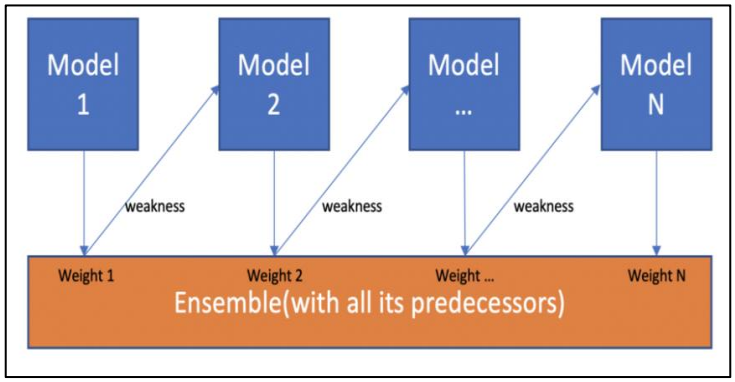


Figure- 17: Working of Voting Classifiers

ADABOOST CLASSIFIER

AdaBoost (Adaptive Boosting) Classifier is an ensemble learning technique that combines multiple weak classifiers to form a strong classifier. It works by fitting a sequence of models, each focusing on the errors made by the previous ones, and adjusting their weights accordingly. The final model is a weighted sum of the individual models' predictions. AdaBoost is particularly effective for binary classification tasks and can improve the performance of weak learners significantly.

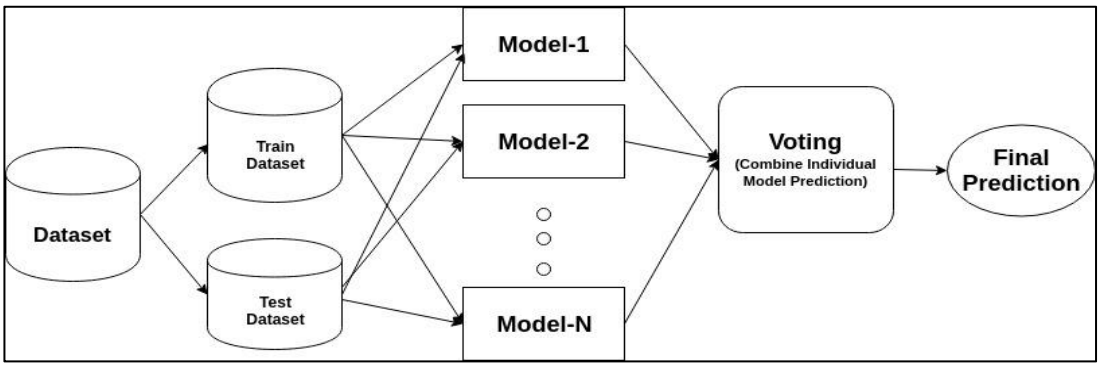


Figure- 18: Working of Adaboost Classifier

BAGGING REGRESSOR

Bagging Regressor (Bootstrap Aggregating) is an ensemble technique used for regression tasks. It trains multiple base regressors on different random subsets of the training data (created by bootstrapping) and then aggregates their predictions, typically by averaging. Bagging reduces variance and helps prevent overfitting, making it a robust method for improving the stability and accuracy of regression models, especially decision trees.

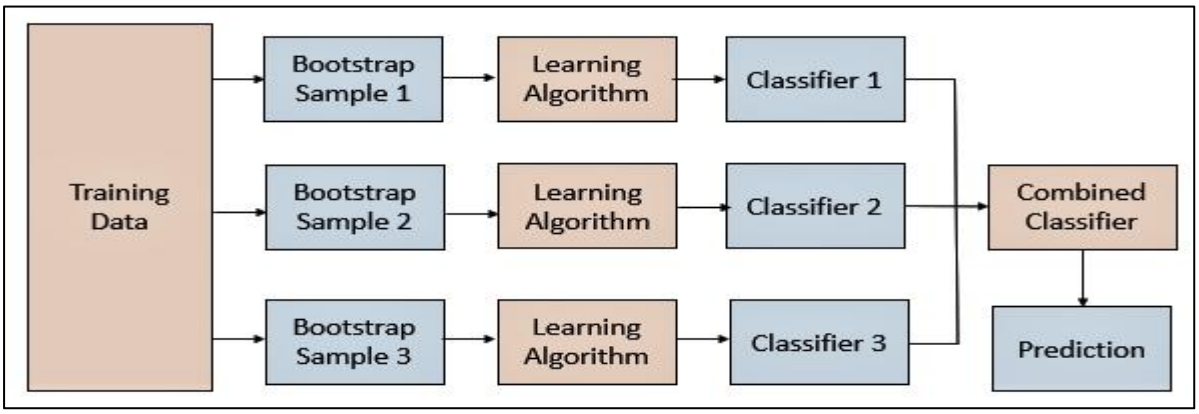


Figure- 19: Model of Bagging Regressor

STOCHASTIC GRADIENT DESCENT CLASSIFIER

Stochastic Gradient Descent (SGD) Classifier is a linear classifier optimized using the stochastic gradient descent algorithm. It is suitable for large-scale and sparse datasets and can handle various loss functions and regularization terms. SGD updates the model parameters incrementally, using one or a few training samples at a time, which makes it highly efficient and scalable. It is particularly useful for online learning scenarios and large datasets where traditional batch learning is impractical.

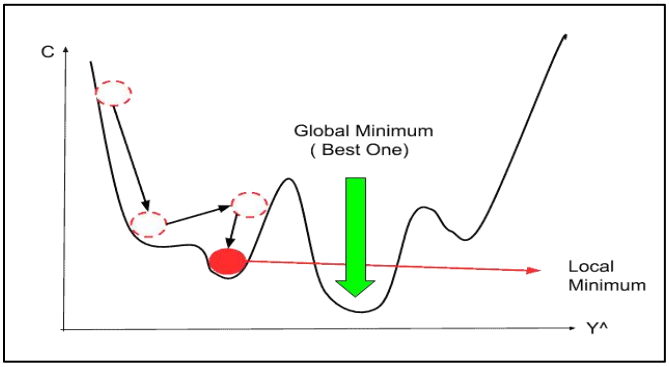


Figure- 20: Working of Stochastic Gradient Descent Classifier

LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant Analysis (LDA) is a classification technique that projects the input data onto a lower-dimensional space to maximize the separation between multiple classes. It assumes that the data for each class follows a Gaussian distribution with the same covariance matrix. LDA finds the linear combinations of features that best separate the classes, making it effective for tasks where the assumption of linear separability holds true.

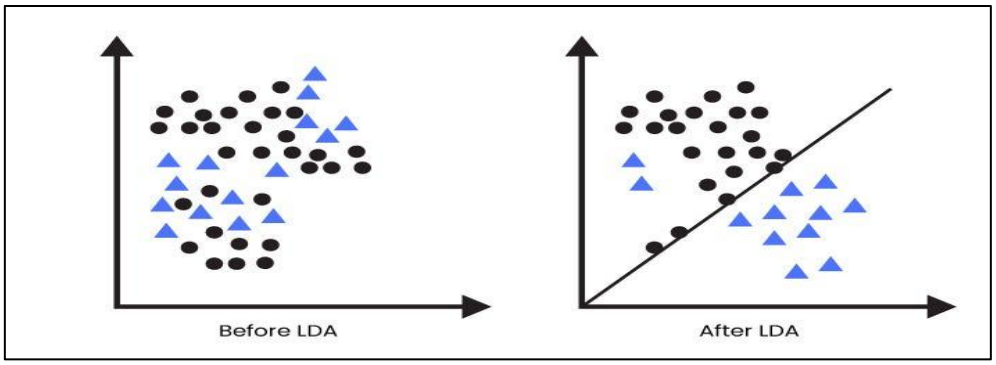


Figure- 20: Working of Linear Discriminant Analysis Classifier

QUADRATIC DISCRIMINANT ANALYSIS

Quadratic Discriminant Analysis (QDA) is similar to LDA but does not assume that the covariance matrices of each class are identical. This allows QDA to capture more complex relationships by modeling quadratic decision boundaries between classes. QDA is more flexible than LDA and can handle non-linear separability, but it requires more data to estimate the parameters accurately and can be more prone to overfitting with small datasets.

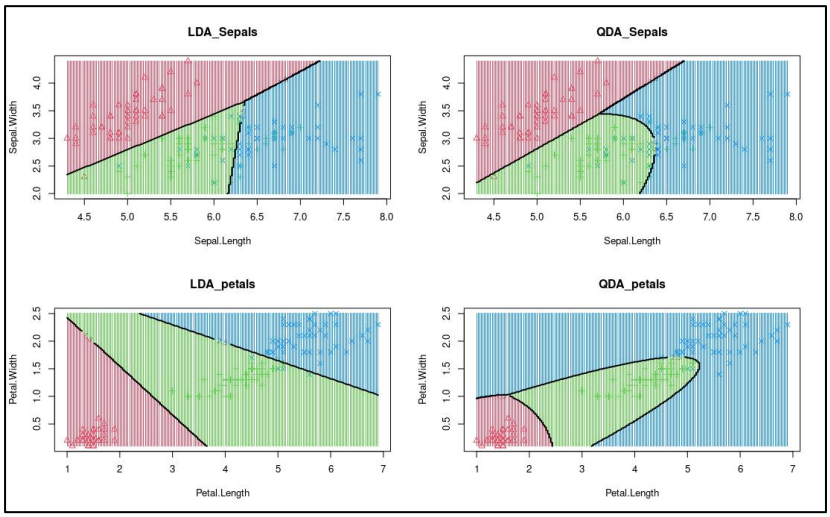


Figure- 22: Working of Quadratic Discriminant Analysis

3.4. GAIT ANALYSIS

LOGISTIC REGRESSION MODEL

Logistic Regression Model is a simple yet effective statistical model used primarily for binary classification tasks. It predicts the probability that a given input belongs to a particular class by applying the logistic (sigmoid) function to a linear combination of the input features. This function outputs a value between 0 and 1, which can be threshold to assign a class label. Logistic Regression assumes a linear relationship between the input features and the log-odds of the target class, making it interpretable and straightforward to implement. Despite its simplicity, it performs well on linearly separable datasets and serves as a solid baseline for more complex models.

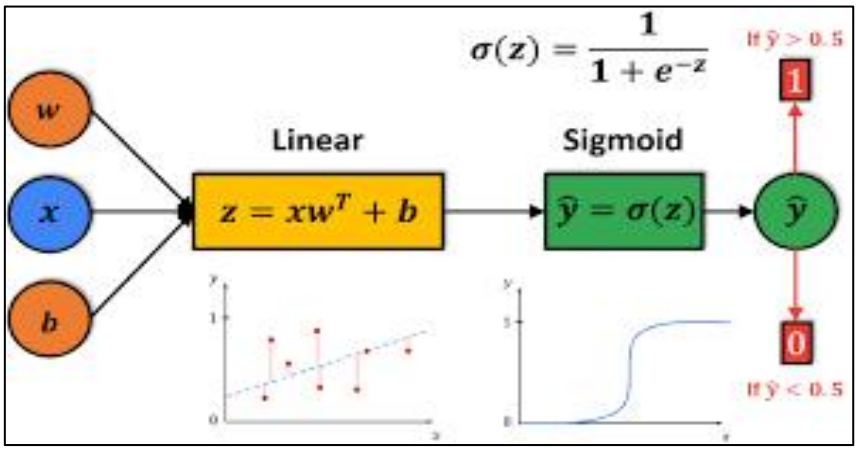


Figure- 23: Working of Logistical Regression Model

BASELINE CAT BOOST MODEL

Baseline Cat Boost Model is an ensemble learning algorithm that uses gradient boosting on decision trees. Cat Boost (Categorical Boosting) is specifically designed to handle categorical features effectively without extensive preprocessing. It builds an ensemble of trees sequentially, where each new tree focuses on reducing the errors of the previous ones. Cat Boost uses ordered boosting to combat overfitting and leverages efficient categorical feature encoding to enhance performance. As a baseline model, Cat Boost is highly efficient, often providing superior performance with minimal hyper parameter tuning, making it a strong starting point for machine learning tasks involving structured data.

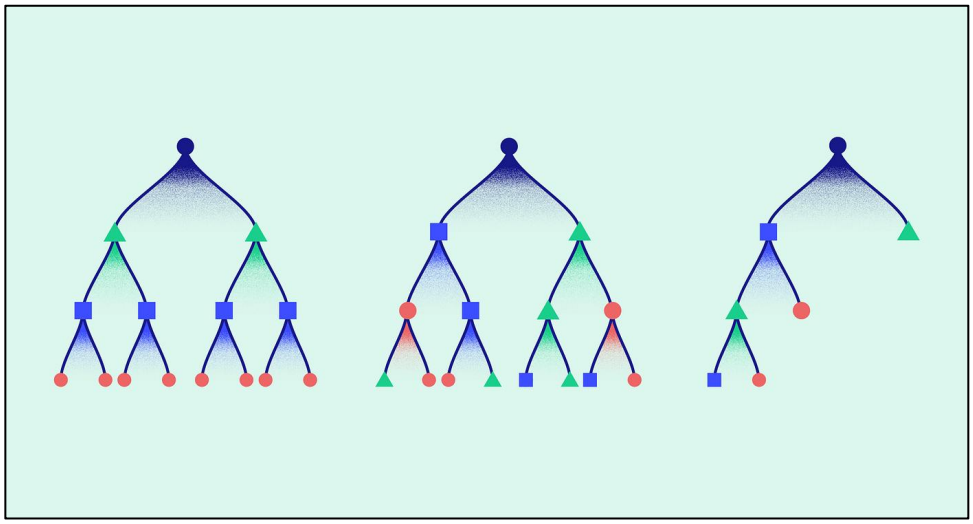


Figure- 24: Working of Baseline Cat Boost Model

LGBM MODEL

LGBM (LightGBM) Model is an advanced gradient boosting framework designed for high efficiency and speed. LightGBM (Light Gradient Boosting Machine) builds decision trees sequentially, like other gradient boosting models, but it introduces novel techniques such as Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce computation and improve accuracy. LGBM is particularly well-suited for large datasets and high-dimensional data, offering fast training times and low memory usage. As a model, LGBM is highly customizable and can achieve state-of-the-art performance with appropriate tuning, making it a popular choice for competitive machine learning tasks.

