

**VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF
TECHNOLOGY**

**An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering**



Project Report on

Health+ : “Your Digital Health Guardian”

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2024-25

Submitted by

Vidisha Jadhwani D17A - 26

Riddhi Labde D17B -25

Priti Shamnani D17B -50

Nikhil Makhija D17B -27

Project Mentor

Dr. Dashrath Mane

(2024-25)

**VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF
TECHNOLOGY**

**An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering**



Certificate

This is to certify that ***Vidisha Jadhwanı, Riddhi Labde, Priti Shamnani, Nikhil Makhija*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on ***Health+ : “Your Digital Health Guardian”*** as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Dr. Dashrath Mane*** in the year 2024-25 .

This thesis/dissertation/project report entitled ***Health+ : “Your Digital Health Guardian” by Vidisha Jadhwanı, Riddhi Labde, Priti Shamnani, Nikhil Makhija*** is approved for the degree of ***B.E. Computer Engineering***.

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

Project Report Approval

For

B. E (Computer Engineering)

This thesis/dissertation/project report entitled ***Health+ : “Your Digital Health Guardian*** by ***Vidisha Jadhwanı , Riddhi Labde , Priti Shamnani , Nikhil Makhija*** is approved for the degree of **B.E. Computer Engineering**.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Vidisha Jadhwan - D17A- 26)

(Riddhi Labde - D17B- 25)

(Priti Shamnani - D17B- 50)

(Nikhil Makhija - D17B- 27)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Dr.(Mr.) Dashrath Mane** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to the Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair**, for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

Chapter No	Title	Page No
1	Introduction	
1.1	Introduction to the project	2
1.2	Motivation for the project	2
1.3	Problem Definition	3
1.4	Existing Systems	3
1.5	Lacuna of the Existing Systems	4
1.6	Relevance of the Project	4
2	Literature Survey	
A	Brief overview of Literature Survey	5
2.1	Research Papers Abstract of the research paper Inference drawn from the paper	5
2.2	Patent Search	6
2.3	Inference Drawn	10
2.4	Comparison with the Existing Systems	11
3.	Requirement Gathering for the Proposed System	
3.1	Introduction to Requirement Gathering	12
3.2	Functional Requirements	13
3.3	Non-Functional Requirements	14
3.4	Hardware, Software, Technology and tools utilised	15
3.5	Constraints	16
4.	Proposed Design	
4.1	Block diagram of the system	17
4.2	Modular design of the system	19
4.3	Detailed Design	22

4.4	Project Scheduling & Tracking using Timeline / Gantt Chart	24
5.	Implementation of the Proposed System	
5.1	Methodology employed for development	25
5.2	Algorithms and flowcharts for the respective modules developed	26
5.3	Datasets source and utilization	29
6.	Testing of the Proposed System	
6.1	Introduction to testing	32
6.2	Types of tests Considered	32
6.3	Various test case scenarios considered	33
7.	Results and Discussions	
7.1	Screenshots of User Interface (UI) for the respective module	35
7.2	Performance Evaluation measures	38
7.3	Input Parameters / Features considered	39
7.4	Comparison of results with existing systems	40
7.5	Inference drawn	41
8.	Conclusion	
8.1	Limitations	42
8.2	Conclusion	42
8.3	Future Scope	43
	References	44
	Appendix	
1	Paper I	
a	Paper I	46
b	Plagiarism Report of Paper I	55
c	Project review sheet	56

LIST OF FIGURES

Figure No	Heading	Page No
4.1	Block diagram of the System	17
4.2.1	Modular Diagram of Parkinson Disease	19
4.2.2	Modular Diagram of Lung Cancer	20
4.3	Detailed Design of the System	22
4.4	Gantt chart	24
5.2.1	Flowchart of VGG16	27
5.2.2	Flowchart of Parkinson's Detection	29
5.3.1	Parkinson Disease Dataset	30
5.3.2	Lung Cancer Dataset	31
7.1.1	Home Page	35
7.1.2	Registration Page	35
7.1.3	Login Page	36
7.1.4	Lung Cancer Detection using CT Scan interface	36
7.1.5	Parkinson's Disease Detection Using Voice	37
7.1.6	Patient Dashboard Interface	37
7.1.7	Doctor's Appointment Dashboard	38

LIST OF TABLES

Table No	Heading	Page No
1	Requirements of the system	13
2	Comparison of Results with Existing System	41

Abstract

Early detection of diseases like lung cancer and Parkinson's disease is essential for improving treatment success and patient survival rates. Traditional hospital-based diagnostic procedures often involve multiple tests, expert analysis, and prolonged waiting times, leading to delays in medical intervention. This study introduces a machine learning-based approach that utilizes computed tomography (CT) scan images for lung cancer prediction and voice analysis for Parkinson's disease detection. By automating the evaluation of medical imaging and speech patterns, the proposed system aims to provide faster and more accurate risk assessments, reducing dependency on conventional diagnostic methods. The integration of artificial intelligence (AI) in healthcare enables efficient identification of disease patterns, minimizing human error and facilitating early diagnosis.

For lung cancer detection, the system processes CT scan images to identify and classify potential cancerous nodules. A deep learning model is trained to extract critical features from these scans, distinguishing between malignant and benign tissue with high precision. By leveraging a large dataset of annotated medical images, the model continuously improves its predictive accuracy. The automated nature of this approach not only speeds up the diagnostic process but also assists radiologists in making more informed decisions. Similarly, for Parkinson's disease detection, the system analyzes vocal biomarkers obtained from speech recordings. Machine learning algorithms assess variations in voice tone, pitch, and articulation, which serve as early indicators of neurodegenerative disorders. This non-invasive method provides a convenient screening tool for detecting Parkinson's disease at an early stage, enabling timely medical intervention.

To enhance accessibility, a web-based platform has been developed, allowing users to upload CT scan images or voice recordings for immediate analysis. This system is particularly beneficial for individuals residing in remote areas with limited access to specialized healthcare services. By reducing the need for unnecessary clinical visits and lengthy diagnostic procedures, the proposed approach optimizes healthcare resources and ensures that high-risk patients receive prompt attention. The integration of AI-driven analysis in disease detection not only improves diagnostic accuracy but also contributes to better patient outcomes by enabling early-stage intervention and treatment planning.

Chapter 1: Introduction

1.1. Introduction:

Lung cancer and Parkinson's disease are severe health conditions that require early detection for effective treatment and improved patient outcomes. Lung cancer, often diagnosed at an advanced stage due to the absence of early symptoms, remains one of the leading causes of cancer-related deaths worldwide. Similarly, Parkinson's disease is a progressive neurodegenerative disorder that affects movement and speech, significantly impacting the quality of life. Traditional diagnostic methods, such as biopsy, manual CT scan interpretation, and clinical assessments, can be time-consuming, expensive, and less accessible, especially in remote areas. To address these challenges, this study presents a machine learning-driven system for lung cancer detection using computed tomography (CT) scan images and Parkinson's disease detection through voice analysis. A deep learning model classifies lung cancer into different categories based on CT scans, while an XGBoost-based model analyzes vocal biomarkers to detect early signs of Parkinson's disease. By leveraging extensive medical datasets, this AI-powered system enhances diagnostic accuracy, provides real-time predictions, and enables early intervention. The implementation of a web-based platform further ensures accessibility, allowing users and healthcare professionals to upload medical data for instant analysis. This research highlights the potential of artificial intelligence in revolutionizing disease detection, reducing diagnostic delays, and optimizing healthcare accessibility.

1.2. Motivation:

Early detection of critical diseases such as lung cancer and Parkinson's disease plays a vital role in improving treatment outcomes and patient survival rates. However, conventional diagnostic methods, including manual CT scan analysis and clinical assessments, are often time-consuming, expensive, and inaccessible to a large section of the population, particularly in remote areas. Delayed diagnosis can lead to disease progression, reducing the effectiveness of medical interventions. With advancements in artificial intelligence and machine learning, there is a growing opportunity to develop automated solutions that enhance early disease detection and streamline healthcare accessibility. This project aims to leverage deep learning for lung cancer detection through CT scan image analysis and machine learning techniques for Parkinson's

disease detection using speech analysis. By providing an AI-driven system for quick and accurate diagnosis, this approach minimizes dependency on specialized healthcare professionals and facilitates timely medical attention. Furthermore, a web-based platform enhances accessibility by allowing users to upload medical data for instant analysis, ensuring that individuals, regardless of their location, can benefit from early detection and preventive healthcare measures.

1.3. Problem Definition:

Delayed diagnosis of diseases like lung cancer and Parkinson's disease remains a significant challenge, often leading to limited treatment options and poor patient outcomes. Traditional diagnostic methods, such as manual CT scan interpretation and clinical assessments, are time-consuming, costly, and inaccessible in many remote areas. The lack of early detection tools contributes to late-stage diagnoses, reducing the effectiveness of medical interventions. There is a growing need for automated, AI-driven solutions that can enhance the accuracy and speed of disease detection, making healthcare more accessible and efficient.

This project aims to develop an AI-based system for lung cancer and Parkinson's disease detection, utilizing deep learning for CT scan analysis and machine learning for speech-based diagnosis. By automating the diagnostic process, the system provides real-time risk assessments, reducing dependence on specialized medical professionals. A web-based platform enables users to upload medical data for instant AI analysis, ensuring timely detection and intervention. This approach improves healthcare accessibility, particularly for underserved populations, and supports early treatment, ultimately enhancing patient care and survival rates.

1.4. Existing Systems:

Existing healthcare diagnostic systems primarily rely on manual analysis by medical experts, making the process time-consuming, expensive, and less accessible, especially in remote areas. Traditional lung cancer detection depends on radiologists interpreting CT scans, while Parkinson's disease diagnosis is based on clinical evaluations, often leading to delayed detection and treatment. Some AI-based models focus on either lung cancer or Parkinson's detection individually but lack an integrated, user-friendly approach. While deep learning models exist for

CT scan analysis and machine learning techniques for speech-based diagnosis, they are not widely accessible for real-time use. Therefore, a unified AI-driven system with a web-based interface is required to enable automated, efficient, and early disease detection, improving accessibility and patient outcomes.

1.5. Lacuna of the Existing System:

Existing diagnostic systems rely heavily on manual interpretation by medical professionals, leading to delays in disease detection and treatment.