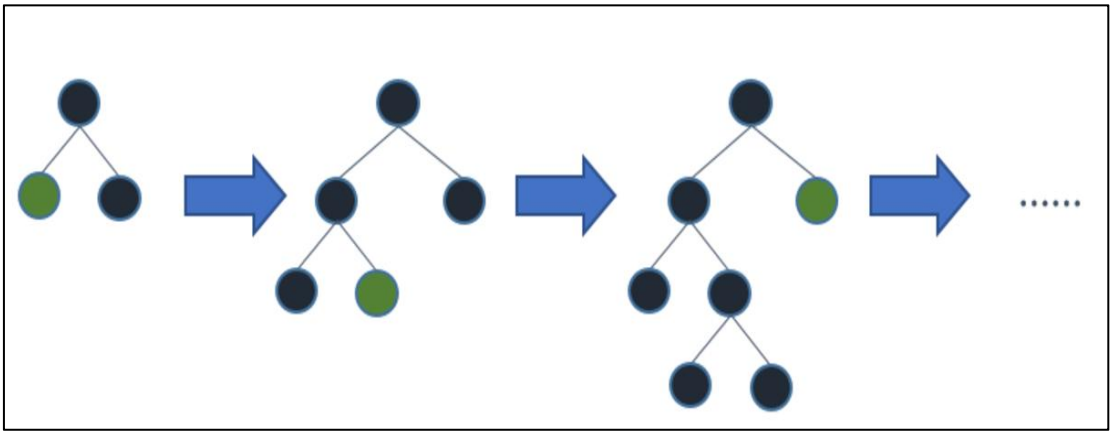


Figure- 25: Working of LGBM Model

CHAPTER 4

METHODOLOGY

4.1 PROCEDURE OUTLINE

Parkinson's disease (PD) is a neurodegenerative disorder that affects movement control, leading to tremors, stiffness, and impaired balance. Traditional diagnostic methods rely on clinical assessments and imaging techniques, but emerging technologies offer innovative solutions for more effective monitoring and diagnosis. The primary focus is on developing a comprehensive system for PD detection, utilizing deep learning techniques.

For motor parameter analysis, the proposed model employs deep learning techniques such as ResNet-34, ResNet-50, DenseNet-121, DenseNet-169, VGG-16, AlexNet, and custom convolutional neural networks (CNNs) on wave and spiral drawing datasets. By training these models on characteristic features extracted from each drawing, we can distinguish between affected and unaffected patients. This integration of advanced machine learning algorithms enables comprehensive monitoring and analysis of PD symptoms.

For acoustic signal analysis of voice input, we use machine learning classifier models such as logistic regression, K neighbors classifier, support vector machine, Gaussian naive Bayes, decision tree classifier, random forest classifier, gradient boosting classifier, and multi-layer perceptron classifier. This involves extracting speech and tremor characteristic features from the given voice signal and mapping them against affected and unaffected patients. This non-motor parameter analysis provides a separate perspective on the comprehensive monitoring and analysis of PD symptoms.

For typing keystroke analysis, we use gradient boosting regressor, voting classifier, AdaBoost classifier, bagging regressor, stochastic gradient descent classifier, linear discriminant analysis, and quadratic discriminant analysis. These models enhance prediction accuracy by combining weak learners and aggregating multiple models' predictions for robustness, improving performance by focusing on misclassified instances and reducing variance through averaging. These models forecast disease progression based on detected activities and trends in physiological parameters.

For gait analysis, we employ logistic regression, the baseline CatBoost model, and the LGBM model. Logistic regression is effective for binary classification, while the baseline CatBoost model handles categorical features efficiently and reduces overfitting. The LGBM model is highly efficient for large datasets, offering fast training and improved accuracy. By continuously refining the machine learning models with new data, we can adapt to individual variations in PD symptoms and provide personalized insights for patients.

Moreover, the ability to send alerts to both patients and caregivers through patient interfacing platforms enhances proactive management of PD symptoms. By promoting early intervention and timely adjustments to treatment plans, these models aim to improve patient outcomes and quality of life. However, challenges such as data privacy, interoperability, and model interpretability must be addressed to ensure the ethical and effective implementation of this technology. The proposed model, combining deep learning algorithms, IoT platforms, and database structures, offers a promising avenue for advancing the diagnosis, monitoring, and detection of Parkinson's disease.

By harnessing the power of technology and data-driven insights, we can revolutionize how we understand and manage this complex neurological condition. The methodology encompasses various stages, starting with data collection from relevant physiological and behavioral signals. This diverse dataset will be pre-processed to remove noise and artifacts, ensuring high-quality input for subsequent analysis.

The models will be adapted and fine-tuned to recognize patterns indicative of Parkinson's disease, leveraging their ability to learn hierarchical representations from complex data. Evaluation of the models' performance will be conducted using validation datasets and standard metrics such as accuracy, precision, recall, log loss, and F1-score. This step ensures that the models meet the required standards for reliable Parkinson's disease detection.

Deployment of the trained models will involve optimizing them for the target platform and ensuring a user-friendly interface for easy integration into clinical settings or wearable devices. Continuous monitoring and maintenance will be essential, with updates and improvements based on user feedback and emerging research in the field. This comprehensive approach combines state-of-the-art deep learning algorithms and web interfacing infrastructure to develop an innovative solution for Parkinson's disease detection.

4.2 DATASET AND EXECUTION

01] Spiral Drawing and Wave Drawing Dataset **- The Michael J. Fox Foundation**

02] Voice Signal Analysis **- University of California, Irvine**

03] Tappy Keystroke **- Charles Sturt University, Australia**

03] Gait Analysis **- Katholieke Universiteit Leuven, Belgium**

STEPS FOR EXECUTING THE PARKINSON'S DISEASE DETECTION SYSTEM

Step 1: Install Required Libraries

Ensure that all necessary libraries for machine learning, deep learning, data handling, and preprocessing are installed. This includes PyTorch, torchvision, librosa, scikit-learn, NumPy, Pandas, OpenCV, and PIL.

Step 2: Import Libraries

Import essential libraries for machine learning, deep learning, audio and image processing, data manipulation, and classification from their respective packages (e.g., PyTorch, librosa, scikit-learn, OpenCV, PIL).

Step 3: Data Collection and Preparation

- Wave and Spiral Drawings: Collect and label datasets with images related to Parkinson's disease indicators.
- Voice Signals: Gather voice recordings and label them accordingly.
- Typing Keystrokes: Collect keystroke patterns and label them.
- Gait Patterns: Obtain gait data and label it with disease indicators.

Step 4: Data Preprocessing

- Wave and Spiral Drawings: Resize images, normalize pixel values, and augment data.
- Voice Signals: Extract features like MFCC's and normalize audio data.
- Typing Keystrokes: Extract features such as typing speed and error rates.
- Gait Patterns: Extract gait features like speed and stride length.
- Split all datasets into training and testing subsets.

Step 5: Input Handling

- Wave and Spiral Drawings: Load images, convert them to tensors, and prepare them for input into the deep learning models.
- Voice Signals: Load audio files, extract features, and prepare feature vectors for input into machine learning models.
- Typing Keystrokes: Load keystroke data, preprocess features, and format them for model input.
- Gait Patterns: Load gait data, extract relevant features, and prepare them for input into machine learning models.

Step 6: Model Initialization

- Wave and Spiral Drawings: Initializing deep learning models such as ResNet-34, ResNet-50, DenseNet-121, DenseNet-169, VGG-16, AlexNet, and custom CNNs.
- Voice Signals: Setting up classifiers like Logistic Regression, K Neighbors, SVM, Gaussian Naive Bayes, Decision Tree, Random Forest, Gradient Boosting, and Multi-layer Perceptron.

- Typing Keystrokes: Choosing models like Gradient Boosting Regressor, Voting Classifier, AdaBoost Classifier, Bagging Regressor, Stochastic Gradient Descent Classifier, Linear Discriminant Analysis, and Quadratic Discriminant Analysis.
- Gait Patterns: Selecting models including Logistic Regression, Baseline CatBoost, and LGB Model.

Step 7: Training the Models

- Wave and Spiral Drawings: Define loss functions and optimizer, then train the deep learning models with the pre-processed data, applying data augmentation techniques.
- Voice Signals: Define loss functions and optimizer, and train the classifiers with audio features, adjusting hyper parameters as needed.
- Typing Keystrokes: Train the models on keystroke features, using appropriate loss functions and optimizer.
- Gait Patterns: Train the models on gait data, defining loss functions and optimizer, and adjusting hyper parameters for optimal performance.

Step 8: Evaluating the Models

- All Inputs: Evaluate each model on its respective testing dataset using metrics such as accuracy, precision, recall, log loss, confusion matrix and F1-score. Compare the performance of different models to select the best performing architectures for each input type.

Step 9: Cloud Integration and Real-Time Monitoring

- All Inputs: Integrate the trained models with a cloud-based platform to enable remote access and real-time monitoring. Set up cloud storage and server management systems for seamless integration with healthcare systems or wearable devices.

Step 10: Deployment and User Interface

- All Inputs: Optimize the models for deployment, ensuring they are efficient and user-friendly. Develop an intuitive user interface to facilitate input lading, easy integration into clinical settings or wearable devices for end-users.

Step 11: Continuous Monitoring and Maintenance

- All Inputs: Continuously monitor the performance of deployed models and gather feedback from users and healthcare providers. Update and refine the models based on new data, emerging research, and user feedback to maintain high accuracy and reliability.

This consolidated approach ensures that each type of input is handled systematically, from data collection to model deployment and maintenance, providing a comprehensive system for Parkinson's disease detection and management.

4.3 PROPOSED MODEL

The proposed model represents an advanced system designed to enhance the detection, monitoring, and detection of Parkinson's disease (PD) through an integrated approach that leverages cutting-edge machine learning and deep learning technologies. This system aims to offer a comprehensive framework for early diagnosis and effective management by analyzing a diverse array of physiological and behavioral data.

Central to the model is its ability to interpret motor symptoms through sophisticated deep learning architectures. Using advanced models such as ResNet-34, ResNet-50, DenseNet, VGG-16, and custom convolutional neural networks (CNNs), the system examines wave and spiral drawings to identify subtle motor control impairments associated with Parkinson's disease. These models are adept at recognizing intricate patterns in the drawings, which can signal early signs of motor dysfunction and help track disease progression with high accuracy.

In addition to motor symptoms, the model integrates voice signal analysis to provide insights into non-motor aspects of Parkinson's disease. Machine learning classifiers, including Logistic Regression, K Neighbors, Support Vector Machines (SVM), and Random Forest,

are utilized to analyze acoustic features such as tremor characteristics. This analysis helps differentiate between affected and unaffected individuals by examining changes in speech patterns and voice quality, offering a broader perspective on the impact of Parkinson's disease. The model also addresses fine motor control through the analysis of typing keystrokes. Models such as Gradient Boosting Regressor and AdaBoost Classifier assess typing speed, accuracy, and error patterns. By evaluating these metrics, the system gains insights into how Parkinson's disease affects motor skills and progression, allowing for a more detailed understanding of the disease's impact on daily activities. Furthermore, gait analysis is a critical component of the model, using Logistic Regression and Light Gradient Boosting Machine (LGBM) models to monitor changes in gait speed and stride length. These indicators are essential for assessing motor function and disease advancement. By analyzing gait patterns, the system provides valuable information about disease progression and helps tailor interventions to individual needs.

The methodology of the proposed model involves several key steps to ensure robust performance. Initially, data from various sources is collected and preprocessed to prepare it for analysis. This preprocessing includes resizing images, normalizing audio features, and standardizing keystroke and gait data. Deep learning and machine learning models are then trained on this preprocessed data, with optimization techniques such as data augmentation and hyper-parameter tuning employed to enhance model accuracy. The models are evaluated using metrics like accuracy, precision, recall, and F1-score to identify the most effective models for each input type. To facilitate practical use, the trained models are integrated with a cloud-based platform, enabling remote access and real-time monitoring. This integration allows for seamless connection with healthcare systems and wearable devices, making the system practical for clinical settings. A user-friendly interface is developed to ensure ease of use for end-users, including both clinicians and patients. This iterative approach ensures that the system adapts to new insights and continues to provide valuable support in managing Parkinson's disease. Overall, the proposed model offers a comprehensive, data-driven solution for Parkinson's disease, combining deep learning and machine learning techniques to deliver an advanced tool for early detection, ongoing monitoring, and personalized treatment. This approach aims to significantly improve the management of Parkinson's disease, enhancing the quality of life for affected individuals.