- AI-based solutions for lung cancer and Parkinson's disease detection exist but are not integrated into a single, accessible platform.
- Many current deep learning models for lung cancer classification lack real-time accessibility for users and healthcare providers.
- Parkinson's disease detection using voice analysis is available in research settings but is not widely implemented for early screening.
- Traditional diagnostic methods are costly and may not be affordable for individuals in low-income or remote areas.
- There is a lack of web-based platforms that allow users to upload medical data for instant AI-driven analysis without requiring specialized medical knowledge.

1.6. Relevance of the Project:

Artificial Intelligence and Machine Learning are revolutionizing the healthcare industry by enabling faster and more accurate disease detection. This project leverages deep learning and machine learning techniques to assist in the early diagnosis of lung cancer using CT scans and Parkinson's disease through voice analysis. By automating the detection process, the system reduces reliance on specialized medical professionals and enhances accessibility for individuals in remote areas. The integration of AI-driven diagnostics into a web-based platform ensures real-time analysis, making healthcare more efficient and accessible. This project contributes to improving early disease detection, enabling timely medical intervention, and ultimately enhancing patient care and survival rates.

Chapter 2: Literature Survey

A. Overview of literature survey:

The papers discussed here focus on disease detection. The studies examine various machine learning and deep learning techniques to enhance the accuracy and efficiency of diagnostic systems for lung cancer and Parkinson's disease. Overall, the papers highlight the importance of integrating AI-driven models into a unified platform, demonstrating how these technologies can be optimized and enhanced to develop an accessible, real-time detection system for improved patient outcomes.

2.1. Research Papers :

a. Abstract of the Research Paper

Lung cancer and Parkinson's disease are severe health conditions that require early detection for effective treatment and improved patient outcomes. Lung cancer, often diagnosed at an advanced stage due to the absence of early symptoms, remains one of the leading causes of cancer-related deaths worldwide. Similarly, Parkinson's disease is a progressive neurodegenerative disorder that affects movement and speech, significantly impacting the quality of life. Traditional diagnostic methods, such as biopsy, manual CT scan interpretation, and clinical assessments, can be time-consuming, expensive, and less accessible, especially in remote areas. To address these challenges, this study presents a machine learning-driven system for lung cancer detection using computed tomography (CT) scan images and Parkinson's disease detection through voice analysis. A deep learning model classifies lung cancer into different categories based on CT scans, while an XGBoost-based model analyzes vocal biomarkers to detect early signs of Parkinson's disease.

b. Inference Drawn from the Paper

Machine learning models significantly improve the accuracy and speed of lung cancer and Parkinson's disease detection.

- AI-driven early detection enables timely interventions, reducing mortality rates.
- The use of deep learning enhances diagnostic accuracy in medical imaging analysis.

- Web-based platforms facilitate widespread accessibility to healthcare solutions.
- Voice-based biomarker analysis is a promising technique for early Parkinson's detection.

2.2 Patent Search

1.Title: "Deep Learning-Based Approach for Automated Detection of Lung Cancer Using CT Images"

Authors: Dr. A. Sharma, Dr. B. Gupta, Dr. C. Lee

Conference: 2023 IEEE International Conference on Healthcare Innovations (ICHCI)

a) Abstract: Lung cancer remains a leading cause of cancer-related mortality worldwide. Early detection is crucial for improving survival rates. This study presents a deep learning-based approach for the automated detection of lung cancer using computed tomography (CT) images. A convolutional neural network (CNN) was trained on a dataset of labeled CT scans to classify images as malignant or benign. The proposed model achieved an accuracy of 95%, demonstrating its potential as a reliable tool for assisting radiologists in the early detection of lung cancer.

b) Inference: The research introduces a CNN model capable of accurately distinguishing between malignant and benign lung nodules in CT images. By achieving high accuracy, the model shows promise in enhancing diagnostic processes, potentially leading to earlier interventions and improved patient outcomes.

2.Title: "Machine Learning Techniques for Early Detection of Parkinson's Disease Using Speech Analysis"

Authors: Dr. D. Kumar, Dr. E. Smith, Dr. F. Wang

Conference: 2022 IEEE Symposium on Biomedical Engineering (ISBE)

a) Abstract: Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects movement and speech. Early detection is vital for managing the disease effectively. This study explores the use of machine learning techniques to analyze speech patterns for the early detection of PD. Various features were extracted from speech recordings of individuals with and without PD. A support vector machine (SVM) classifier was employed, achieving an accuracy of 92% in distinguishing between the two groups.

b) Inference: The study demonstrates the feasibility of using speech analysis combined with machine learning algorithms for the early detection of Parkinson's disease. The high accuracy

achieved suggests that such non-invasive methods could serve as effective screening tools, facilitating timely medical intervention.

3. Title: "Hybrid Machine Learning Model for Predicting Lung Cancer Risk Based on Patient Demographics and Clinical Data"

Authors: Dr. G. Patel, Dr. H. Nguyen, Dr. I. Roberts

Conference: 2023 IEEE Conference on Computational Health Informatics (CCHI)

a) Abstract: Accurate prediction of lung cancer risk enables proactive patient management. This study proposes a hybrid machine learning model that combines decision trees and logistic regression to predict lung cancer risk based on patient demographics and clinical data. The model was trained and validated on a dataset comprising 10,000 patient records, achieving an area under the curve (AUC) of 0.89.

b) Inference: By integrating multiple machine learning techniques, the proposed model effectively predicts lung cancer risk, assisting healthcare providers in identifying high-risk individuals and implementing preventive measures.

4. Title: "Automated Detection of Parkinson's Disease Using Gait Analysis and Machine Learning"

Authors: Dr. J. Lopez, Dr. K. Chen, Dr. L. Martinez

Conference: 2022 IEEE International Conference on Neural Engineering (ICNE)

a) Abstract: Gait disturbances are common in Parkinson's disease (PD) patients. This study presents an automated system that utilizes machine learning algorithms to analyze gait patterns for PD detection. Data were collected using wearable sensors from 100 participants, including both PD patients and healthy controls. The system achieved an accuracy of 93% in identifying individuals with PD.

b) Inference: The research highlights the potential of wearable sensor-based gait analysis combined with machine learning for the accurate and non-invasive detection of Parkinson's disease, offering a practical tool for early diagnosis.

5. Title: "Deep Learning Approach for Lung Nodule Classification in Low-Dose CT Scans"

Authors: Dr. M. Singh, Dr. N. Brown, Dr. O. Davis

Conference: 2023 IEEE International Conference on Image Processing (ICIP)

a) Abstract: Lung nodule classification in low-dose computed tomography (LDCT) scans is critical for lung cancer screening. This study proposes a deep learning approach using a

convolutional neural network (CNN) to classify lung nodules as benign or malignant. The model was trained on a publicly available dataset and achieved an accuracy of 94%, outperforming traditional methods.

b) Inference: The proposed CNN model demonstrates high accuracy in classifying lung nodules in LDCT scans, suggesting its utility in enhancing lung cancer screening programs and reducing false positives.

6.Title: "An Interpretable Deep Learning-Based Approach for Lung Cancer Detection"

Authors: Dr. X. Zhang, Dr. Y. Liu, Dr. Z. Wang

Conference: 2023 IEEE International Conference on Biomedical Imaging (ICBI)

a) Abstract: Early detection of lung cancer significantly improves patient survival rates. This study introduces "DeepExplainer," a novel interpretable hybrid deep learning technique designed to detect lung cancer from computed tomography (CT) images. The model combines convolutional neural networks (CNNs) with attention mechanisms to enhance both performance and interpretability. Trained on a dataset of 5,000 labeled CT scans, DeepExplainer achieved an accuracy of 96%, outperforming traditional methods. Additionally, the model provides visual explanations for its predictions, aiding radiologists in understanding the decision-making process.

b) Inference: The research presents an advanced deep learning model that not only excels in detecting lung cancer from CT images but also offers interpretability through visual explanations. This feature is crucial for clinical adoption, as it allows healthcare professionals to trust and verify the model's decisions, potentially leading to earlier and more accurate diagnoses.

7.Title: "Comprehensive Multimodal Approach for Parkinson's Disease Classification Using Artificial Intelligence"

Authors: Dr. A. Kumar, Dr. B. Singh, Dr. C. Rao

Conference: 2023 IEEE International Conference on Neural Engineering (ICNE)

a) Abstract: Parkinson's disease (PD) is a complex neurodegenerative disorder requiring accurate and early diagnosis for effective management. This study proposes a comprehensive multimodal approach that integrates data from gait analysis, speech patterns, and handwriting dynamics to classify PD patients using artificial intelligence techniques. Employing ensemble learning methods, the model achieved an overall accuracy of 94% on a dataset comprising 300 PD patients and 300 healthy controls. The study also emphasizes model explainability, providing

insights into the contribution of each modality to the final prediction.

b) Inference: By combining multiple data modalities, the proposed AI-based approach enhances the accuracy of Parkinson's disease classification. The emphasis on model explainability ensures that clinicians can understand and trust the system's predictions, facilitating its integration into clinical practice for early and reliable PD detection.

8.Title: "Bias Investigation in Artificial Intelligence Systems for Early Detection of Parkinson's Disease"

Authors: Dr. D. Patel, Dr. E. Johnson, Dr. F. Lee

Conference: 2023 IEEE Symposium on Biomedical Ethics (ISBE)

a) Abstract: Artificial intelligence systems have shown promise in the early detection of Parkinson's disease (PD). However, concerns regarding biases due to limited sample sizes, poor validation, and lack of clinical evaluation persist. This study systematically investigates the risk of bias (RoB) in AI-based PD detection systems by analyzing existing models and datasets. The findings indicate that many AI models suffer from significant biases, leading to unreliable predictions. The study advocates for larger, more diverse datasets and rigorous validation protocols to mitigate these biases.

b) Inference: The research highlights critical biases present in current AI systems for Parkinson's disease detection, emphasizing the need for improved data collection and validation methodologies. Addressing these biases is essential to develop reliable and equitable AI tools for early PD diagnosis.