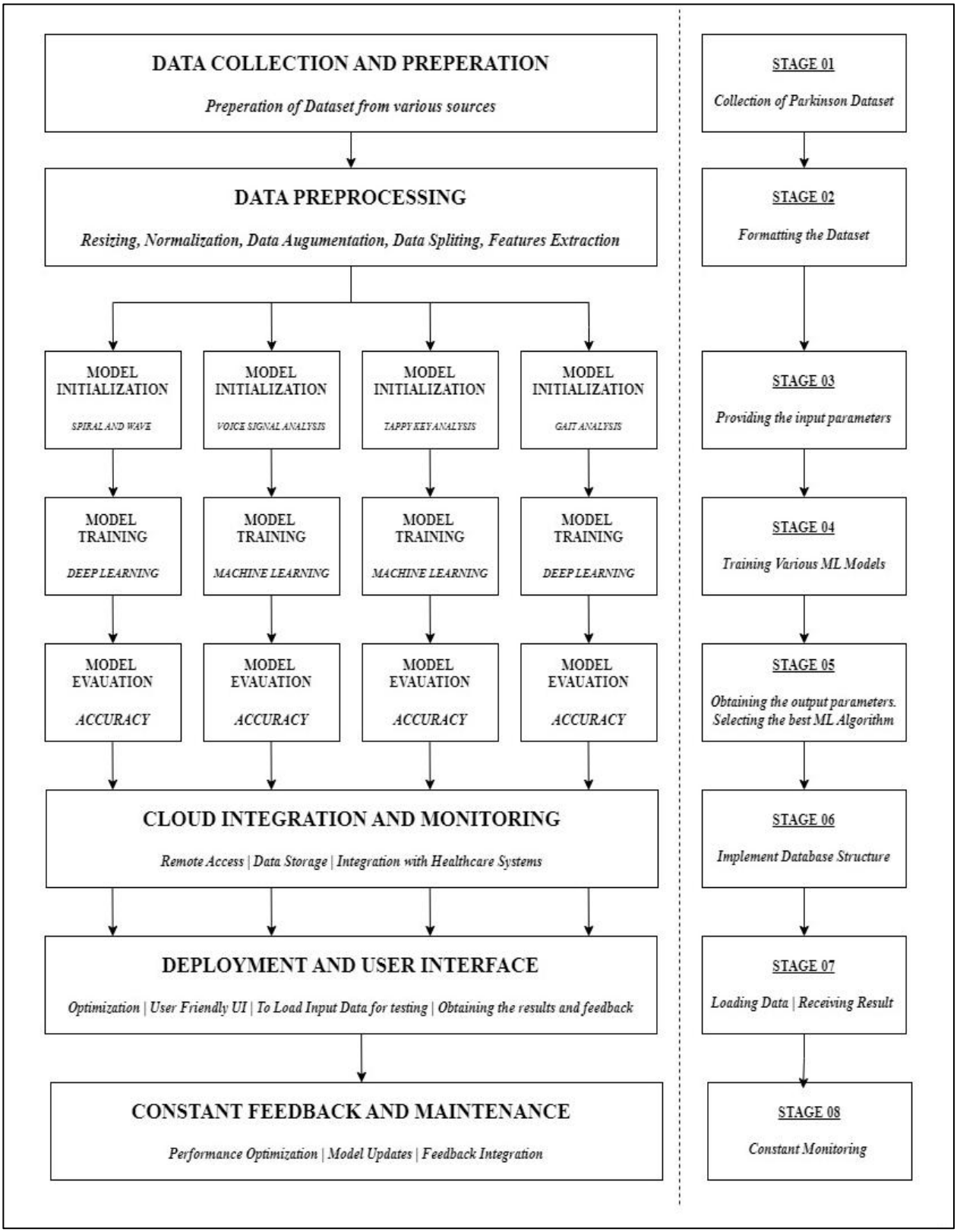


Figure- 26: Working Methodology

CHAPTER 5

SYSTEM IMPLEMENTATION

5.1 MODEL REQUIREMENTS

DATASET

SPIRAL AND WAVE DATASET: The speed and pen-pressure while sketching a spiral are lower among Parkinson's disease (PD) patients with higher severity of the disease. However, the correlation between these features and the severity level (SL) of PD has been reported to be 0.4. There is a need for identifying parameters with a stronger correlation for considering this for accurate diagnosis of the disease. This study has proposed the use of the Composite Index of Speed and Pen-pressure (CISP) of sketching as a feature for analyzing the severity of PD. A total of 28 control group (CG) and 27 PD patients (total 55 participants) were recruited and assessed for Unified Parkinson's Disease Rating Scale (UPDRS). They drew guided Archimedean spiral on an A3 sheet. Speed, pen-pressure, and CISP were computed and analyzed to obtain their correlation with severity of the disease. The correlation of speed, pen-pressure, and CISP with the severity of PD was −0.415, −0.584, and −0.641, respectively. Mann–Whitney U test confirmed that CISP was suitable to distinguish between PD and CG, while non-parametric k-sample Kruskal–Wallis test confirmed that it was significantly different for PD SL-1 and PD SL-3. This shows that CISP during spiral sketching may be used to differentiate between CG and PD and between PD SL-1 and PD SL-3 but not SL-2.

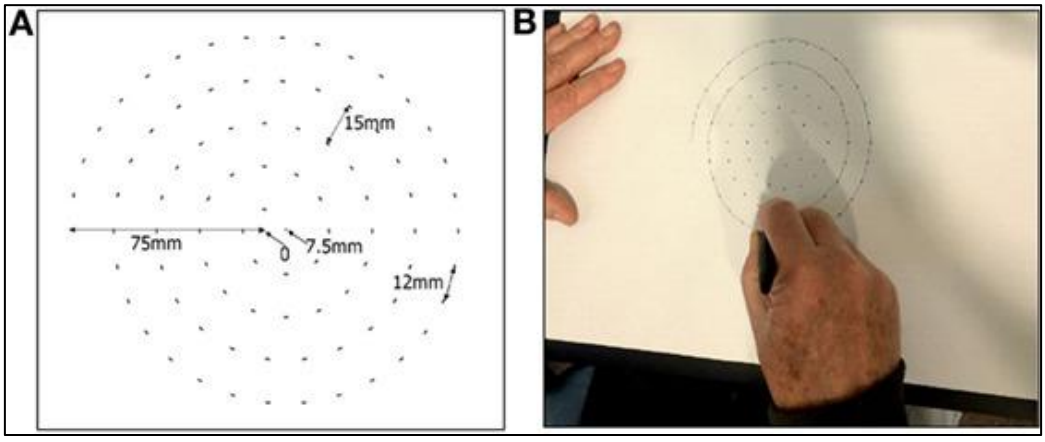


Figure- 27: Spiral Dataset Generation

VOICE SIGNAL ANALYSIS:

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient, the name of the patient is identified in the first column..

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966

MDVP:Shimmer	...	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
0.04374	...	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.813031	0.266482	2.301442	0.284654
0.06134	...	0.09403	0.01929	19.085	1	0.458359	0.819521	-4.075192	0.335590	2.486855	0.368674
0.05233	...	0.08270	0.01309	20.651	1	0.429895	0.825288	-4.443179	0.311173	2.342259	0.332634
0.05492	...	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.117501	0.334147	2.405554	0.368975
0.06425	...	0.10470	0.01767	19.649	1	0.417356	0.823484	-3.747787	0.234513	2.332180	0.410335

Figure- 28: Voice Dataset Example

TAPPY KEYSTROKE ANALYSIS:

This is the keystroke dataset for the study titled 'High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing'. This research report is currently under review for publication by PLOS ONE. The dataset contains keystroke logs collected from over 200 subjects, with and without Parkinson's Disease (PD), as they typed normally on their own computer (without any supervision) over a period of weeks or months (having initially installed a custom keystroke recording app, Tappy). This dataset has been collected and analyzed in order to indicate that the routine interaction with computer keyboards can be used to detect changes in the characteristics of finger movement in the early stages of PD

	HoldTime	Direction	LatencyTime	ID	Hand	FlightTime	\
0	121.1	RR	277.3	927ZLCPEJM	R	246.1	
1	121.1	RR	375.0	927ZLCPEJM	R	253.9	
2	113.3	LR	480.5	927ZLCPEJM	R	375.0	
3	78.1	RL	273.4	927ZLCPEJM	L	160.2	
4	121.1	LR	144.5	927ZLCPEJM	R	39.1	

	binIndex
0	(9608760, 9608850]
1	(9608760, 9608850]
2	(9608760, 9608850]
3	(9608760, 9608850]
4	(9608760, 9608850]

	ID	BirthYear	Gender	Parkinsons	Tremors	DiagnosisYear	Sided	\
0	QEYNMBJ8T0	1940.0	Male	True	False	2009	None	
1	QDV7XQWLI7	1938.0	Male	True	False	2016	None	
2	7QVQD5IJEH	1939.0	Male	True	False	2010	Right	
3	6LB9FQABZQ	1956.0	Female	False	False	------	None	
4	LEFLOMUTBN	1943.0	Male	True	False	2016	None	
...								
1	Don't know	NaN	False	False	False	True		
2	Don't know	Medium	False	False	False	True		
3	Don't know	------	False	False	False	False		
4	Don't know	Mild	True	False	False	False		

Figure- 29: Tappy Keystroke Dataset Example

GAIT ANALYSIS

This database contains measures of gait from 93 patients with idiopathic PD (mean age: 66.3 years; 63% men), and 73 healthy controls (mean age: 66.3 years; 55% men). The database includes the vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground. Underneath each foot were 8 sensors (Ultraflex Computer Dyno Graphy, Infotronic Inc.) that measure force (in Newtons) as a function of time. The output of each of these 16 sensors has been digitized and recorded at 100 samples per second, and the records also include two signals that reflect the sum of the 8 sensor outputs for each foot. For details about the format of the data, please see this note.

With this information, one can investigate the force record as a function of time and location, derive measures that reflect the center-of-pressure as a function of time, and determine timing measures (e.g., stride time, swing time) for each foot as functions of time.

Thus, one can study the stride-to-stride dynamics and the variability of these time series. This database also includes demographic information, measures of disease severity and other related measures.

	Time	AccV	AccML	AccAP	StartHesitation	Turn	Walking
0	0	-9.533939	0.566322	-1.413525	0	0	0
1	1	-9.536140	0.564137	-1.440621	0	0	0
2	2	-9.529345	0.561765	-1.429332	0	0	0
3	3	-9.531239	0.564227	-1.415490	0	0	0
4	4	-9.540825	0.561854	-1.429471	0	0	0

- AccV: Vertical
- AccML: Mediolateral
- AccAP: Anteroposterior

Figure- 30: Gait Dataset Parameters

5.2 INPUT PARAMETERS

Spiral and Wave Analysis

Parkinson Affected Dataset

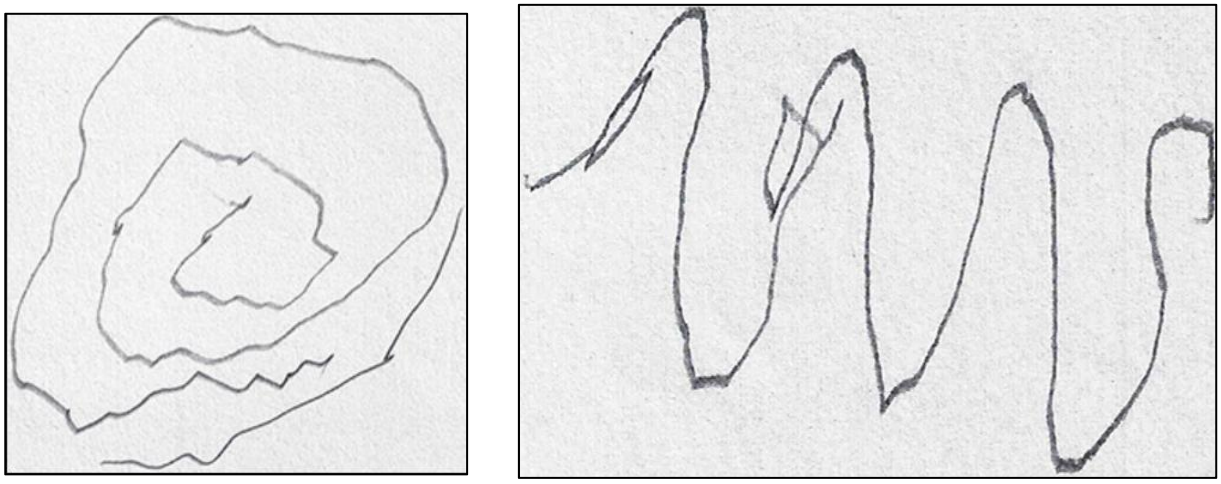


Figure- 31: Spiral and Wave Parkinson Affected Dataset Examples

Healthy Dataset

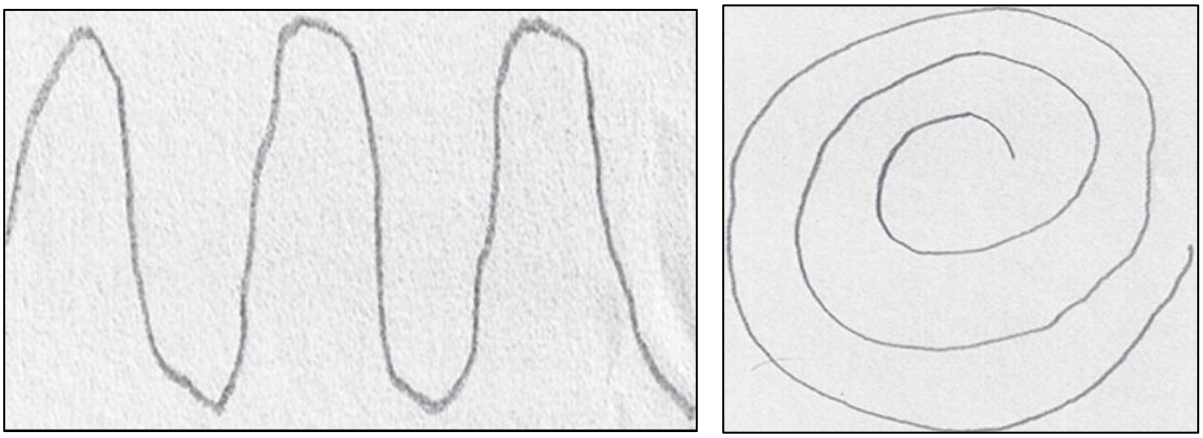


Figure- 32: Spiral and Wave Healthy Dataset Examples

Collect a dataset of wave and spiral drawings, labeling them with indicators of Parkinson's disease presence or not. Organize and label the dataset accordingly.

Voice Signal

name	0
MDVP:Fo(Hz)	0
MDVP:Fhi(Hz)	0
MDVP:Flo(Hz)	0
MDVP:Jitter(%)	0
MDVP:Jitter(Abs)	0
MDVP:RAP	0
MDVP:PPQ	0
Jitter:DDP	0
MDVP:Shimmer	0
MDVP:Shimmer(dB)	0
Shimmer:APQ3	0

Shimmer:APQ5	0
MDVP:APQ	0
Shimmer:DDA	0
NHR	0
HNR	0
status	0
RPDE	0
DFA	0
spread1	0
spread2	0
D2	0
PPE	0

Figure- 33: Voice Signal Dataset Parameters

Collect a dataset of voice recordings, labeling them with Parkinson's disease indicators. Organize and label the dataset.Preprocess audio data by extracting features such as MFCCs, pitch, and tremor characteristics. Split the dataset into training and testing subsets.