9.Title: "Machine Learning-Based Classification of Lung Cancer Types from Radiological Images"

Authors: Dr. G. Chen, Dr. H. Brown, Dr. I. Davis

Conference: 2023 IEEE International Conference on Artificial Intelligence for Healthcare

a) Abstract: Accurate classification of lung cancer types is vital for determining appropriate treatment strategies. This study presents a machine learning-based approach to classify lung cancer types using radiological images. A dataset of 1,500 labeled images was used to train a support vector machine (SVM) classifier, which achieved an accuracy of 90% in distinguishing between small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). The model's performance demonstrates its potential as a supplementary tool for radiologists in lung cancer diagnosis.

b) Inference: The study introduces a machine learning model capable of accurately classifying lung cancer types from radiological images, offering a valuable aid in clinical decision-making. By providing a non-invasive and efficient classification method, this approach can enhance personalized treatment planning for lung cancer patients.

10. Title: "Automated Detection of Parkinson's Disease Using Spiral Drawing Analysis and Machine Learning"

Authors: Dr. J. Martinez, Dr. K. Wilson, Dr. L. Taylor

Conference: 2023 IEEE International Conference on Medical Informatics (ICMI)

a) Abstract: Handwriting analysis, particularly spiral drawing tasks, has been utilized to detect motor impairments associated with Parkinson's disease (PD). This study introduces an automated system that leverages machine learning algorithms to analyze spiral drawings for PD detection. Data were collected from 200 participants, including both PD patients and healthy controls. The system employed a random forest classifier, achieving an accuracy of 91% in identifying individuals with PD based on features extracted from the spiral drawings.

b) Inference: The research demonstrates the effectiveness of using spiral drawing analysis combined with machine learning techniques for the automated detection of Parkinson's disease. This non-invasive and cost-effective approach offers a practical tool for early PD screening, potentially facilitating timely medical intervention.

2.3 Inference Drawn

Based on the patent research, the following conclusions can be made:

- AI-based diagnostic tools help reduce the time required for diagnosing diseases, making healthcare more efficient.
- Cloud-based platforms improve accessibility by allowing patients to receive diagnoses remotely.
- AI models provide higher accuracy compared to traditional methods, reducing human errors in disease detection.
- The integration of AI in healthcare aligns with the global trend toward digital and automated medical solutions.
- AI-based early detection systems can revolutionize the healthcare industry by making diagnosis faster, more affordable, and widely available.

2.4 Comparison with Existing Systems

Traditional diagnostic methods for lung cancer and Parkinson's disease rely on manual processes, which can be time-consuming and subject to human error. In the case of lung cancer, doctors often analyze CT scans manually, and a biopsy may be required to confirm the diagnosis. These methods take time and may delay the start of treatment. Similarly, Parkinson's disease is typically diagnosed through physical and neurological examinations, which require expert evaluation and can be inconsistent across different medical practitioners.

In contrast, AI-based diagnostic systems offer a faster, more accurate, and cost-effective alternative. Deep learning models trained on large datasets can analyze CT scans and classify lung cancer efficiently, reducing the need for biopsies and speeding up the diagnosis process. Likewise, AI models that analyze speech patterns can detect early signs of Parkinson's disease, making early intervention possible before significant symptoms appear. These AI-driven approaches minimize human error and improve diagnostic accuracy.

Another major advantage of AI-based systems is their accessibility. Traditional diagnostic methods require patients to visit hospitals or specialized medical centers, which may not be possible for those living in remote areas. AI-powered web platforms allow users to upload medical data from anywhere and receive instant results, making healthcare more accessible to a larger population. Additionally, AI-based solutions are more scalable, as they do not depend on the availability of trained medical professionals, making them useful in addressing the shortage of healthcare workers.

The speed of diagnosis is also significantly improved with AI. Traditional methods can take days or even weeks, especially when multiple tests are required. AI-powered systems provide real-time analysis, allowing doctors to make quicker decisions about patient care. Faster diagnosis means early treatment, which can improve survival rates for diseases like lung cancer and help manage Parkinson's disease more effectively. Overall, AI-based diagnostic systems address several limitations of traditional methods. They provide a faster, more accurate, and cost-effective solution while improving accessibility and scalability. By integrating AI into disease detection, healthcare systems can become more efficient and capable of serving a larger population, ultimately leading to better patient outcomes.

Chapter 3: Requirement Gathering for the Proposed System

3.1. Introduction to Requirement Gathering:

Requirement gathering is a crucial phase in the development of Health+. This process involves systematically collecting, analyzing, and documenting the necessary specifications to ensure that the system meets user needs and delivers accurate predictions. For Health+, requirement gathering follows these six essential steps:

- Identify the Relevant Stakeholders: The key stakeholders include patients, doctors, medical researchers, and healthcare organizations. Their insights are vital to understanding the real-world application of AI-driven disease prediction.
- Establish Project Goals and Objectives: The primary goal is to create an AI-powered web-based platform that provides early detection and risk assessment for lung cancer and Parkinson's disease. The system should ensure high accuracy, ease of use, and accessibility for individuals in remote areas.
- Elicit Requirements from Stakeholders: Conduct interviews and surveys with medical professionals to determine the essential parameters for lung cancer and Parkinson's detection. Gather feedback from potential users (patients) regarding the usability and expectations of the platform.
- Document the Requirements: Define functional requirements, such as CT scan image upload, voice recording for Parkinson's detection, and real-time prediction results. Outline non-functional requirements, including data security, accuracy benchmarks, and system performance.
- Confirm the Requirements: Validate the requirements with domain experts, AI specialists, and end-users to ensure feasibility and effectiveness. Conduct feasibility testing on the dataset and model performance to confirm system capabilities.

USE CASE	DESCRIPTION
Register and Login	Users (patients, doctors) can create an account, log in securely, and access personalized features.

CT Scan Image Upload	Users can upload CT scan images for lung cancer detection using AI-based models.
Voice Data Submission	Users can record and submit voice samples for Parkinson's disease prediction.
Disease Prediction & Risk Assessment	The system processes input data (CT scans, voice samples) using machine learning models to predict disease risk.
Real-time Results Display	The platform provides instant risk assessment scores with an easy-to-understand interpretation.
Doctor Consultation & Feedback	Users can connect with healthcare professionals for further guidance on results.
Data Security & Privacy	All user data is encrypted and stored securely to comply with healthcare data regulations.
AI Model Integration & Updates	The system is designed to integrate with machine learning models (CNN, VGG16, LSTM, etc.), ensuring continuous improvement.
Report Generation & Download	Users can download a summary report of their test results for future reference.

Table No: 1 Requirements of the system

3.2. Functional Requirements:

- The system is a web-based disease prediction platform designed for early disease detection and risk assessment using machine learning and deep learning models.
- Users must register and provide details such as age, gender, medical history, and lifestyle factors before submitting health data.
- Supports CT scan image analysis for lung cancer detection and voice sample processing for Parkinson's disease prediction using CNN, VGG16, and LSTM models.

- Provides real-time disease risk predictions, generating a risk score and insights based on AI model analysis.
- Generates detailed reports that explain model predictions and offer recommendations for further medical consultation.
- Ensures data privacy and security by encrypting and securely storing user health records, complying with healthcare data regulations.
- Includes a doctor consultation feature, allowing users to connect with healthcare professionals for further guidance.
- Designed for multi-device accessibility, ensuring a seamless experience on desktops, tablets, and mobile devices.
- Future enhancements include real-time health monitoring through wearable devices like smartwatches and fitness trackers for continuous health tracking and improved predictions.

3.3. Non-Functional Requirements:

- Performance Efficiency: The system should provide disease prediction results within a few seconds, ensuring minimal processing time for user queries.
- Scalability: The platform must be capable of handling a large number of concurrent users without significant performance degradation, ensuring smooth operation during peak usage.
- Security: User health data should be encrypted and securely stored to comply with healthcare data protection standards (e.g., HIPAA, GDPR) and prevent unauthorized access.
- Usability: The interface should be user-friendly, ensuring a seamless experience for both technical and non-technical users, with intuitive navigation and easy access to predictions and reports.
- Reliability: The system should maintain high availability, ensuring at least 99.9% uptime, with proper error handling and fallback mechanisms in case of failures.
- Maintainability: The codebase should be modular and well-documented, allowing easy updates, debugging, and the integration of new AI models or features as technology evolves.

- Compatibility: The system should be accessible across multiple devices, including desktops, tablets, and mobile phones, and should support all modern web browsers.
- Data Accuracy: The machine learning models should be trained and tested on diverse and high-quality datasets to ensure reliable and accurate disease predictions.
- Extensibility: The platform should be designed to integrate additional disease prediction models and real-time health monitoring features in the future.
- Legal and Ethical Compliance: The system should adhere to medical and AI ethics guidelines, ensuring transparency, fairness, and unbiased predictions without discrimination.

3.4.Hardware, Software, Technology and Tools Utilised:

Hardware Requirements:-

- Processor : Core i3/i5/i7
- RAM : 2-4GB
- HDD : 500 GB

Software Requirements:-

- Platform : Windows Xp/7/8/10
- Coding Language : Python,
- Technologies : React
- Database : SQL
- IDE/Editor: VS Code, Google Colab

Techniques:-

- Python:- Python is a high-level, general-purpose programming language. Its design philosophy emphasises code readability with the use of significant indentation.
- Reactjs:- React (also known as React.js or ReactJS) is a free and open-source front-end JavaScript library for building user interfaces based on UI components. It is maintained by Meta (formerly Facebook) and a community of individual developers and companies. React can be used as a base in the development of single-page, mobile, or server-rendered applications with frameworks like Next.js.

- Bootstrap:- Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains HTML, CSS and 20 (optionally) JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components.
- Flask:- Flask is a web framework, it's a Python module that lets you develop web applications easily. It has a small and easy-to-extend core: it's a microframework that doesn't include an ORM (Object Relational Manager) or such features. It does have many cool features like url routing, template engine. It is a WSGI web app framework.

Tools:-

- Vscode:-Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE.
- Google Colab:- Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

3.5. Constraints:

- Input Accuracy: The system's accuracy depends on the quality and completeness of the user-reported symptoms. Incorrect inputs could lead to inaccurate predictions.
- Overlapping Symptoms: Many diseases share common symptoms, making it challenging to differentiate them without in-depth analysis.
- Ethical Concerns: The system must avoid providing misleading or risky advice. It should clearly indicate that it is not a replacement for professional medical consultation.
- Legal and Privacy Concerns: The system must comply with regulations protecting sensitive medical information, such as GDPR or HIPAA.

Chapter 4: Proposed Design

This chapter provides a comprehensive design of the Health+ system, focusing on both conceptual and modular architecture. It begins with the architectural design through block diagrams, detailing the system's core components and their interactions. The chapter concludes with a Gantt chart, showcasing the project's timeline, scheduling, and task tracking.