Tappy Keystroke Analysis

Preprocess keystroke data by extracting relevant features such as Hand, Direction, Hold Time, Latency, Flight and Press Time. Split the dataset into training and testing subsets.

```
MIT taps initial dataframe:

      ID Hand Direction HoldTime LatencyTime FlightTime pressTime
6   1029    L        RL   0.1163      0.4649     0.3719    7.2105
9   1029    L        LL    0.171      0.3793      0.195    7.8201
10  1029    S        LS   0.1089      0.2563     0.0852    8.0764
11  1029    R        SR   0.1138      0.4716     0.3627     8.548
12  1029    L        RL   0.1974      0.2149     0.1011    8.7629
```

Figure- 34: Typing Dataset Parameters and Examples

Gait Analysis

Collect a dataset of gait patterns with Parkinson's disease indicators such as Time, Start Hesitation, Vertical, Mediolateral, Anteroposterior Acceleration, Turning and more. Organize and label the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 162907 entries, 0 to 162906
Data columns (total 9 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   Time             162907 non-null  int64
 1   AccV             162907 non-null  float64
 2   AccML            162907 non-null  float64
 3   AccAP            162907 non-null  float64
 4   StartHesitation  162907 non-null  int64
 5   Turn             162907 non-null  int64
 6   Walking          162907 non-null  int64
 7   Valid            162907 non-null  bool
 8   Task             162907 non-null  bool
dtypes: bool(2), float64(3), int64(4)
memory usage: 9.0 MB
```

Figure- 35: Gait Analysis Dataset Parameters

5.3 LISTED OUTCOMES

- Accuracy Score: Measures the ratio of correctly predicted instances to the total instances in the dataset. It provides a simple measure of how often the classifier is correct across all classes. However, it can be misleading in cases of imbalanced datasets where some classes are much more frequent than others.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

- F1 Score: Combines precision and recall into a single metric, balancing the trade-off between the two. It balances precision and recall, making it useful for situations where both false positives and false negatives are important. It is especially helpful when dealing with imbalanced datasets.

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

- Log Loss: Measures the performance of a classification model by quantifying the difference between the predicted probability and the actual label. Lower log loss values indicate better model performance, with 0 being perfect.

$$\text{Log Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

- Precision Score: Indicates the proportion of true positive predictions among all positive predictions made.It measures how many of the predicted positives are actually true positives. This is crucial in scenarios where false positives are costly or undesirable.

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

- Confusion Matrix: A table used to evaluate the performance of a classification algorithm, showing the number of true positives, true negatives, false positives, and false negatives. Helps in understanding the types of errors made by the classifier. It provides a detailed breakdown of the classification performance, showing the types and counts of errors. It's essential for understanding how the model performs across different classes.

	Predicted Positive	Predicted Negative
Actual Positive	True Positives (TP)	False Negatives (FN)
Actual Negative	False Positives (FP)	True Negatives (TN)

- Recall Score: Measures the proportion of actual positives that were correctly identified by the model.It measures how many of the actual positives were correctly identified by the model. This is crucial in scenarios where missing a positive instance (false negatives) is costly or undesirable.

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

5.4 SELECTION CRITERIA

When selecting the best machine learning model, it is essential to align the choice of evaluation metrics with the specific needs of the project and the characteristics of the data.

For datasets with balanced class distributions, where each class is represented equally, the Accuracy Score is a straightforward and effective metric. It provides a clear measure of the model's overall correctness in its predictions. However, to ensure a more comprehensive evaluation, it is advisable to also consider the F1 Score. This metric helps to confirm that the model maintains a good balance between precision and recall, addressing potential issues where a high accuracy might mask deficiencies in handling specific classes.

In scenarios with imbalanced class distributions, where some classes occur much more frequently than others, the F1 Score becomes crucial. This metric provides a balanced measure of precision and recall, offering a clearer picture of how well the model performs, especially for the minority class. It is also important to evaluate Precision and Recall individually to understand the model's performance on the minority class. For models that need to provide probabilistic predictions, Log Loss is essential. It measures how well the predicted probabilities align with the actual outcomes, giving valuable insights into the quality of the probability estimates.

When probability estimation is a key requirement, Log Loss should be the primary focus. This metric assesses the accuracy of the probabilistic predictions, which is crucial when the model's confidence levels are important. Alongside Log Loss, evaluating Accuracy or F1 Score is useful to ensure that the overall classification performance meets the required standards.

To gain a thorough understanding of model errors, the Confusion Matrix is an invaluable tool. It provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, which helps in diagnosing specific types of errors made by the model. Using the Confusion Matrix in conjunction with Precision, Recall, and F1 Score allows for a comprehensive evaluation of model performance and highlights areas where improvements might be needed.

In summary, selecting the best model involves choosing metrics that are aligned with the project's objectives and the nature of the data. For balanced datasets, accuracy may be sufficient, while for imbalanced data or when the cost of errors varies, emphasis should be placed on precision, recall, and F1 Score. Log Loss is critical for evaluating probabilistic models, and the Confusion Matrix provides detailed insights into classification errors.

5.5 INTEGRATION OF TRAINED MODELS

Integrating machine learning models from datasets such as voice, spiral, wave, keystroke, and gait involves a structured approach to maximize the utility of each dataset. Here's a concise summary for your report:

Preprocessing and Feature Extraction are critical initial steps. For the voice dataset, preprocess the audio data, extract features like Mel-frequency cepstral coefficients (MFCCs) and pitch, and use models such as RNNs or SVM. In the spiral dataset, normalize the drawings, extract features such as stroke length and curvature, and apply CNNs. The wave dataset requires preprocessing and normalization, focusing on features like amplitude and frequency, with CNNs or RNNs being suitable. For the keystroke dataset, normalize timing and pressure data, extract features like typing speed, and use LSTMs or gradient boosting machines. The gait dataset involves normalizing data and extracting features such as stride length, with RNNs, CNNs, or SVMs being appropriate.

Feature Fusion entails combining features from each model into a unified vector. Collect relevant features from each model, concatenate them, and normalize the combined set to ensure consistency across different data types. Ensemble Learning techniques aggregate predictions from the models. Use methods like voting or averaging for simpler integration, or stacking and blending for more sophisticated approaches. In stacking, train a meta-model on the base models' predictions to optimize performance, while blending offers straightforward aggregation.

Integration and Deployment require developing a unified pipeline for data preprocessing, feature extraction, model inference, and result aggregation. Ensure smooth data flow and conduct thorough testing to validate the integrated model's performance. After validation, deploy the model and monitor its performance for continuous improvement.

User Interaction via Web Interface involves creating a web platform where users can input their data—voice recordings, spiral drawings, keystroke patterns, and gait metrics. This interface should facilitate data upload and provide real-time predictions on Parkinson's disease likelihood, with clear instructions and secure data handling.

Cloud Integration is essential for scalable data processing and storage. Deploy the machine learning models and the web interface on cloud platforms to handle large volumes of data and ensure high availability and reliability. Cloud services offer scalability to manage varying workloads and support seamless updates and maintenance, enhancing the overall efficiency and accessibility of the system.

Interpretation and Visualization of results are essential for understanding the model's performance. Analyze how predictions from different datasets contribute to the final outcome and use visualization tools to effectively communicate the model's insights and performance metrics. This approach ensures a comprehensive system that leverages diverse data sources to deliver accurate predictions while providing users with a practical interface for testing.

5.6 SOFTWARE REQUIREMENTS

- Python: Python is the core programming language for developing machine learning models and data processing scripts. It provides a wide range of libraries and frameworks for machine learning, such as TensorFlow, PyTorch, and scikit-learn, making it an essential tool for implementing and training models.
- VSCode: Visual Studio Code (VSCode) is a versatile code editor that supports multiple programming languages and offers a rich set of extensions for Python development. It provides features like debugging, code linting, and integrated terminal access, facilitating an efficient development workflow.
- Jupyter: Jupyter Notebooks are used for interactive development and visualization of data. They allow for the creation of documents that combine live code, equations, visualizations, and narrative text. Jupyter is particularly useful for data exploration, experimentation, and sharing results.

- Flask: Flask is a lightweight web framework for Python that is used to build and deploy the web interface for user interaction. It handles HTTP requests, integrates with the machine learning models, and serves the user interface, enabling users to input their data and receive predictions.

- Blynk IoT Cloud Architecture: Blynk provides a cloud-based platform for Internet of Things (IoT) applications. It offers tools for building IoT dashboards and managing devices. In the context of this project, Blynk can be used for integrating IoT sensors, collecting real-time data, and visualizing it on a web or mobile interface.

- HTML | CSS: HTML (HyperText Markup Language) and CSS (Cascading Style Sheets) are essential for designing and styling the web interface. HTML structures the content on the web pages, while CSS is used for layout, colors, and overall appearance. Together, they ensure a user-friendly and visually appealing interface.

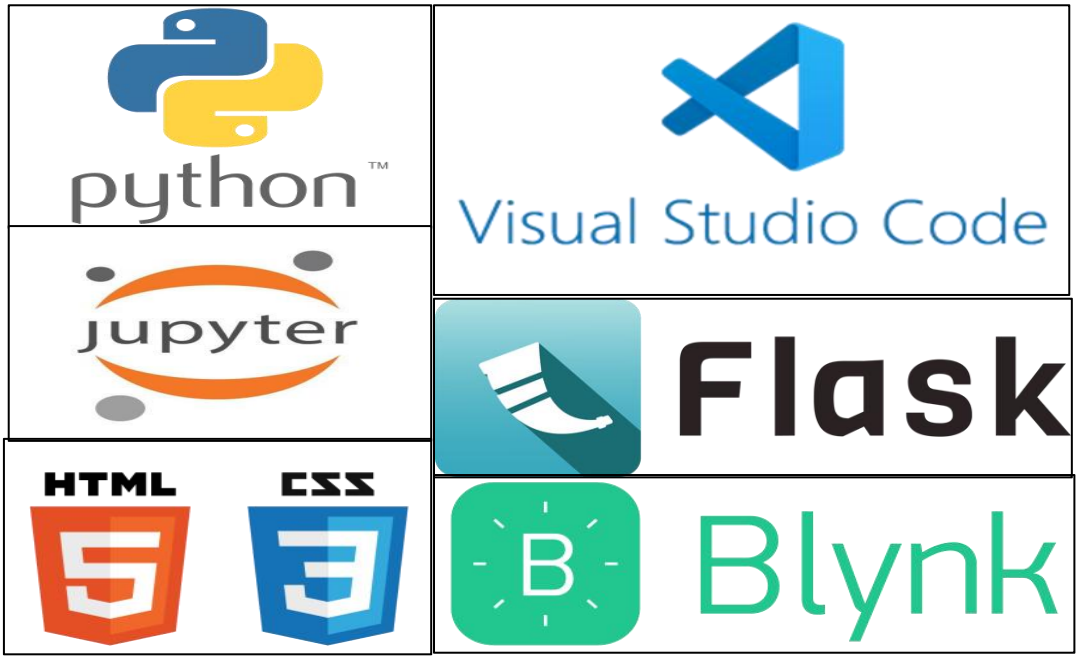


Figure- 36: Software's Used

This software stack supports the development, deployment, and user interaction aspects of the machine learning project, ensuring robust model integration and seamless user experience.

CHAPTER 6

RESULTS AND DISCUSSION

6.1 SPIRAL AND WAVE ANALYSIS OUTCOMES:

TRAIN LOSS, VALID LOSS, ACCURACY AND TIME

- RESENT34

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	1.655003	0.854319	0.500000	00.41
1	1.399725	0.661609	0.800000	00:27
2	1.071367	0.638107	0.750000	00:29
3	0.969734	0.942346	0.700000	00:25
4	0.847144	1.020449	0.733333	00:30
5	0.861470	0.918422	0.800000	00:24
6	0.819885	0.867286	0.783333	00:25
7	0.757911	0.878943	0.816667	00:29
8	0.758251	0.895106	0.816667	00:24
9	0.717758	0.809857	0.816667	00:26

Table 1: Resnet34 Outputs

- RESNET50

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	0.916377	0.636152	0.683333	00:36
1	1.056380	0.656015	0.700000	00:32
2	1.157664	0.843288	0.700000	00:35
3	1.158303	1.030600	0.733333	00:32
4	1.044749	0.711210	0.750000	00:34
5	0.995304	0.806717	0.716667	00:33
6	0.985888	0.680070	0.766667	00:32
7	0.981679	0.773015	0.700000	00:35
8	0.933008	0.956950	0.700000	00:32
9	0.850253	0.782971	0.666667	00:35

Table 2: Resnet50 Outputs

- DENSENET121

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	0.985671	0.694144	0.733333	00:29
1	0.754824	0.484870	0.733333	00:29
2	0.678787	0.607804	0.750000	00:31
3	0.763258	0.969738	0.800000	00:31
4	0.735078	0.797341	0.816667	00:32
5	0.694312	0.628933	0.800000	00:28
6	0.662225	0.690574	0.850000	00:29
7	0.647854	0.712517	0.850000	00:31
8	0.626096	0.776893	0.850000	00:30
9	0.565133	0.800234	0.850000	00:32

Table 3: Densenet121 Outputs

- DENSENET169

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	1.228958	0.906763	0.616667	00:35
1	1.051276	0.599612	0.700000	00:38
2	0.969538	1.124437	0.616667	00:34
3	0.895675	1.049062	0.700000	00:37
4	0.855953	0.915649	0.750000	00:36
5	0.763573	1.179581	0.716667	00:35
6	0.702714	0.689903	0.766667	00:37
7	0.679061	0.748759	0.766667	00:34