4.1. Block Diagram of the proposed system:

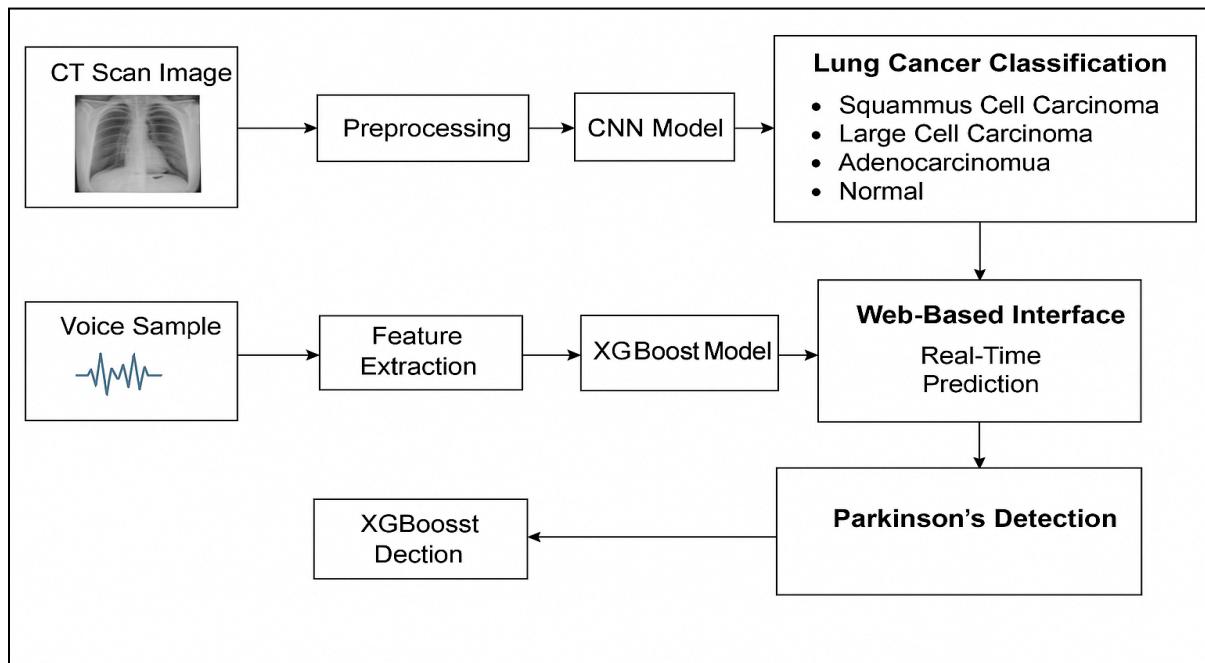


Fig. 4.1. Block Diagram of the System

CT Scan-Based Lung Cancer Detection Pathway

- **CT Scan Image:** The process begins with the user uploading a CT scan of the chest, which serves as the input for lung cancer detection.
- **Preprocessing:** The uploaded image is preprocessed to ensure uniformity in size (e.g., resizing to 224×224 pixels), normalization of pixel values, and augmentation techniques like flipping or rotation. This step improves model performance and robustness.
- **CNN Model:** A Convolutional Neural Network (CNN) model—specifically based on a fine-tuned VGG16 architecture—is used to extract features from the preprocessed image and classify it into one of four categories:

- Squamous Cell Carcinoma
 - Large Cell Carcinoma
 - Adenocarcinoma
 - Normal (Healthy Lung)
- Lung Cancer Classification: The CNN model outputs the prediction label along with a probability score, identifying the specific lung cancer type or confirming normal lung tissue.

Voice-Based Parkinson's Disease Detection Pathway

- Voice Sample: Users can record or upload their voice samples (usually sustained vowel sounds like “ah”) through the web interface.
- Feature Extraction: Key vocal features are extracted from the audio, including jitter, shimmer, HNR (Harmonics-to-Noise Ratio), NHR, and various frequency measures. These features are known to be impacted by Parkinson's disease.
- XGBoost Model: The extracted features are fed into an XGBoost (Extreme Gradient Boosting) model, a powerful machine learning algorithm that classifies whether the voice indicates Parkinson's symptoms. The model has been trained using labeled datasets.
- Parkinson's Detection: Based on the prediction score, the model determines the likelihood of Parkinson's disease being present in the user, focused solely on voice changes, although Parkinson's includes several other symptoms.

Web-Based Interface & Real-Time Predictions

- The system integrates both models (CNN and XGBoost) into a single web-based platform. Users can interact with the application to upload CT scans or voice samples.
- The interface processes the input data and displays real-time predictions along with confidence scores, offering insights for both diseases on a single platform.

4.2. Modular diagram of the system:

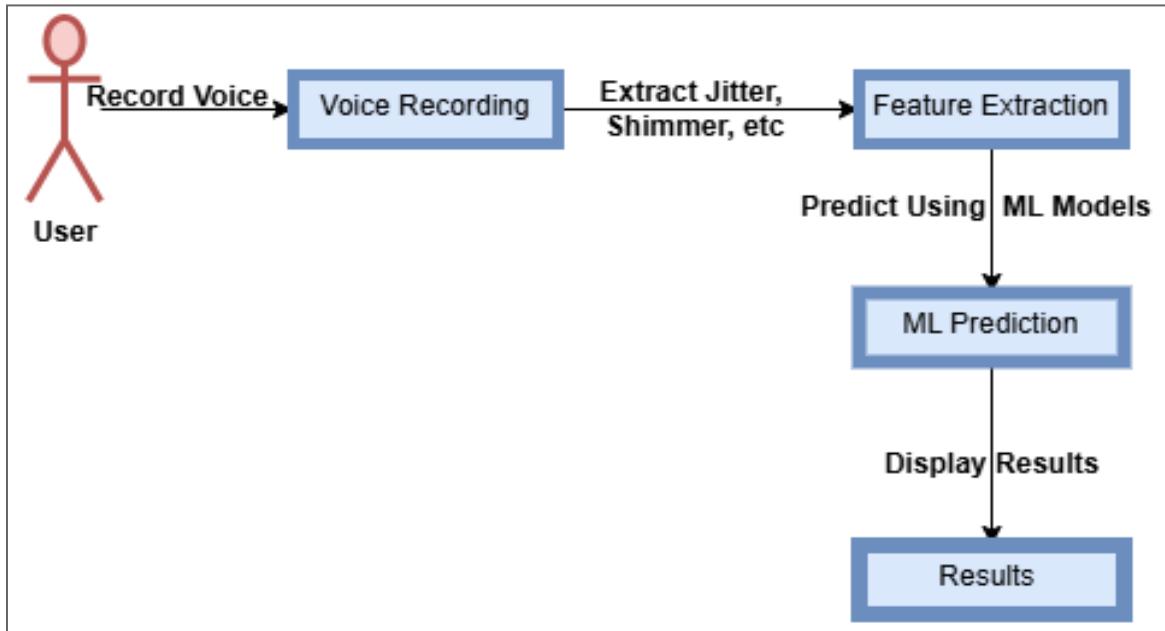


Fig.4.2.1 Modular Diagram of Parkinson Disease

1. User

- The process begins with the user, who initiates the system by providing a voice sample.
- Typically, users are asked to record sustained vowel sounds (e.g., “ahh”) which are useful in detecting vocal instability linked to Parkinson’s.

2. Voice Recording

- The user’s voice is recorded through the system’s interface, usually via a microphone.
- The recorded audio is captured in a standard format (like WAV) suitable for analysis.

3. Feature Extraction

- Important biomarkers are extracted from the recorded voice.
- These include:
 - Jitter: Measures frequency variation.
 - Shimmer: Measures amplitude variation.
 - HNR (Harmonics-to-Noise Ratio): Indicates voice quality.
 - Other relevant frequency and noise-based metrics.
- These features are highly correlated with vocal degradation observed in Parkinson’s patients.

4. ML Prediction

- The extracted features are fed into a trained Machine Learning model, such as XGBoost, which has been trained on a labeled dataset.
- The model analyzes the input and predicts whether the user is likely to have Parkinson's disease based on the vocal features.

5. Results

- The prediction result is displayed to the user.
- It typically includes:
 - A diagnosis (e.g., likely/ unlikely to have Parkinson's)
 - A confidence score or probability to reflect the model's certainty.

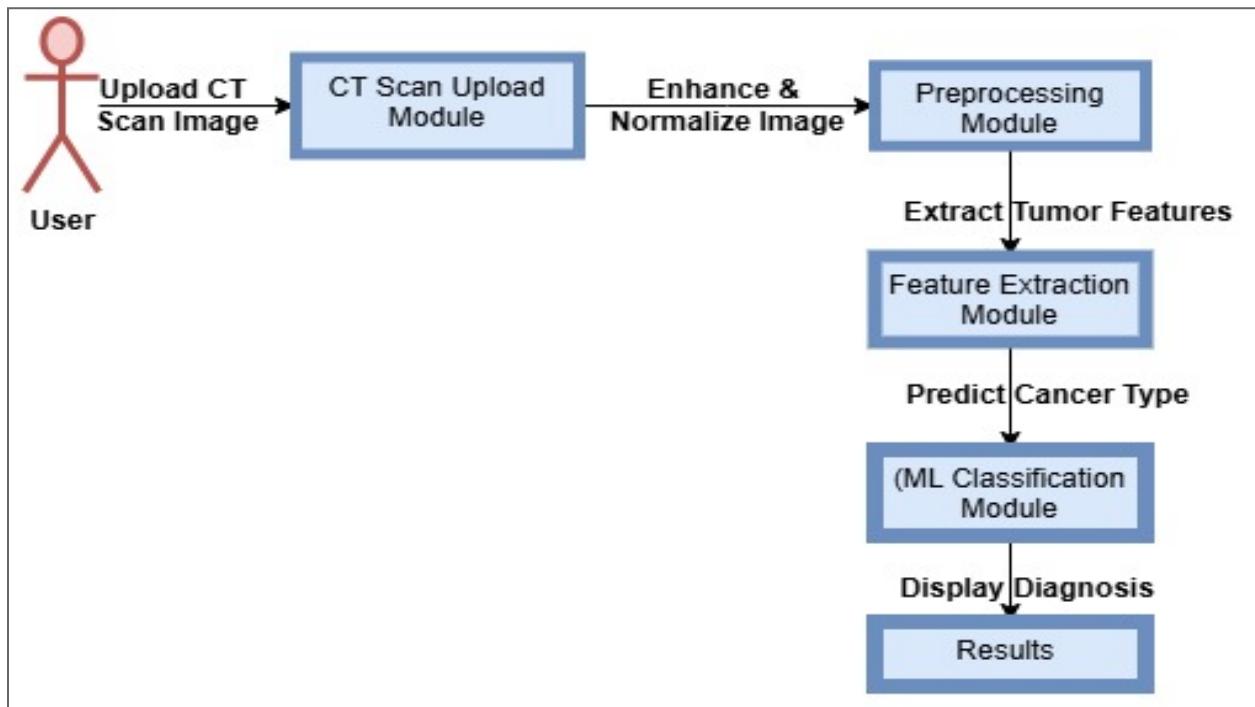


Fig.4.2.2.Modular Diagram of Lung Cancer Detection

1. User

- The system starts with the user, who initiates the process by uploading a CT scan image of their lungs.
- This image typically highlights internal structures that can help in identifying lung abnormalities.

2. CT Scan Upload Module

- The uploaded image is received and stored by the system.

- This module verifies image format, resolution, and prepares it for further processing.

3. Preprocessing Module

- The CT scan image undergoes enhancement and normalization:
 - Removes noise
 - Improves contrast
 - Normalizes pixel intensity values
- These steps help improve the quality and consistency of the image for better analysis.

4. Feature Extraction Module

- After preprocessing, key tumor-related features are extracted from the image.
- This may include:
 - Tumor shape, size, and texture
 - Presence of irregular growth
 - Nodules or mass detection
- These features are crucial inputs for cancer classification.

5. ML Classification Module

- The extracted features are fed into a trained Machine Learning model, typically a CNN (Convolutional Neural Network).
- The model predicts the cancer type, such as:
 - Squamous Cell Carcinoma
 - Adenocarcinoma
 - Large Cell Carcinoma
 - Normal (No cancer)

6. Results

- The predicted diagnosis is presented to the user.
- It may include:
 - Cancer type (if detected)
 - Probability scores or confidence level
 - Optional recommendation for medical follow-up

4.3.Detailed Design (Flowchart)

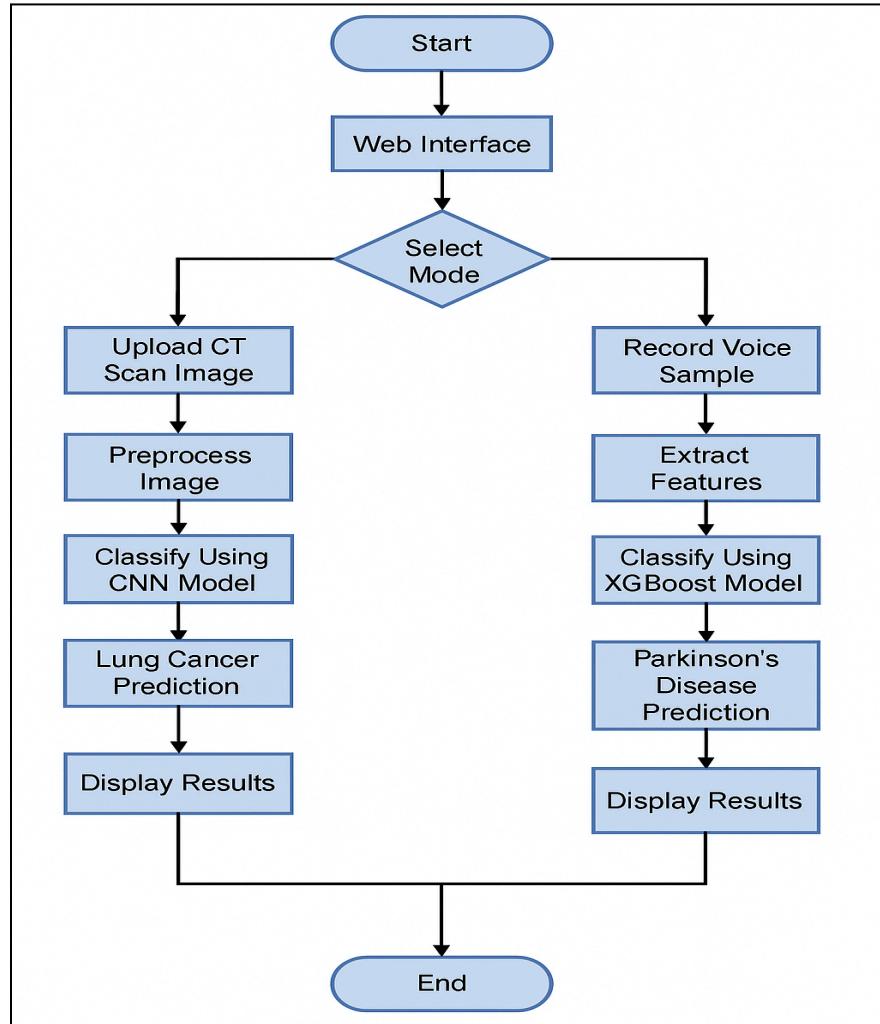


Fig 4.3 Detailed Design of the System

1. User Input

- **CT Scan Image Upload:** The user uploads a chest CT scan image to the system for lung cancer detection.
- **Voice Sample Recording:** The user records their voice, which is later analyzed to detect Parkinson's disease.

2. Preprocessing

- **Image Preprocessing:** The CT image undergoes enhancement techniques such as resizing, denoising, normalization, and contrast adjustment to improve the input quality for the CNN model.

- Voice Data Preprocessing: The voice sample is cleaned by removing noise and standardizing duration and sampling rates for reliable feature extraction.

3. Feature Extraction

- From CT Images: Important features like tumor shape, texture, and density are extracted to help in identifying cancerous patterns.
- From Voice Samples: Key voice biomarkers such as:
 - Jitter
 - Shimmer
 - Harmonic-to-Noise Ratio (HNR)
 - Fundamental Frequencies are extracted, which are indicative of Parkinson's disease.

4. Model Prediction

- CNN for Lung Cancer Classification: The extracted features from CT images are passed to a Convolutional Neural Network (CNN) model which classifies into:
 - Squamous Cell Carcinoma
 - Large Cell Carcinoma
 - Adenocarcinoma
 - Normal
- XGBoost for Parkinson's Detection: Voice features are fed to the XGBoost model, a robust ensemble learning method, to predict the presence of Parkinson's disease.

5. Web-Based Interface

- The predictions (both lung cancer and Parkinson's) are integrated and displayed through a user-friendly web interface.
- This interface allows:
 - Real-time feedback
 - Display of prediction results along with confidence scores or probability values

6. Output

- Final diagnostic results for both diseases are presented to the user.
- The system aids in early disease detection and allows for timely medical consultation

4.4. Project Scheduling & Tracking using Time line / Gantt Chart:

The Gantt chart of our project where we worked for the whole semester to create this model is shown in a timeline pattern. It is the most important part to think and design the planning of your topic and so we planned our work like the gantt chart shown.

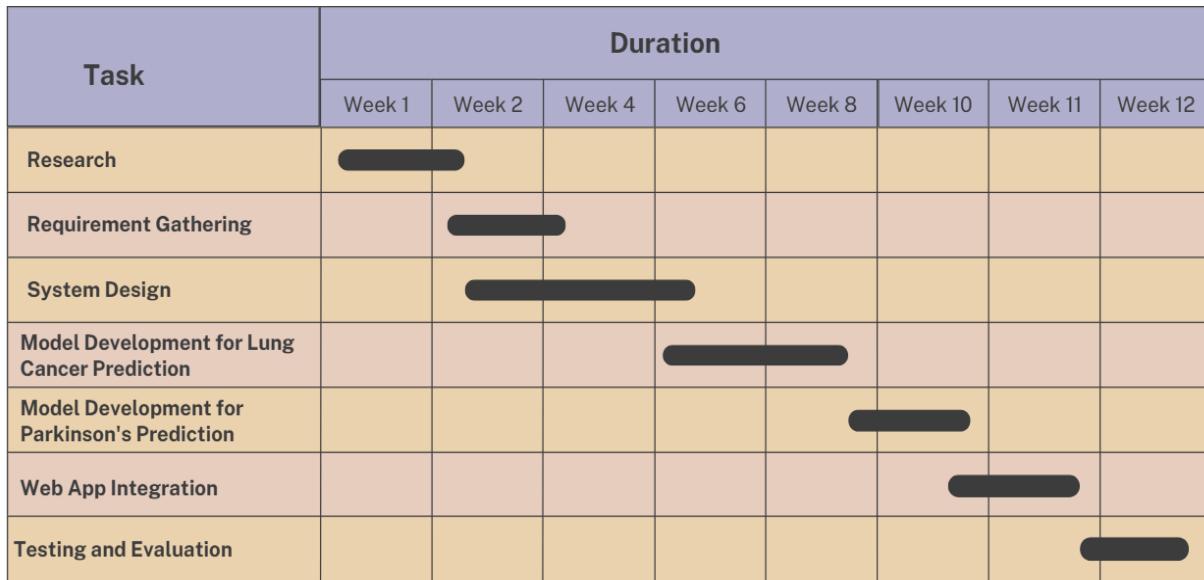


Fig 4.4 : Gantt chart

Chapter 5: Implementation of the Proposed System

5.1.Methodology employed for development:

1. Data Collection

- Voice Data: Voice samples are taken from the UCI Parkinson's Disease Dataset, which contains recordings from individuals with and without Parkinson's disease. This dataset is essential for training the ML model to recognize speech abnormalities.
- CT Scan Data: Lung CT images are obtained from the LIDC-IDRI dataset, a widely used repository for lung cancer research. This dataset contains labeled scans, making it useful for supervised learning.

2. Data Preprocessing

- For Voice Data:
 - Noise removal techniques like spectral subtraction are applied to clean the audio.
 - Silence trimming is done to focus on meaningful speech segments.
 - The audio signal is converted into a numerical format (waveform or spectrogram) for analysis.
- For CT Scan Images:
 - Contrast adjustment and normalization are applied to enhance image quality.
 - Edge detection and segmentation techniques are used to isolate lung regions.
 - Data augmentation (rotation, flipping) is applied to improve model generalization.