8	0.612276	0.801563	0.766667	00:37
9	0.568734	0.816435	0.766667	00:35

Table 4: Densenet169 Outputs

- VGG16

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	1.233069	0.557125	0.766667	01:31
1	1.057547	0.424269	0.883333	01:25
2	1.024386	0.565602	0.850000	01:24
3	0.937139	0.651897	0.816667	01:26
4	0.898746	0.466686	0.883333	01:25
5	0.860076	0.561753	0.833333	01:25
6	0.760136	0.663044	0.816667	01:25
7	0.758178	0.634949	0.833333	01:28
8	0.661982	0.644653	0.833333	01:25
9	0.655634	0.663215	0.833333	01:24

Table 5: VGG16 Outputs

- ALEXNET

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	1.360816	0.882201	0.600000	00:06
1	1.202358	0.555810	0.750000	00:05
2	1.062525	0.589067	0.766667	00:07
3	0.884177	0.800327	0.733333	00:06
4	0.900412	0.877342	0.750000	00:04
5	0.859992	0.841593	0.766667	00:05
6	0.793817	0.562621	0.816667	00:06
7	0.781968	0.594367	0.816667	00:04
8	0.790172	0.677045	0.800000	00:05
9	0.797909	0.756086	0.800000	00:06

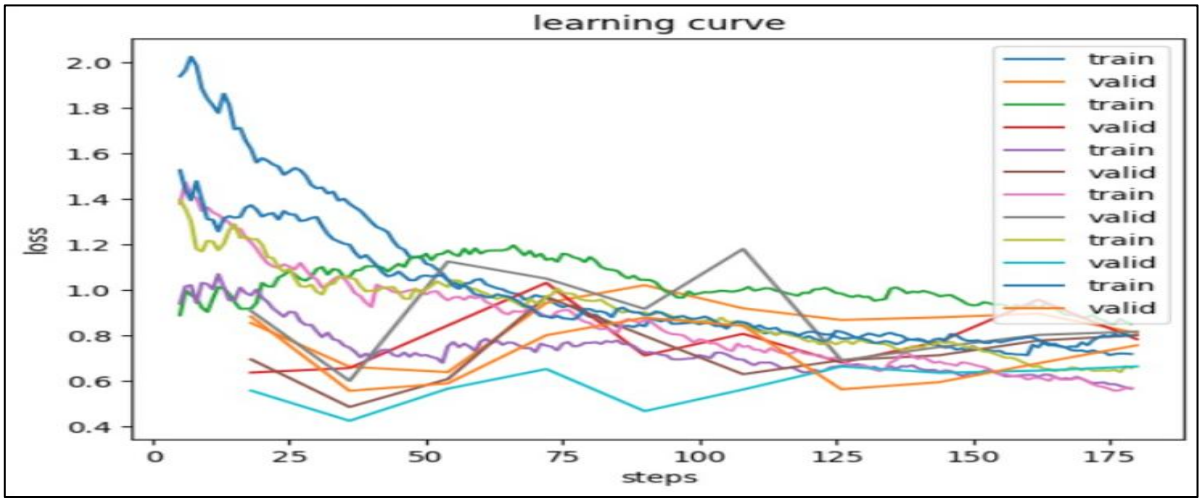
Table 6: AlexNet Outputs

➢ CUSTOM CONVOLUTION NEURAL NETWORK

EPOCH	TRAIN LOSS	VALID LOSS	ACCURACY	TIME
0	1.360816	0.882201	0.5139	00:46
1	1.202358	0.555810	0.5417	00:45
2	1.062525	0.589067	0.5000	00:47
3	0.884177	0.800327	0.5000	00:46
4	0.900412	0.877342	0.5000	00:44
5	0.859992	0.841593	0.5347	00:45
6	0.793817	0.562621	0.6944	00:46

Table 7: Custom Neural Network Outputs

➢ LEARNING CURVE OF VARIOUS METHODS



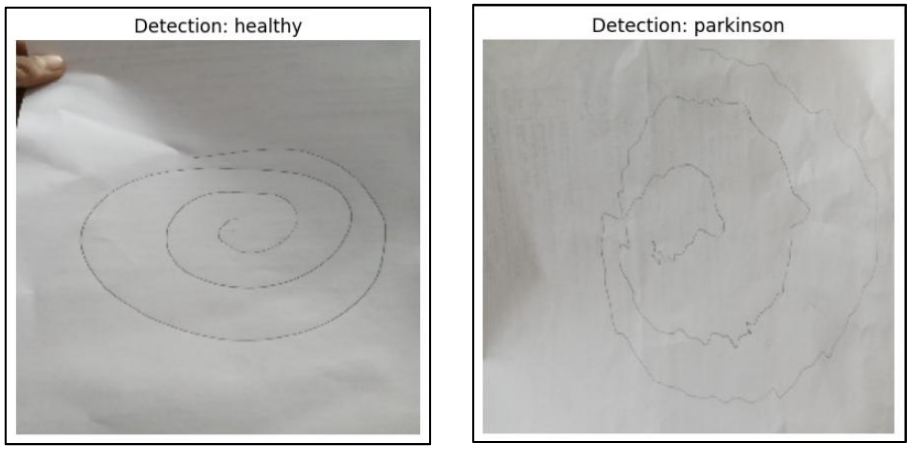


Figure- 37: Spiral Learning Curve and Prediction Result

LEARNING RATE AND CONFUSSION MATRIX

- RESNET34

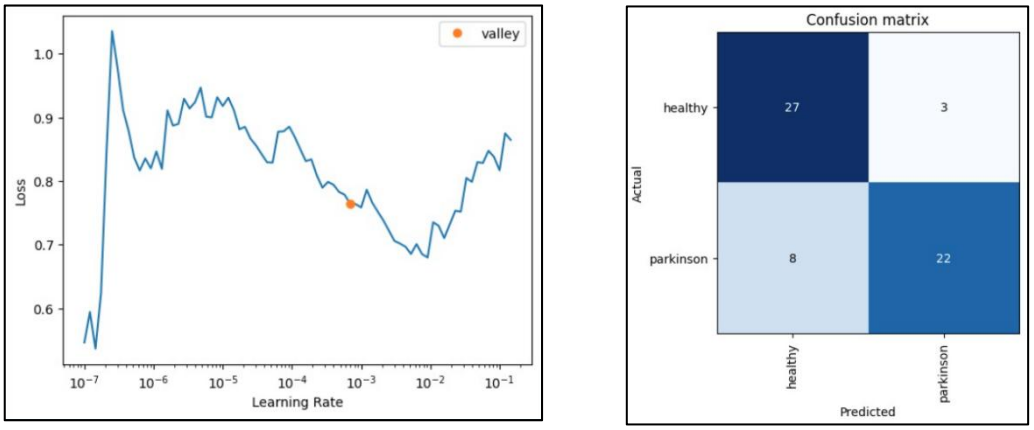


Figure- 38: Learning Rate and Confusion Matrix of Resnet34

- RESENT50

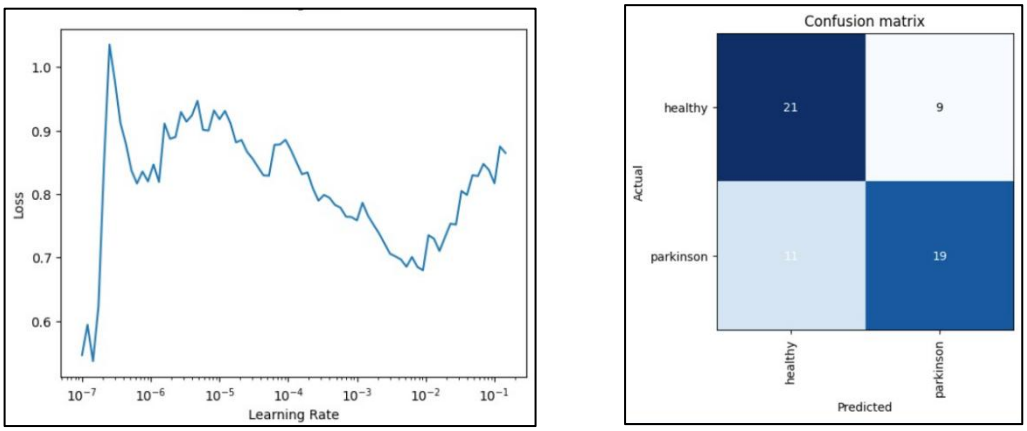


Figure- 39: Learning Rate and Confusion Matrix of Resnet50

- DENSENET121

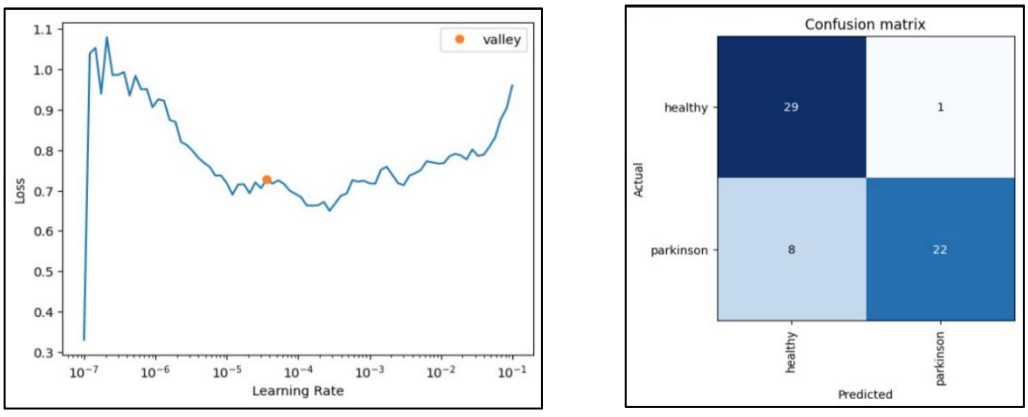
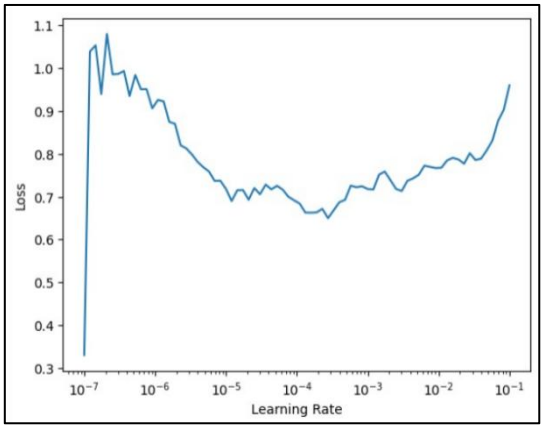


Figure- 40: Learning Rate and Confusion Matrix of Densenet121

- DENSENET169



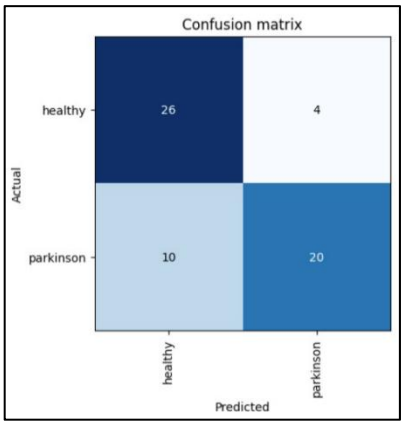
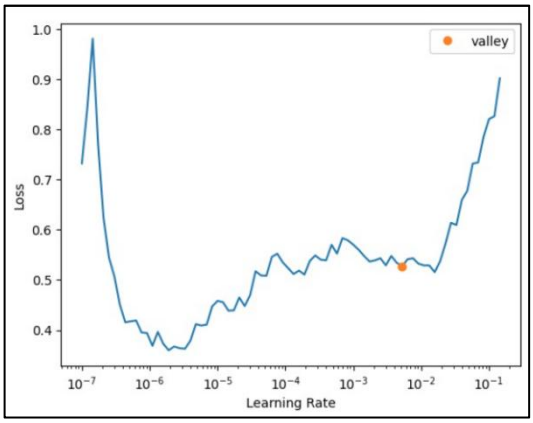


Figure- 41: Learning Rate and Confusion Matrix of Densenet169

- VGG16



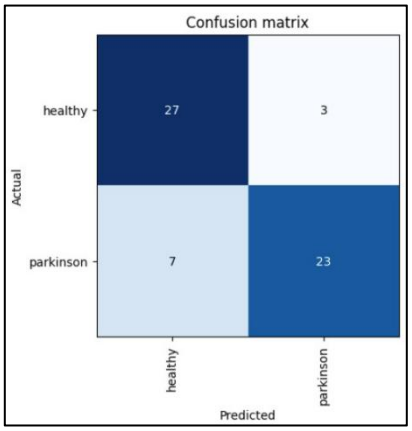
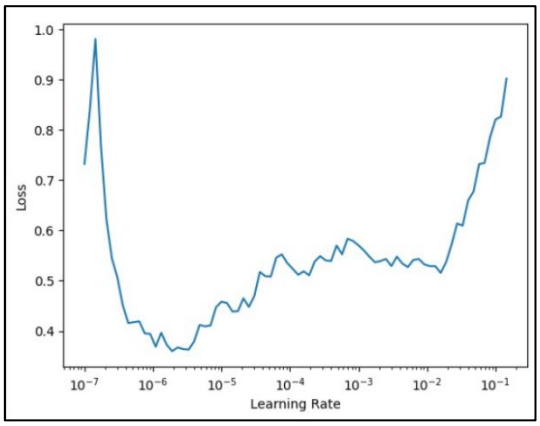


Figure- 42: Learning Rate and Confusion Matrix of VGG16

- ALEXNET



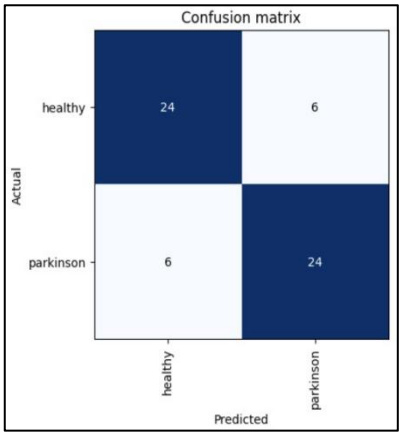


Figure- 43: Learning Rate and Confusion Matrix of AlexNet

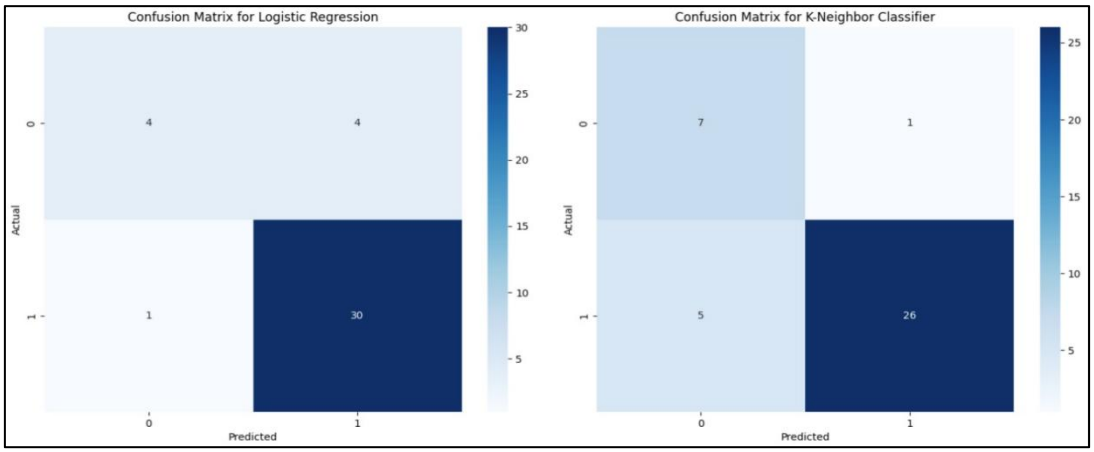
6.2 VOICE SIGNAL ANALYSIS OUTCOMES:

ESSENTIAL PARAMETERS

OUTCOME ALGORITM	ACCURACY	F1 SCORE	PRECESION	LOG LOSS	RECALL SCORE	MEAN ABSOLUTE ERROR
Logistic Regression	0.82051	0.88524	0.9	6.46937	0.87096	0.17948
K-Neighbor Classifier	0.79487	0.85185	1.0	7.39356	0.74193	0.20512
Support Vector Machine	0.89743	0.93939	0.88571	3.69678	1.0	0.10256
Gaussian Naive Bayes	0.61538	0.68085	1.0	13.86294	0.51612	0.38461
Decision Tree Classifier	0.74358	0.81481	0.95652	9.24196	0.70967	0.25641
Random Forest Classifier	0.84615	0.9	0.93103	5.54517	0.87096	0.15384
Gradient Boosting Classifier	0.82051	0.88135	0.92857	6.46937	0.83870	0.17948
Neural Network MLP	0.82051	0.88135	0.92857	6.46937	0.83870	0.17948

Table 8: Voice Signal Parameters and Outputs

CONFUSSION MATRIX



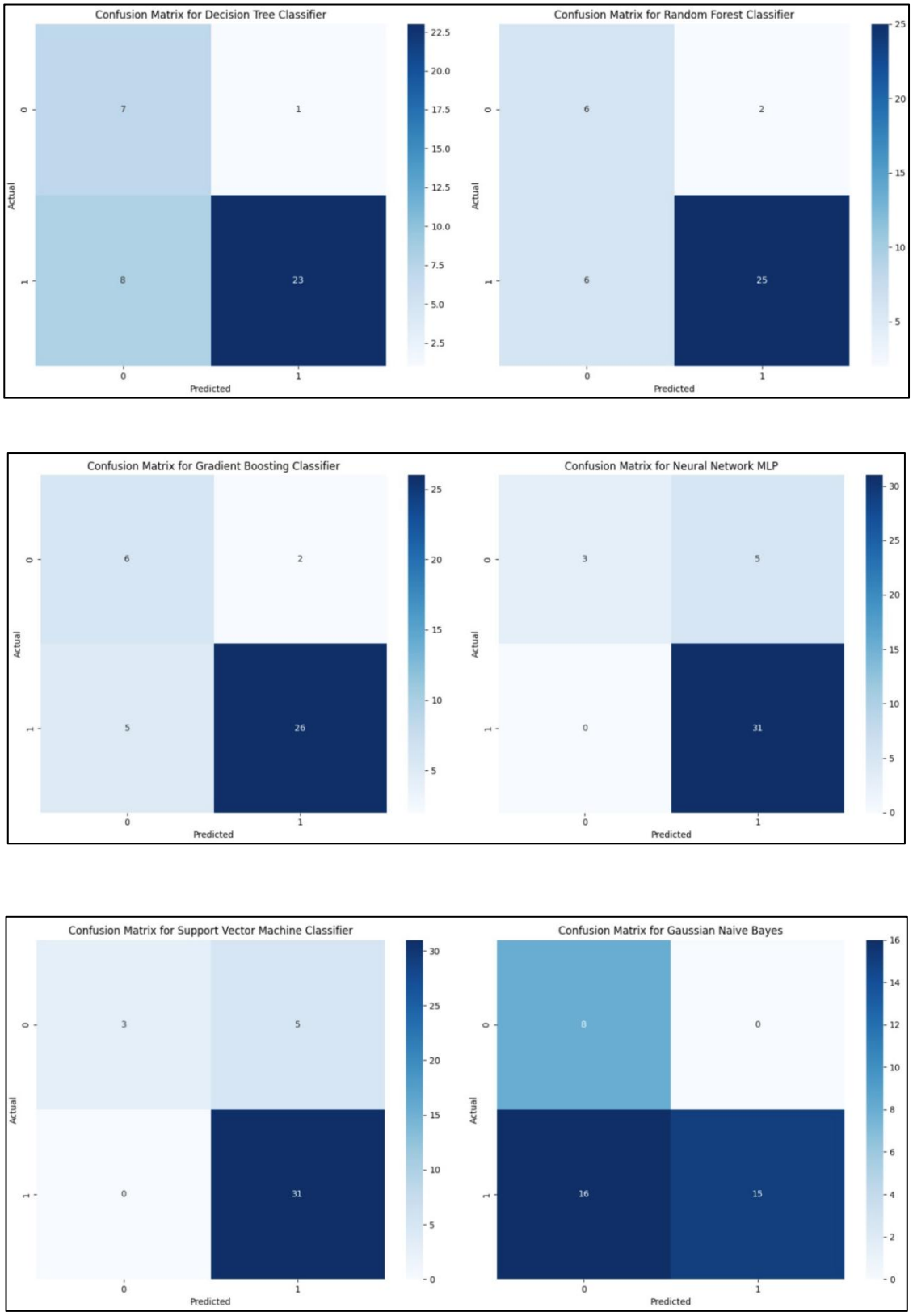


Figure- 44: Confusion Matrix of algorithms trained for voice signal dataset

6.3 TAPPY KEYSTROKE ANALYSIS OUTCOMES:

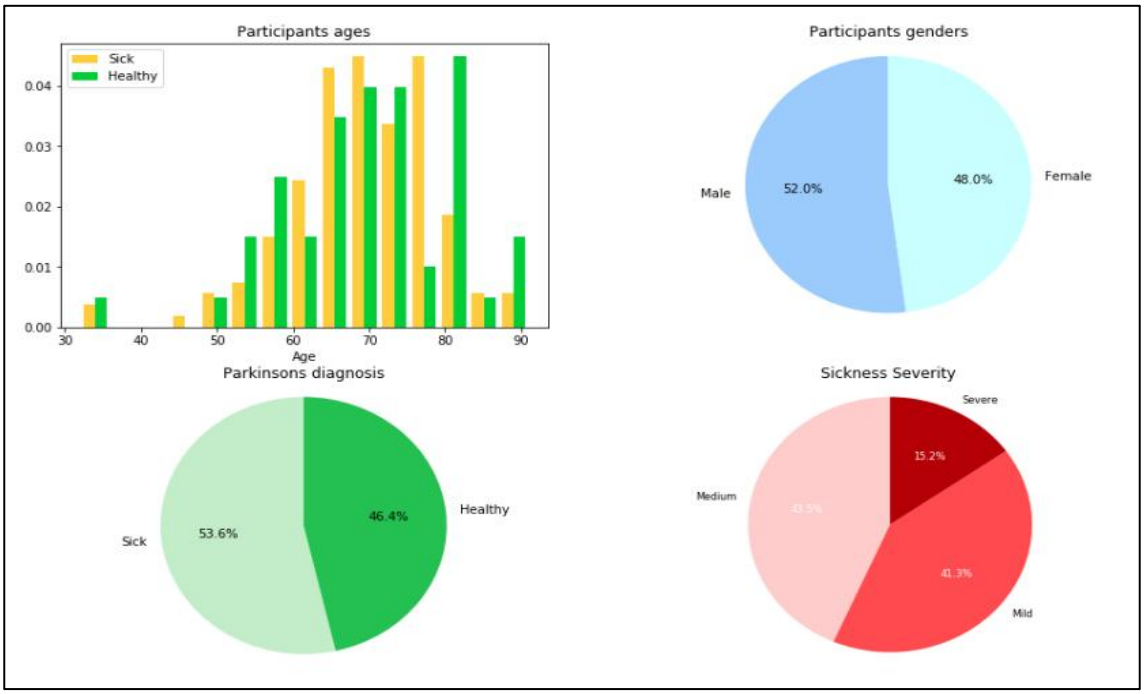


Figure- 45: Specifications of Participants and their corresponding parameters

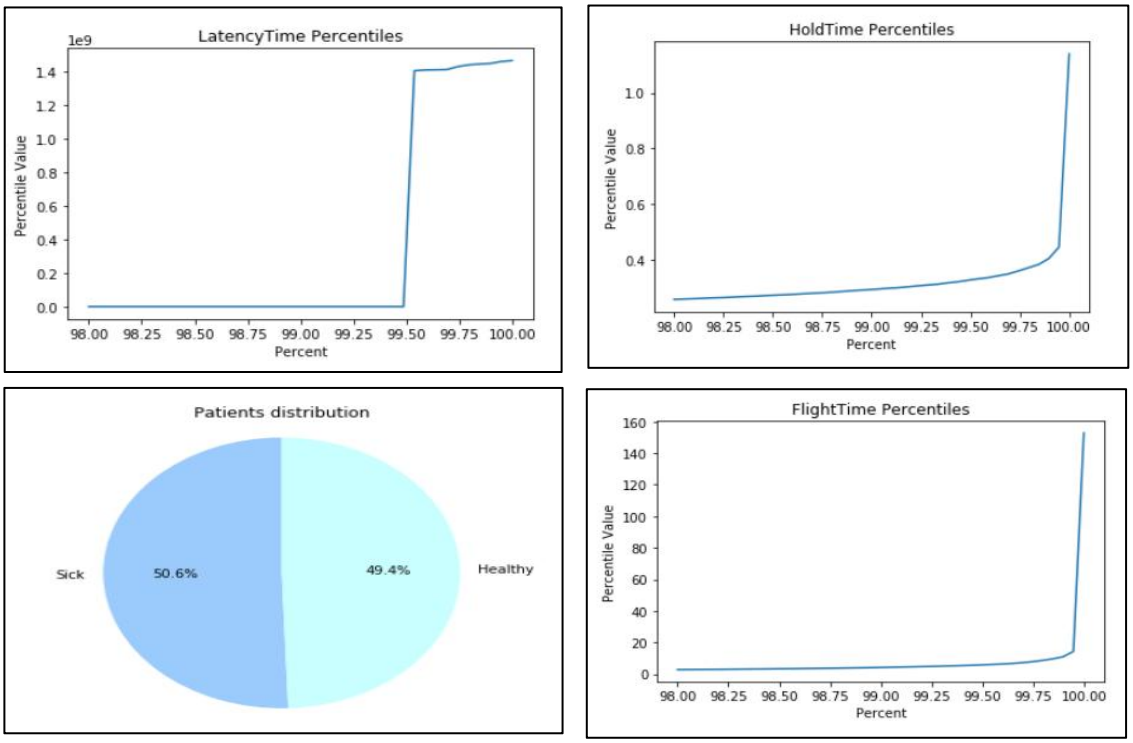


Figure- 46: Latency, Hold and Flight Time Outcomes with Patient Distribution

Algorithm | Outcome	Train Accuracy	Test Accuracy
Logistic Regression	0.75	0.5018
K-Neighbor Classifier	0.6538	0.58
Support Vector Machine	0.6346	0.6364
AdaBoost Classifier	1.0	0.5564
Gradient Boost Classifier	1.0	0.5364
Random Forest Classifier	1.0	0.5764

Table 9: Train and Test Accuracy of various Algorithms for Gait Analysis

6.4 GAIT ANALYSIS OUTCOMES:

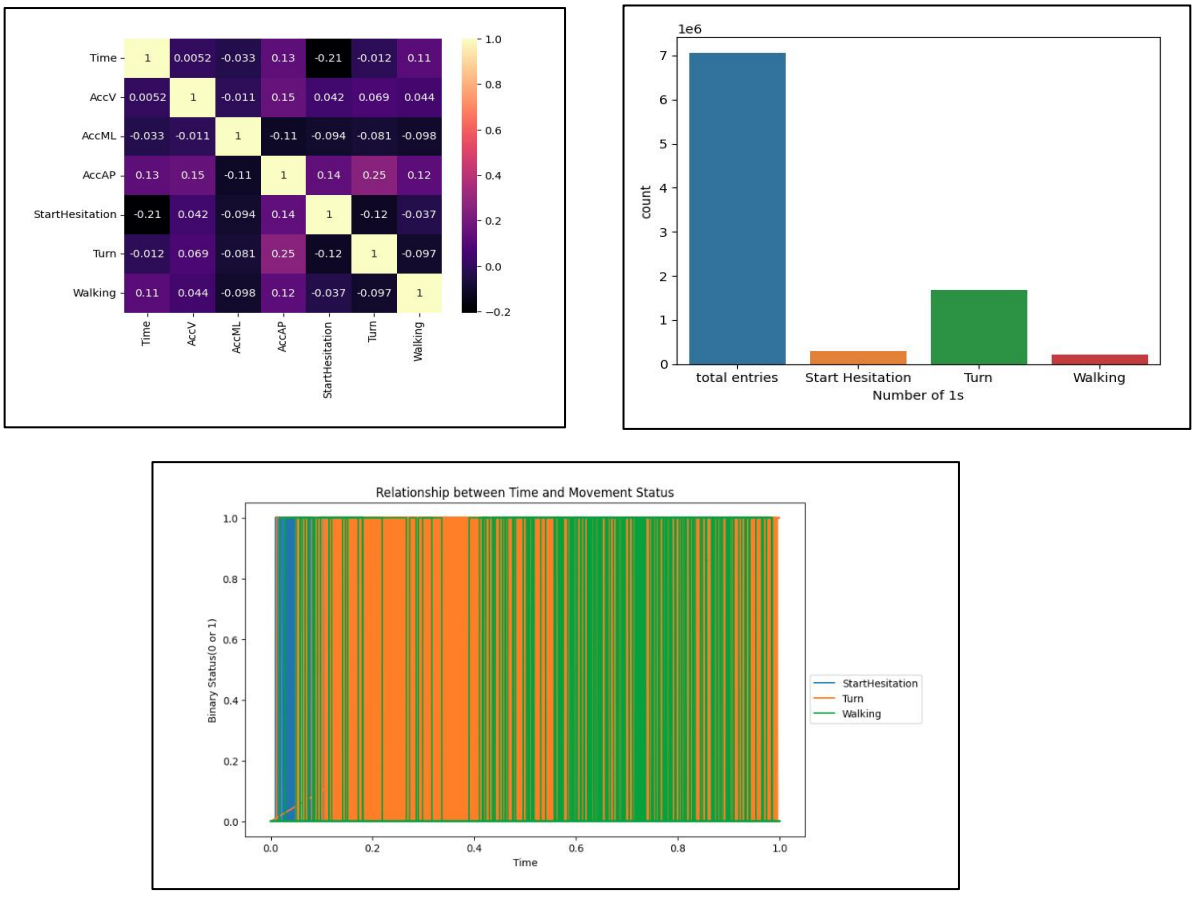
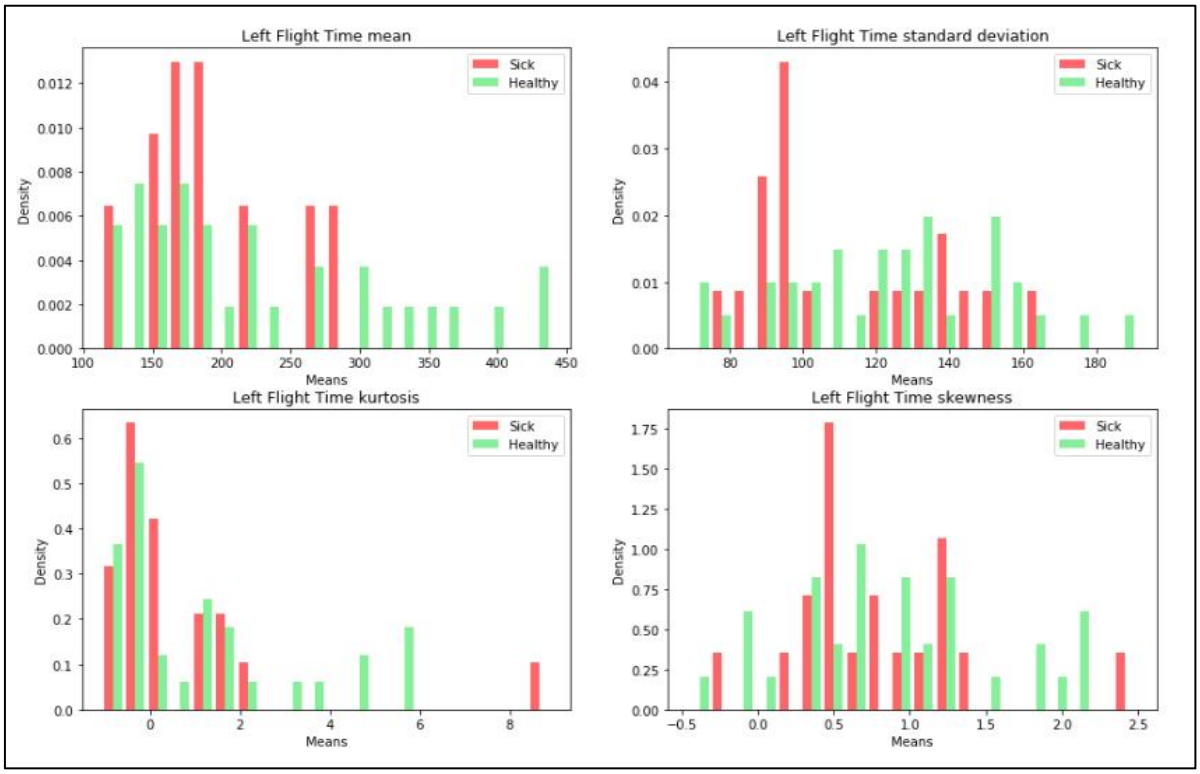


Figure- 47: Gait Analysis Outcomes



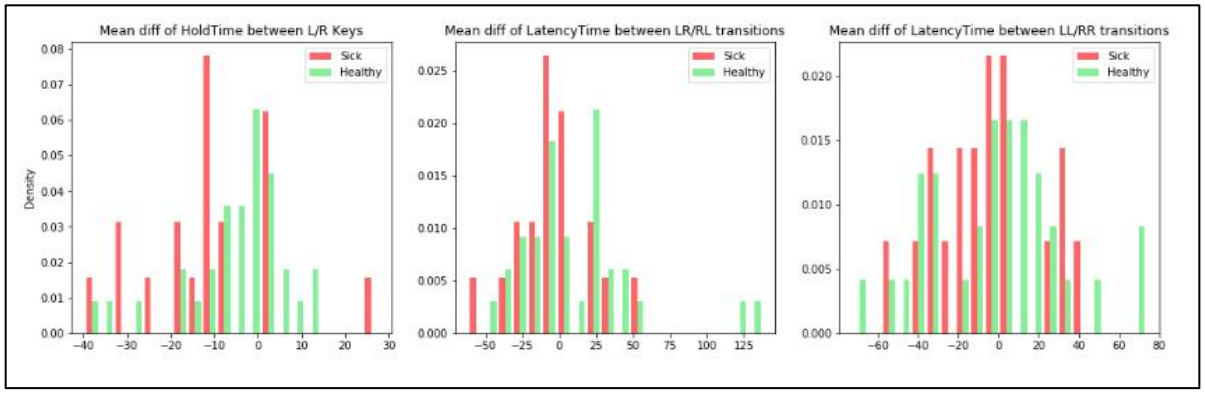


Figure- 48: Mean Difference of Different Parameters

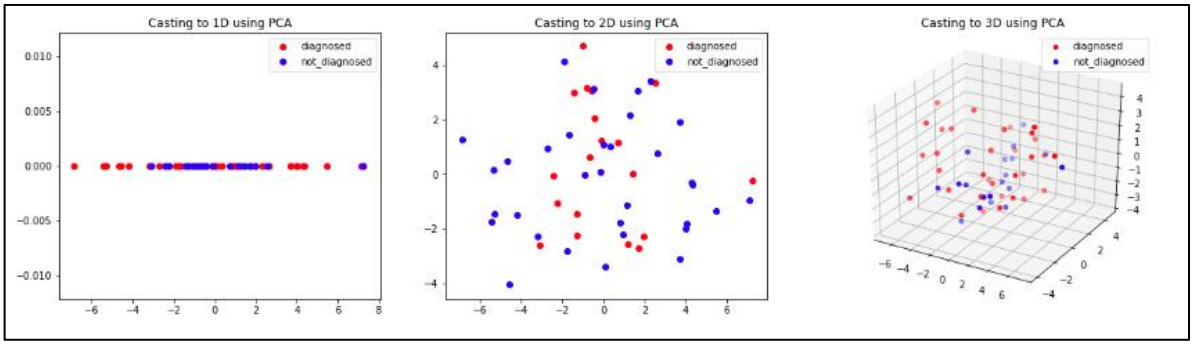


Figure- 49: Casting to Different Dimensions

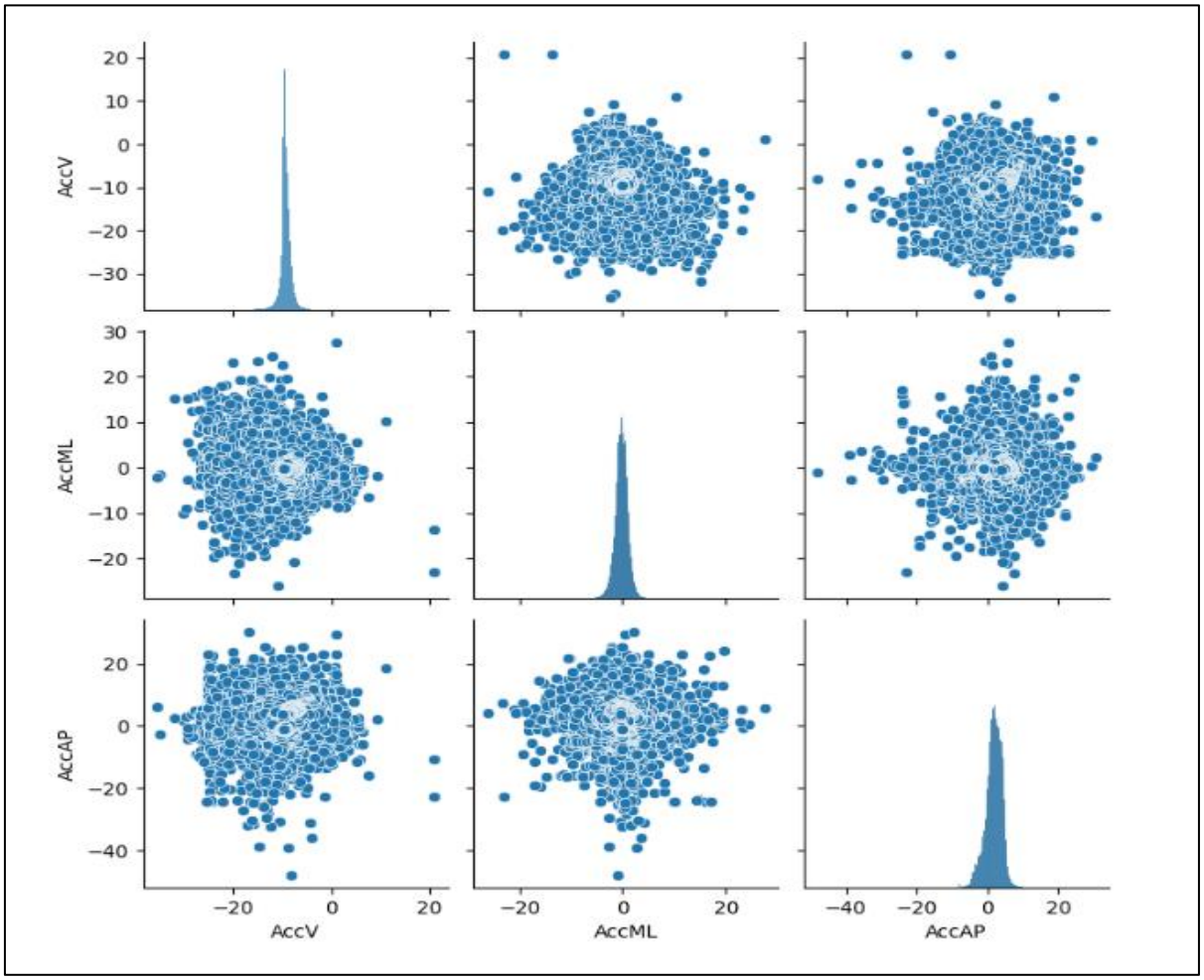


Figure- 50: Peak points plot of acceleration Vertical, Mediolateral and Anteroposterior components

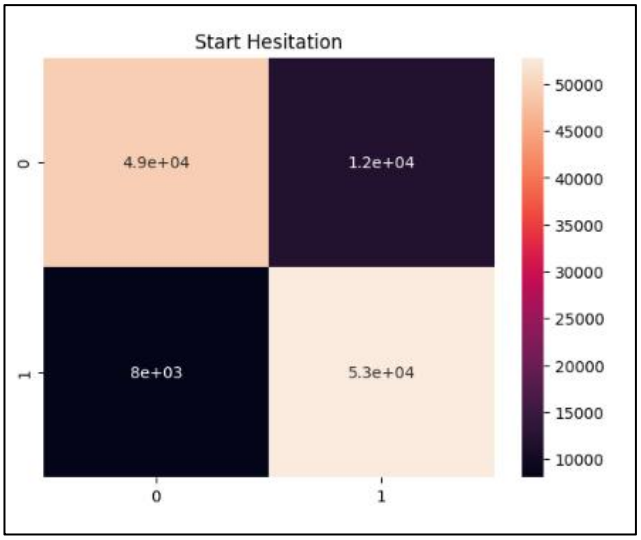


Figure- 51: Confusion Matrix for Start Hesitation

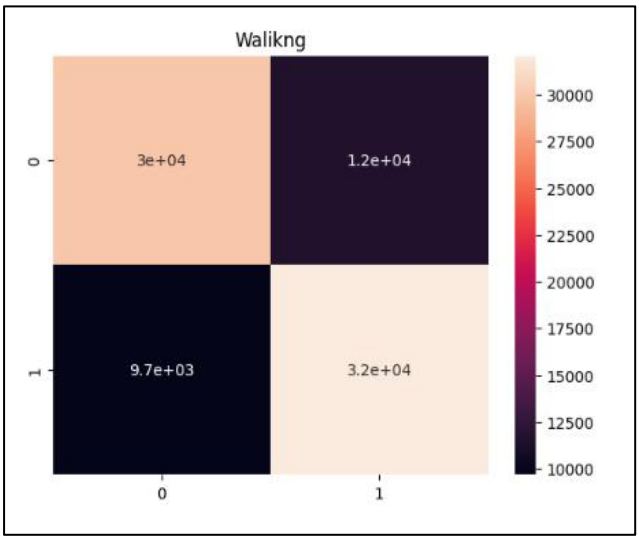


Figure- 52: Confusion Matrix for Walking

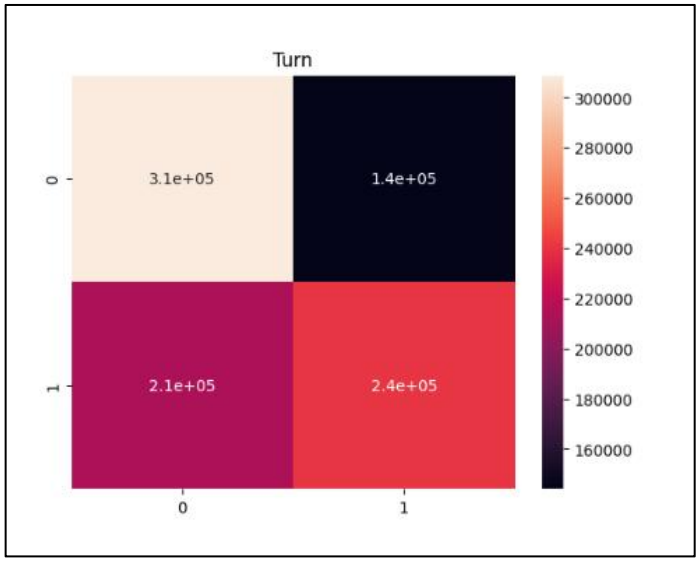


Figure- 53: Confusion Matrix for Turn

START HESITATION

	Precision	Recall	F1-score	Support
0	0.86	0.81	0.83	61262
1	0.82	0.87	0.84	60854
Accuracy			0.84	122116
Macro Average	0.84	0.84	0.84	122116
Weighted Average	0.84	0.84	0.84	122116

TURN

	Precision	Recall	F1-score	Support
0	0.59	0.68	0.63	452791
1	0.63	0.53	0.57	45354
Accuracy			0.61	906245
Macro Average	0.61	0.61	0.60	906245
Weighted Average	0.61	0.61	0.60	906245

WALKING

	Precision	Recall	F1-score	Support
0	0.76	0.72	0.74	41371
1	0.74	0.77	0.75	41765
Accuracy			0.74	83136
Macro Average	0.75	0.74	0.74	83136
Weighted Average	0.75	0.74	0.74	83136

Table 10: Measure of accuracy and average values during start turn and walking conditions

6.5 USER-WEB INTERFACE:

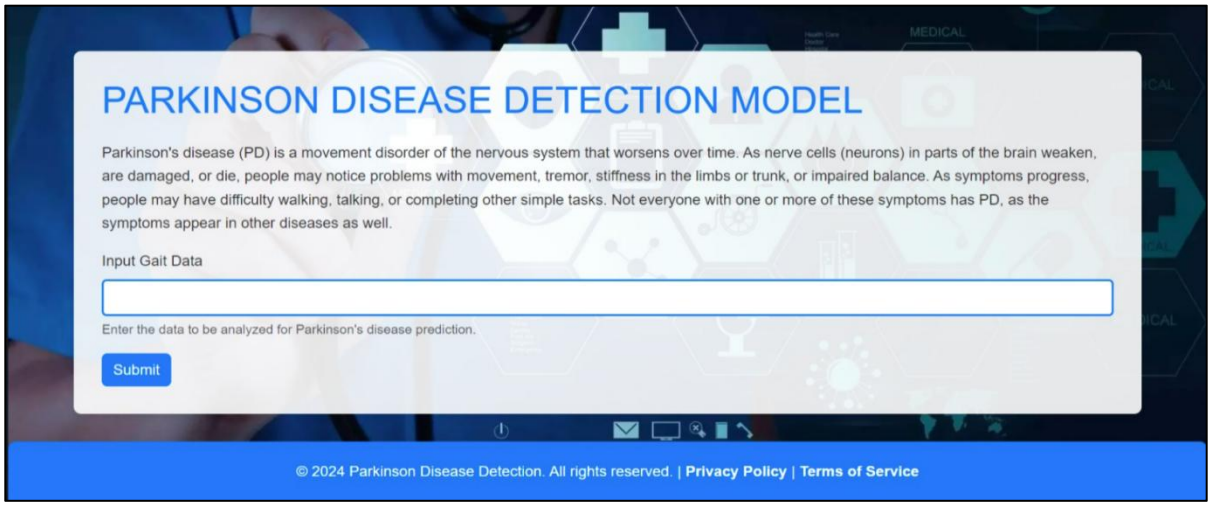


Figure - 54: Web Interface for Gait Data Input

Figure - 55: Web Interface for Waves and Spiral drawing input

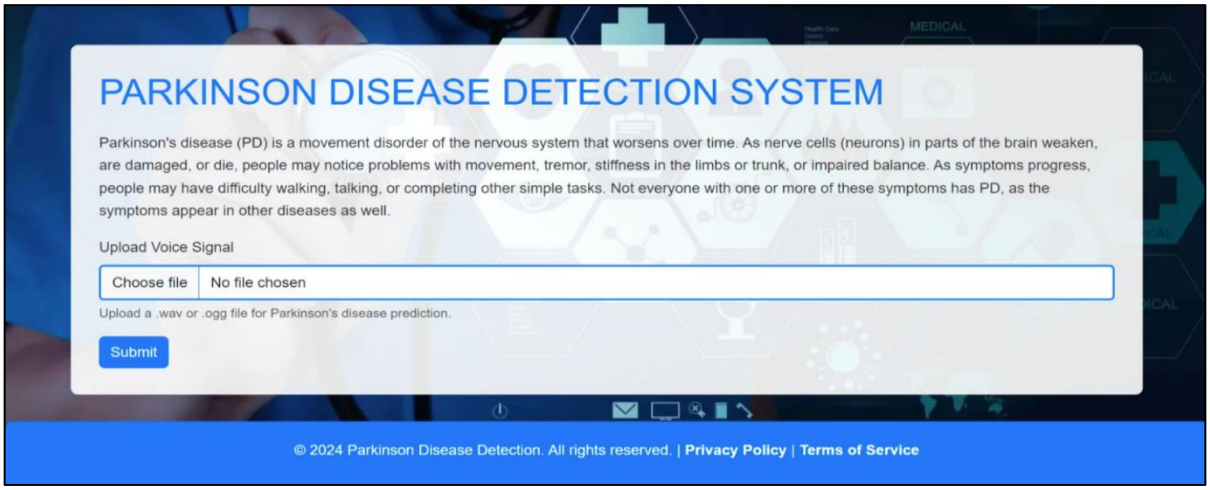


Figure - 56: Web Interface for Voice Signal Input

KEY TAKEAWAYS:

Based on a comprehensive study, the best machine learning models for Parkinson's disease prediction across various datasets have been identified:

For spiral and wave analysis, ResNet-34 is the most effective model, achieving a peak accuracy of 81.67% with a validation loss of 0.810 and a training time of 26 seconds per epoch. It balances accuracy and training time, and mitigates vanishing gradient issues through residual learning. ResNet-50 was also considered but had a lower accuracy of 66.67% and a longer training time of 35 seconds per epoch. In voice signal analysis, the Support Vector Machine (SVM) stands out with an accuracy of 89.74%, an F1 score of 93.94%, precision of 88.57%, recall of 100%, and a mean absolute error of 10.26%. The high recall score indicates its effectiveness in identifying positive instances, crucial for early detection.

For tappy keystroke analysis, the Gradient Boosting Regressor and Voting Classifier are recommended. The Gradient Boosting Regressor enhances performance by leveraging weak learners, while the Voting Classifier combines predictions from multiple models to improve accuracy. Both models enhance robustness and prediction accuracy through their ensemble nature. In gait analysis, LightGBM (Light Gradient Boosting Machine) emerged as the preferred model due to its fast training times, low memory usage, and suitability for large datasets and high-dimensional data. LightGBM's efficiency and accuracy make it ideal for handling complex gait data.