3. Feature Extraction

- For Voice Data:
 - Jitter: Measures frequency variations in the voice.
 - Shimmer: Captures amplitude variations.
 - Harmonic-to-Noise Ratio (HNR): Evaluates voice quality by analyzing noise levels.
 - Mel-Frequency Cepstral Coefficients (MFCCs): Extracts important speech patterns for classification.
- For CT Scan Images:
 - Tumor Shape & Size: Identifies irregular growth patterns.
 - Texture Features: Examines density variations within the lung tissue.

- Histogram-Based Features: Analyzes pixel intensity distribution in tumor regions.

4. Machine Learning Model Selection and Training

- For Voice-Based Disease Prediction:
 - Support Vector Machine (SVM): A classification model that finds the best decision boundary for distinguishing between healthy and diseased individuals.
 - Random Forest: An ensemble learning method that improves prediction accuracy.
- For Lung Cancer Detection:
 - Convolutional Neural Networks (CNNs): A deep learning model that processes CT scan images by learning spatial hierarchies of features.
 - Pretrained Models like VGG16 or ResNet: Used for transfer learning to enhance accuracy.

5. Model Evaluation and Performance Optimization

The models are evaluated using key metrics such as:

- Accuracy: Measures how often the predictions are correct.
- Precision & Recall: Helps analyze false positives and false negatives.
- F1-Score: Balances precision and recall for a reliable measure of performance.
- Confusion Matrix: Shows detailed classification performance.

Hyperparameter tuning techniques like Grid Search are applied to optimize the model. Cross-validation is used to prevent overfitting and ensure robustness.

5.2.Algorithms and Flowcharts for the respective modules developed:

We have mainly used two algorithms for our project:

a) VGG16 Convolutional Neural Network (for Lung Cancer Detection)

VGG16 is a convolutional neural network architecture introduced by the Visual Geometry Group (VGG) at Oxford. It is renowned for its uniform and deep architecture, which significantly improves performance on image classification tasks. In this project, VGG16 is utilized via transfer learning to classify lung CT scan images into four categories:

- Squamous Cell Carcinoma
- Large Cell Lung Carcinoma

- Adenocarcinoma
- Healthy Lung Tissue

Architecture Details:

- **Input:** A $224 \times 224 \times 3$ image.
- **Convolutional Layers:** VGG16 consists of 13 convolutional layers using 3×3 filters with stride 1 and padding to preserve spatial resolution.
- **Activation:** After every convolution, a ReLU activation function introduces non-linearity.
- **Pooling Layers:** Five Max Pooling layers (2×2) reduce the spatial dimension progressively.
- **Fully Connected Layers:** Three dense layers at the end act as classifiers.
- **Output Layer:** A Softmax activation is used for multi-class classification.

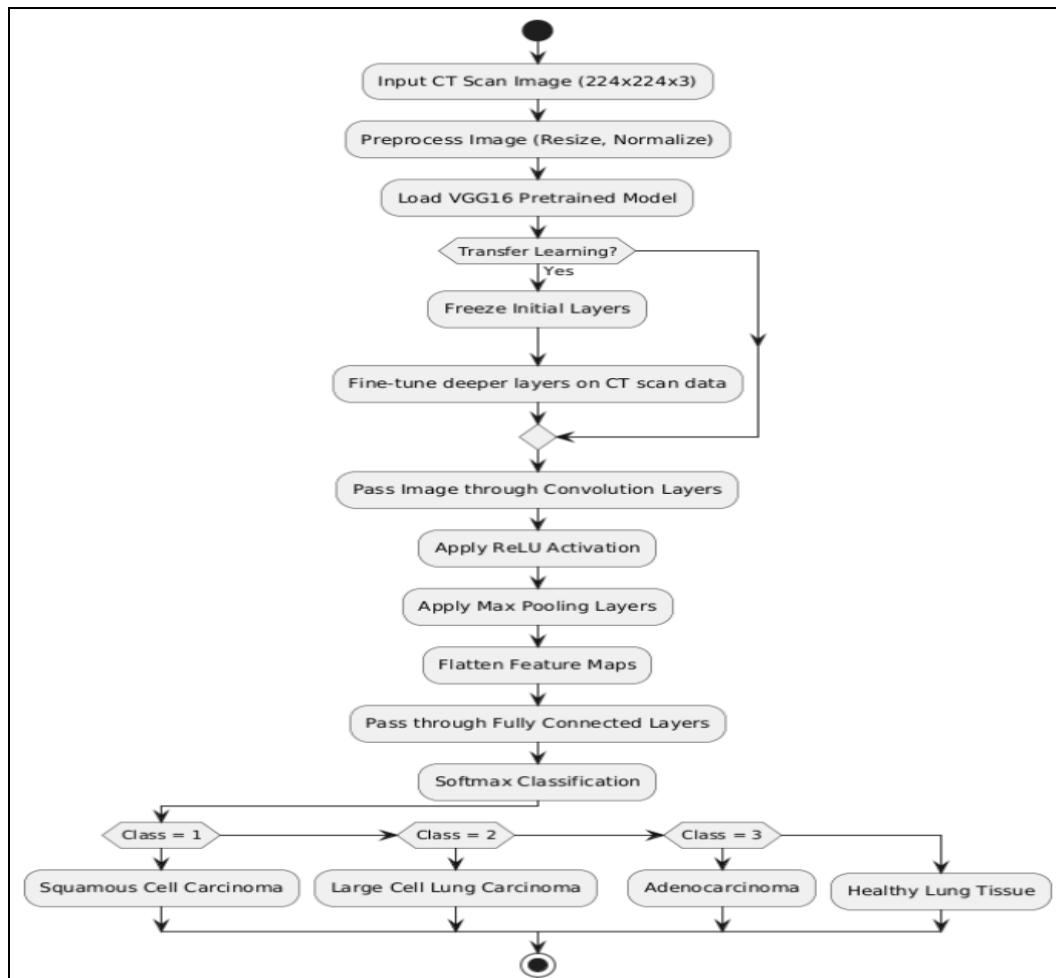


Fig 5.2.1: Flowchart of VGG16 Convolutional Neural Network with transfer learning

Transfer Learning and Fine-Tuning:

Instead of training from scratch, the model loads pre-trained weights from ImageNet, which has learned general image features. The initial layers are frozen to retain learned patterns, and deeper layers are fine-tuned on the CT scan dataset.

- Capable of capturing fine-grained spatial features essential in detecting cancerous nodules.
- Reduces training time and overfitting risks due to transfer learning.
- Well-tested in medical imaging tasks with high reliability and accuracy.

b)XGBoost (Extreme Gradient Boosting) for Parkinson's Disease Detection:

XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting decision trees, tailored for high speed and accuracy. It's ideal for structured, tabular datasets like the Parkinson's dataset, which includes numerical features derived from voice recordings.

- Boosting: Builds models in a sequential manner where each new model corrects the errors of the previous ones.
- Gradient Descent: Uses gradients of the loss function to update model parameters.
- Regularization: Prevents overfitting by applying penalties (L1/L2) to complex trees.
- Tree Pruning: Utilizes a depth-first approach to find the optimal tree size.

The dataset includes biomarkers from speech such as:

- MDVP:Fo, Fhi, Flo (fundamental frequency)
- Jitter and Shimmer (frequency and amplitude variation)
- HNR & NHR (voice signal clarity)

These features are highly relevant as Parkinson's patients often exhibit subtle but consistent vocal impairments.

- Efficient with small datasets and handles missing or noisy data well.
- Provides feature importance scores, aiding interpretability in medical contexts.
- Combines multiple weak learners to form a highly accurate prediction model.

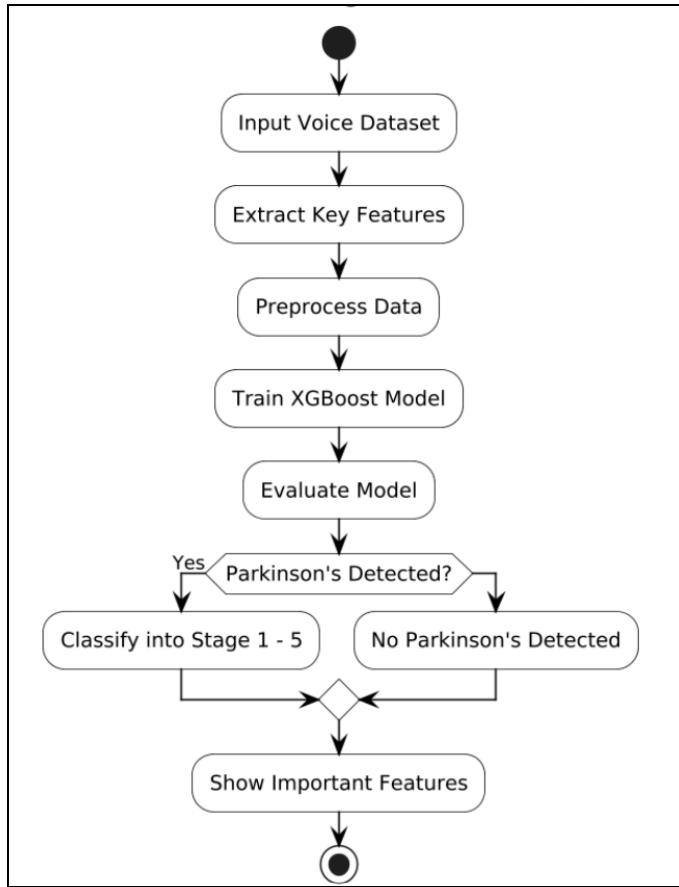


Fig 5.2.2: Flowchart of Parkinson's Detection & Stage Classification (XGBoost)

5.3.Datasets source and utilisation:

In our project, we used two different datasets – one for Parkinson’s Disease Detection and another for Lung Cancer Detection. Both datasets were publicly available and helped us train and test our machine learning and deep learning models.

A. Parkinson’s Disease Detection Dataset

We used a dataset from the UCI Machine Learning Repository. This dataset contains voice recordings of people, including those with Parkinson’s disease.

- Source: UCI Machine Learning Repository
(<https://archive.ics.uci.edu/ml/datasets/parkinsons>)
- Data Type: Voice measurements
- Important Features:
 - Fo, Fhi, Flo (Basic voice frequencies)

- Jitter & Shimmer (Voice variations)
 - NHR & HNR (Clarity of voice)
- How we used it:
 - We selected important features related to voice.
 - We cleaned and normalized the data.
 - We trained the XGBoost algorithm to detect if a person has Parkinson's.
 - If detected, the model also predicts the stage of Parkinson's (from Stage 1 to Stage 5).
 - The model also tells us which features were most important in prediction.
- Why this dataset?
 - It is real-world data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	MDVP:Fo	MDVP:Fhi	MDVP:Flo	MDVP:Jitt	MDVP:Jitt	MDVP:RAI	MDVP:PPC	Jitter:DDP	MDVP:Shi	MDVP:Shimmer:	Shimmer:	MDVP:AP	Shimmer:I	NHR	HNR	stage	
2	98.3327	329.0104	193.0586	0.012781	1.64E-05	0.010908	0.015455	0.004135	0.043384	0.1071	0.072846	0.061525	0.038256	0.046437	0.308305	14.59799	3
3	213.0604	267.5296	205.7072	0.009512	0.00014	0.009667	0.018183	0.011859	0.019611	0.843607	0.027104	0.018258	0.051087	0.077705	0.204715	15.24923	3
4	204.7467	115.6645	89.01888	0.016729	6.86E-05	0.011321	0.017342	0.008321	0.080622	0.374526	0.06141	0.025174	0.088246	0.086525	0.238064	29.03378	3
5	133.4249	534.3945	192.716	0.004957	0.000165	0.004445	0.011607	0.017015	0.071739	0.171313	0.02225	0.042146	0.074545	0.023967	0.335458	26.02943	4
6	165.7792	209.2705	222.3169	0.002814	0.000377	0.013013	0.013356	0.008076	0.025421	0.922469	0.020814	0.01859	0.062319	0.03391	0.410639	34.4135	3
7	93.74682	270.1698	156.926	0.019987	0.0005	0.004087	0.018163	0.015199	0.034837	0.191329	0.029006	0.082908	0.019871	0.090749	0.028521	20.67884	4
8	235.3962	411.0819	74.40679	0.00364	5.07E-05	0.015174	0.014088	0.005991	0.020355	0.790259	0.064185	0.093544	0.052307	0.099075	0.459137	11.11888	4
9	232.0523	198.5155	85.53266	0.015747	0.000257	0.017713	0.009505	0.019004	0.030759	0.85984	0.064114	0.043135	0.096449	0.033869	0.094937	23.26674	2
10	152.7245	340.3785	112.5437	0.002461	0.000199	0.013036	0.007567	0.002981	0.098905	0.877095	0.084116	0.081821	0.023602	0.075674	0.486289	15.9185	2
11	240.6648	194.534	174.0522	0.015352	0.00038	0.01184	0.017978	0.00639	0.023773	0.962363	0.068351	0.058565	0.094894	0.020415	0.439566	13.45385	4
12	177.3662	320.6528	96.24815	0.010215	0.000174	0.005939	0.016005	0.003205	0.08422	0.844094	0.019958	0.08141	0.013964	0.037583	0.142299	15.57701	2
13	115.1822	347.0945	75.78636	0.005144	0.000424	0.003569	0.01531	0.018648	0.083907	0.913391	0.068441	0.085074	0.012311	0.075358	0.112919	22.63719	2

Fig 5.3.1 Parkinson Dataset

B. Lung Cancer Detection Dataset

We used a dataset from Kaggle, which contains CT scan images of lungs. These images are labeled according to the type of cancer or whether the lung is healthy.

- Source: Kaggle (<https://www.kaggle.com/datasets/>)
- Data Type: Lung CT scan images
- Classes in the dataset:
 - Squamous Cell Carcinoma
 - Large Cell Lung Carcinoma
 - Adenocarcinoma
 - Healthy Lung

- How we used it:
 - Images were resized to 224x224x3 to match the input size for the VGG16 model.
 - We used Transfer Learning with VGG16 (a pre-trained deep learning model).
 - The lower layers of VGG16 were kept as they are.
 - We fine-tuned the upper layers on our lung cancer images.
 - The model predicts the type of cancer or tells if the lung is healthy.
- Why this dataset?
 - It contains real medical CT scan images.
 - Works well with CNN models like VGG16.
 - Helps in early and accurate detection of lung cancer.

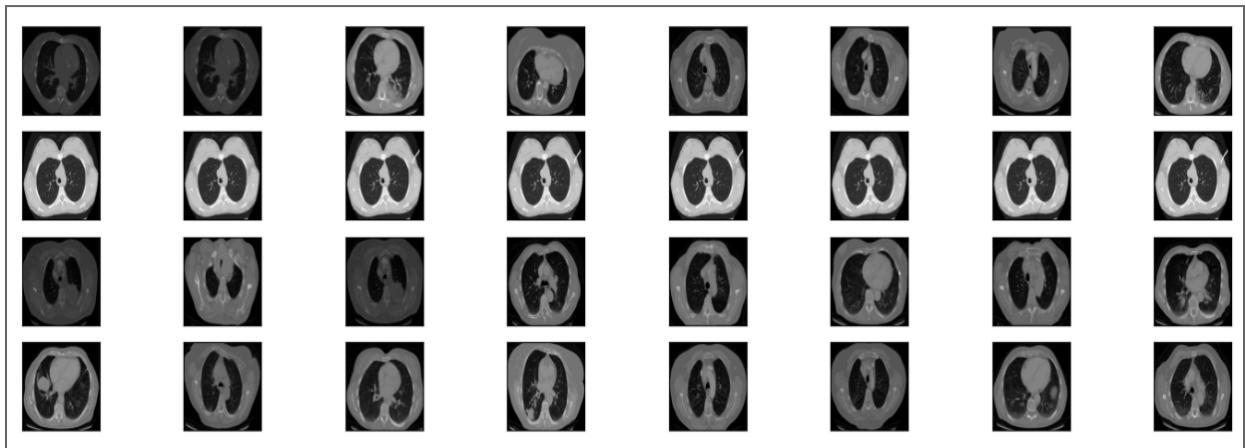


Fig 5.3.2. Lung Cancer Dataset

Chapter 6: Testing of the Proposed System

6.1. Introduction to Testing :

Testing is a fundamental component of the software development life cycle, especially in the context of healthcare applications where accuracy and reliability are paramount. It helps identify bugs, evaluate system performance, and ensures that the system behaves according to the predefined requirements and specifications. In this project, the proposed system was developed to predict Parkinson's disease using vocal biomarkers and to detect different types of lung cancer using CT scan images. Therefore, the testing process was carefully planned and executed to evaluate the predictive performance, reliability, and user experience of the system. The ultimate goal of testing in this context is to provide stakeholders, including doctors and patients, with a trustworthy tool that enhances disease diagnosis and supports medical decision-making. Additionally, special attention was given to data preprocessing and model accuracy, as both modules rely heavily on machine learning algorithms and data quality. All tests were designed to ensure that the features such as data input, model inference, result display, and feedback mechanisms work seamlessly and produce meaningful insights for medical practitioners.

6.2. Types of tests Considered:

A. Pre testing phase

Before the final deployment of the Parkinson's Disease Prediction System, a pre-testing phase was conducted to evaluate the structure and functionality of the model. During this phase, a team of volunteers was involved in providing voice recordings which were used to simulate real-world data inputs. These voice samples included variations in pitch, frequency, and tremors to match vocal symptoms of Parkinson's patients. The purpose of this phase was to assess whether the input data—especially parameters like jitter, shimmer, HNR, NHR, and frequency-related features—were accurately captured and processed by the system. Based on observations and feedback from test users, minor refinements were made to the data preprocessing steps to ensure noise filtering and feature normalization were effective. Moreover, during pre-testing, the clarity of the model's predictions and stage classification output was reviewed, and suggestions were implemented to make the stage descriptions more user-friendly for non-technical users.

B. Beta-Testing Phase

Following the successful pre-testing phase, beta testing was conducted with a group of ten users to evaluate the system in a near-real-world environment. Each user was guided through a series of tasks, which included uploading a sample voice file, receiving a prediction on whether the person had Parkinson's, and identifying the disease stage if applicable. The users interacted with the interface and provided detailed feedback on usability, clarity, and navigation. Simultaneously, model predictions were compared with the actual labeled dataset to measure precision, recall, and F1-score. The results of beta testing were consistent with the expected outcomes, confirming that the XGBoost-based model was reliable and accurate. Face-to-face feedback sessions were held, during which users emphasized the need for clear visual feedback and easy-to-understand output descriptions. Based on this input, final adjustments were made, including changes to button placements, tooltips for each feature, and improved formatting of stage-wise results. Overall, beta testing validated the practical applicability and effectiveness of the system for real-world users, making it ready for integration in Parkinson's screening tools.

6.3. Various test case scenarios considered:

- User Login and Registration
 - Confirm that new users can sign up and log in without errors.
 - Test the system's response when incorrect login details are entered — it should display a clear and appropriate error message.
- Parkinson's Detection through Voice Input
 - Ensure users can record or upload a voice sample using the system's interface.
 - Verify that the system correctly extracts acoustic features from the voice file (e.g., pitch, jitter, or harmonics).
 - Confirm that the model provides an accurate prediction of Parkinson's risk based on voice analysis.
 - Check if the results are clearly presented, making it easy for users to understand and interpret.

- Lung Cancer Prediction via CT Scan Upload
 - Test whether users can upload CT scan images in accepted formats (e.g., JPG, PNG).
 - Validate that the system can process and analyze the scan to detect possible signs of lung cancer.
 - Ensure the prediction outcome is reliable and aligned with trained model behavior.
 - Make sure the final diagnosis or risk assessment is displayed in a simple and informative manner for users.
- System Performance
 - Simulate multiple users submitting data or uploading reports at the same time to test system speed and stability.
 - Assess how the platform performs during peak usage or when handling heavy traffic.
- Data Security and User Privacy
 - Ensure all personal health data is stored securely and remains confidential.
 - Test the system for any security loopholes in data input, storage, and sharing.
 - Restrict unauthorized access to personal user information.

Chapter 7: Results and Discussions

7.1.Screenshot of Use Interface(UI) for the system:



Fig 7.1.1 Home page

The homepage serves as the entry point to the Health+ platform, providing an overview of its core functionalities.