The implementation includes Jupyter Notebooks for interactive development and visualization, Flask for building the user interface and handling HTTP requests, and Blynk IoT Cloud Architecture for collecting and visualizing real-time data. HTML and CSS are used for designing and styling the web interface, ensuring a user-friendly and visually appealing experience. These models and tools collectively offer a robust solution for early prediction and monitoring of Parkinson's disease, leveraging advanced machine learning techniques and real-time data analysis for effective clinical application. This approach facilitates early intervention and enables personalized treatment strategies, ultimately improving clinical outcomes for Parkinson's disease patients.

CHAPTER 7

CONCLUSION

The project presents a ground-breaking approach to the early detection and management of Parkinson's disease (PD) through a multifaceted strategy incorporating voice analysis, wave and spiral drawing assessments, typing keystroke dynamics, and gait analysis. This comprehensive method addresses the limitations of current diagnostic practices, which often rely on subjective clinical assessments and are prone to misdiagnosis or delays.

By leveraging advanced data integration and machine learning methodologies, the project enhances the accuracy and efficiency of PD detection. The integration of speech and tremor characteristics, fine motor control assessments, and gait analysis provides a holistic understanding of the disease's progression, enabling earlier and more precise diagnoses. Additionally, the incorporation of typing keystroke features adds another layer of data, further improving the system's sensitivity and specificity.

The development of a user-friendly web interface ensures that both patients and healthcare providers can easily interact with the system. Patients can upload their voice recordings, gait data, typing keystroke data, and spiral and wave drawings, receiving real-time feedback and visualizations of their condition. Healthcare providers can access detailed reports and analytics, facilitating remote monitoring and personalized treatment plans. This seamless interaction enhances the accuracy of PD severity assessments and empowers patients to actively participate in their healthcare management.

Ultimately, the project's goal is to improve patient outcomes and quality of life by enabling earlier detection and intervention. Timely diagnosis allows for the initiation of appropriate treatments, symptom management strategies, and lifestyle modifications, which can slow disease progression and optimize patient care. The innovative approach outlined in this project combines physiological measurements and motor control assessments to revolutionize PD detection and management, offering significant potential to impact the lives of individuals affected by this debilitating disease.

CHAPTER 8

FUTURE SCOPE

The future scope of this project encompasses several promising avenues for further development and refinement. One of the most significant advancements would be the integration with wearable devices. By incorporating data from smartwatches, fitness trackers, and other wearable technologies, the system can continuously monitor patients' motor and non-motor symptoms in real-time. This continuous monitoring can provide dynamic and responsive disease management, allowing for timely interventions and adjustments in treatment plans. Real-time data collection from wearable devices can also help in identifying early signs of disease progression, enabling proactive measures to be taken to slow down the disease's impact.

Another critical area for expansion is the inclusion of additional data sources. Future iterations of the system could integrate genetic information, environmental factors, and biomarkers from biological fluids such as blood, cerebrospinal fluid, and saliva. This holistic approach would enhance the comprehensiveness of the diagnostic process, potentially uncovering new biomarkers and risk factors associated with Parkinson's disease. By broadening the scope of data, the system can offer a more precise and personalized diagnosis, improving patient outcomes. This integration would also facilitate the identification of individual risk profiles, enabling tailored preventive strategies.

The development of personalized treatment plans is another exciting prospect. By leveraging advanced machine learning algorithms, the system could tailor medication regimens, physical therapy programs, and lifestyle recommendations to each patient's unique needs and disease progression. This personalized approach ensures that patients receive the most effective treatments, minimizing side effects and maximizing therapeutic benefits. Personalized treatment plans could also adapt over time based on continuous monitoring data, ensuring that the interventions remain effective as the disease evolves.

Extensive clinical trials and validation studies are essential for further refining and validating the diagnostic algorithms. Collaborations with healthcare institutions and research organizations will be crucial to ensure the system's accuracy, reliability, and clinical utility.

These partnerships can facilitate large-scale studies that provide robust data, ultimately leading to regulatory approvals and widespread adoption of the system in clinical settings. Engaging in multi center trials would also help in understanding the system's performance across diverse populations and healthcare settings, ensuring its generalizability and robustness.

Efforts should also be made to adapt and implement the system in diverse healthcare environments worldwide. Addressing potential challenges related to data privacy, regulatory requirements, and integration with existing healthcare infrastructure is crucial for global implementation. By making the system accessible to a broader population, it can significantly impact Parkinson's disease management on a global scale. This involves not only technical adaptations but also cultural and language modifications to ensure the system's usability in various regions.

Incorporating advanced artificial intelligence (AI) and deep learning techniques can further enhance the system's capabilities. These technologies can detect subtle patterns and anomalies in the data, improving diagnostic accuracy and the ability to predict disease progression and treatment responses. Continuous advancements in AI will ensure the system remains at the forefront of medical innovation. AI can also facilitate the discovery of new therapeutic targets by analyzing complex datasets and identifying novel disease mechanisms.

Patient education and support are also vital components of the future scope. Developing educational resources and support tools for patients and caregivers can help them better understand their condition, manage symptoms, and maintain a high quality of life. Empowering patients with knowledge and support can lead to more proactive and engaged healthcare management. This can be achieved through the creation of user-friendly mobile apps and online platforms that provide access to educational materials, community support groups, and tools for tracking symptoms and treatment progress.

Lastly, ongoing collaboration with healthcare providers is essential to ensure the system meets the needs of clinicians and integrates seamlessly into clinical workflows. Regular feedback from users will guide continuous improvements and updates to the system, ensuring it remains user-friendly and effective.

BIBLIOGRAPHY

1. *M. Aghzal and A. Mourhir, "Early Diagnosis of Parkinson's Disease based on Handwritten Patternsusing Deep Learning," 2020 Fourth InternationalConference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 2020, pp. 1-6, doi: 10.1109/ICDS50568.2020.9268738.*

2. *A. Hussain and A. Sharma, "Machine Learning Techniques for Voice-based Early Detection of Parkinson's Disease," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida,India, 2022,pp.1436-1439,doi: 10.1109/ICACITE53722.2022.9823467.*

3. *M. Nithya, V. Lalitha, K. Paveethra and S. Kumari, "Early Detection of Parkinson's Disease using Machine Learning & Image Processing," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore,India,2022,pp.1-4,doi:10.1109/ICCCI 54379. 2022.9740961.*

4. *K. Polat, "A Hybrid Approach to Parkinson Disease Classification Using Speech Signal: The Combinationof SMOTE and Random Forests," 2019 ScientificMeeting on Electrical Electronics & BiomedicalEngineering and Computer Science (EBBT), Istanbul,Turkey,2019,pp.1-3,doi: 10.1109/EBBT.2019.8741725.*

5. *M. Bharti, S. S. Gupta, V. Kumar, and R. Bhatnagar, "A Comprehensive Study of Parkinson's Disease Detection Using Machine Learning Techniques," in 2018International Conference on Communication,Computing and Internet of Things (IC3IoT), 2018, pp. 564 -569*

6. *A. Bhattacharya, S. Roy, S. Ghosh, and D. K. Bhattacharya, "Parkinson's Disease Diagnosis fromSpiral Test using Machine Learning," IEEE Transactions on Neural Systems and RehabilitationEngineering, vol. 28,no. 5, pp. 1154-1162, 2020.*

7. *M. E. Khan, S. Lee, and D. Lee, "ECG-based Parkinson's disease detection using a convolutional neural network," in 2019 International Conference on Information and Communication Technology Convergence (ICTC), 2019, pp. 160-165.*

8. *A. K. M. Shahjahan, M. M. Kabir, and M. M. Hasan, "Early detection of Parkinson's disease using machine learning techniques," in 2021 7th International Conference on Advances in Electrical Engineering (ICAEE), 2021, pp. 1-6.*

9. *M. Rahman, M. Hossain, and T. Ahmed "Classification of Parkinson's disease from spiral test data using machine learning techniques," in 2021 3rd International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2021, pp. 1-6.*

10. *M. H. U. Bhuiyan, M. A. M. Sohel, S. Ahmed, and S. Islam, "Parkinson's disease detection using transfer learning from deep convolutional neural network," in 2021 11th International Conference on Communications Systems and Networks (COMSNETS), 2021, pp. 570-573.*

11. *S. Paul, S. M. Aziz, and M. R. A. Bhuiyan, "An ensemble deep learning approach for Parkinson'sdisease detection using EEG and EMG signals," in 2021 IEEE International Conference on Electro/Information Technology (EIT), 2021, pp. 0116-0121.*

12. *2021 3rd International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2021, pp. 1-6.*

13. *Shikha Tripathi, MSc, Teresa Arroyo-Gallego, PhD, and Luca Giancardo, PhD "Keystroke-Dynamics for Parkinson's Disease Signs Detection in An At-Home Uncontrolled Population: A New Benchmark and Method." PMCID: PMC9904385, NIHMSID: NIHMS1860974, PMID: 35767495.*

14. *Alexandra-Georgiana Andrei; Alexandra-Maria Tăuțan; Bogdan Ionescu "Parkinson's Disease Detection from Gait Patterns" IEEE / 2019 E-Health and Bioengineering Conference (EHB) / Iasi, Romania / 21-23 November 2019/ 28 January 2020 / DOI: 10.1109/EHB47216.2019.8969942*

15. *T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi, "Parkinson's Disease Diagnosis Using Machine Learning and Voice," 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, DOI: 10.1109/SPMB.2018.8615607.*

16. *Muskan Bahrani,Meet Chhabria, Kaustubh Kharche,Sakshi Shinde,Dr Prashant Kanade " Voice-Based Parkinson's Disease Prediction Using Machine Learning ", © 2024 IJNRD | Volume 9, Issue 4 April 2024| ISSN: 2456-4184 | IJNRD.ORG*

17. *D. Ambujam Vigneswari; J Aravinth, "Parkinson's disease Diagnosis using Voice Signals by Machine Learning Approach" / IEEE, 27-28 August 2021, Bangalore, India, DOI: 10.1109/RTEICT52294.2021.9573689.*

18. *Ana Paula Janner Zanardi, Edson Soares da Silva, Rochelle Rocha Costa, Elren Passos-Monteiro, Ivan Oliveira dos Santos, Luiz Fernando & Leonardo Alexandre Peyré-Tartaruga, "Gait parameters of Parkinson's disease compared with healthy controls: a systematic review and meta-analysis", DOI:10.1038/s41598-020-80768-2.*

19. *M.S. Roobini; Yaragundla Rajesh Kumar Reddy; Udayagiri Sushmanth Girish Royal; Amandeep K Singh; K Babu, "Parkinson's Disease Detection Using Machine Learning",IEEE, 10-11 March 2022, Chennai, India, 12 May 2022, DOI: 10.1109/IC3IOT53935.2022.9768002.*

20. *Sabyasachi Chakraborty; Satyabrata Aich; Jong-Seong-Sim; Eunyoung Han; Jinse Park; Hee-Cheol Kim " Parkinson's Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach ", 2020, Phoenix Park, Korea (South) DOI:10.23919/ICACT48636.2020.9061497*

21. *Shivangi; Anubhav Johri; Ashish Tripathi, "Parkinson Disease Detection Using Deep Neural Networks", IEEE, 2019 Twelfth International Conference on Contemporary Computing (IC3), 08-10 August 2019, Noida, India, DOI: 10.1109/IC3.2019.8844941.*

BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT

AVALAHALLI, DODDABALLAPUR MAIN ROAD, BANGALORE-64

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

STUDENTS PROFILE

SL NO	NAME	USN	PHOTOGRAPH	MAIL ID	PHONE NUMBER
01	Adithya R Hiremath	1BY21EC004		adithyahiremathr3@gmail.com	6366063535
02	Basavaprabhu S H	1BY21EC027		bshhalapeti2002@gmail.com	8197889414
03	Bharath Vasishta R	1BY21EC029		bharathvasishtar3@gmail.com	7892126372
04	G P Varshini	1BY21EC045		varshinigp25@gmail.com	7975010351