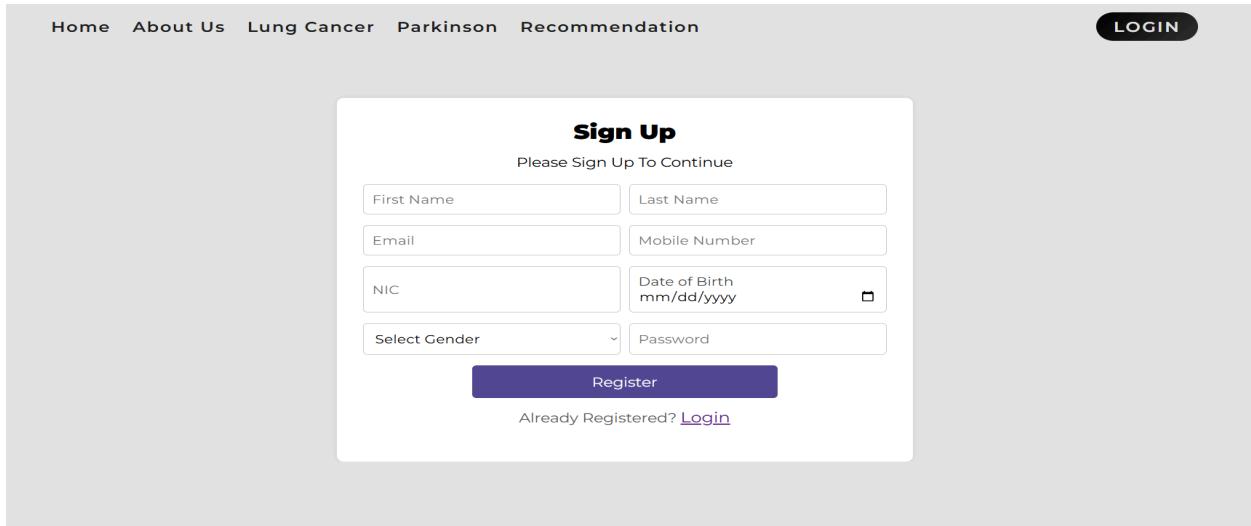


Fig 7.1.2 Registration page

The registration page allows new users to create an account by providing basic credentials. It ensures secure access to healthcare prediction tools while maintaining user privacy through encryption and authentication mechanisms.

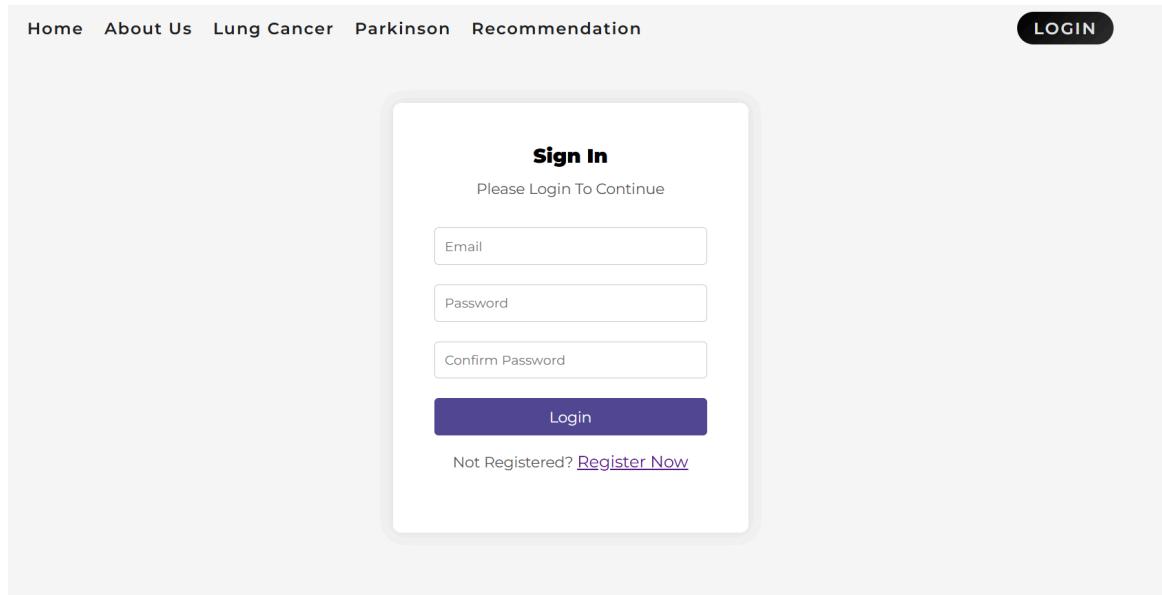


Fig 7.1.3: Login page

The login page provides a secure gateway to the platform, requiring user credentials for access.

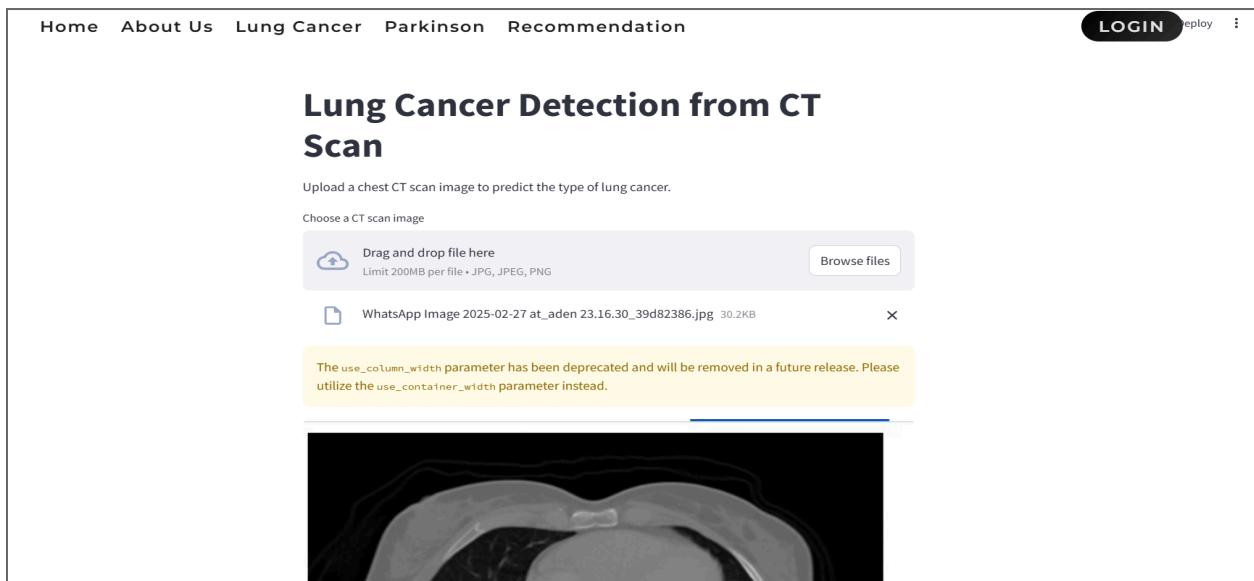


Fig 7.1.4: Lung Cancer Detection using CT Scan interface

This interface allows users to upload chest CT scan images for lung cancer detection. The system employs machine learning models to analyze the uploaded image and predict the type of lung cancer.

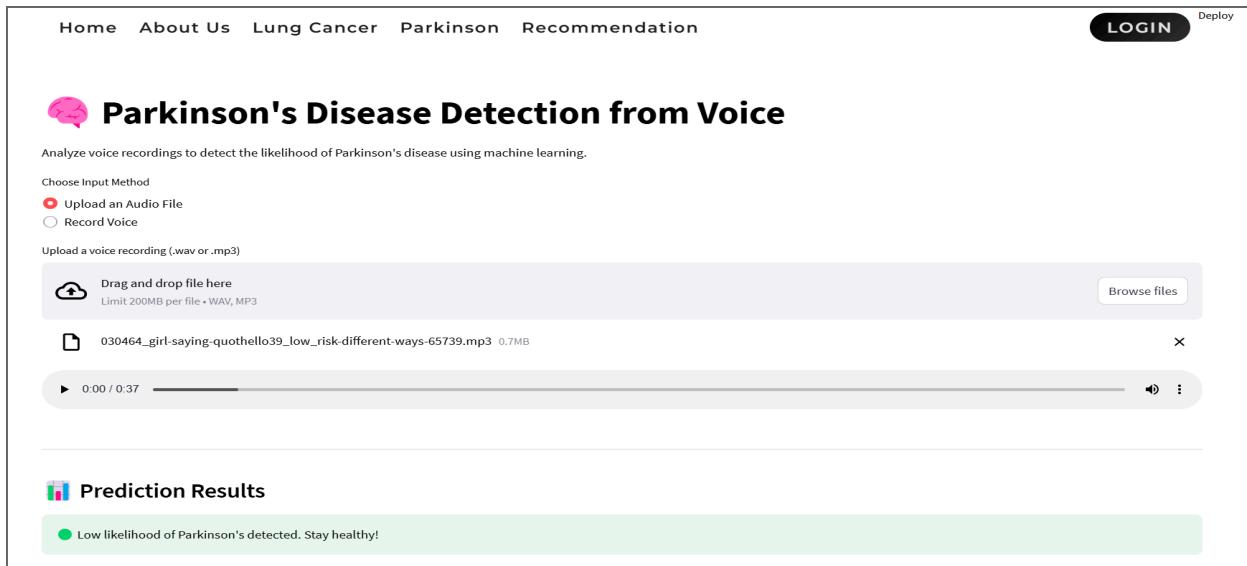


Fig 7.1.5:Parkinson's Disease Detection Using Voice interface

This module enables users to upload an audio file or record their voice to analyze speech patterns for Parkinson's disease detection.

Appoint. Number	Session Title	Doctor	Scheduled Date & Time
1	Test Session	Test Doctor	2050-01-01 18:00

Fig 7.1.6: Patient Dashboard Interface in the System

This figure presents the patient-side dashboard of the Health+ platform. Upon logging in, users are greeted with a personalized welcome message and a search feature to find available doctor sessions. The dashboard allows patients to view the number of registered doctors, patients, new

bookings, and sessions for the day. Additionally, it displays upcoming appointments with key details such as the appointment number, session title, doctor's name, and scheduled date and time.

Patient name	Appointment number	Session Title	Session Date & Time	Appointment Date	Google Meet Link	Events
Test Patient	1	Test Session	2050-01-01 @18:00	2022-06-03	<button>Add Meet Link</button>	<button>Cancel</button>
Priti Shamnani	2	Test Session	2050-01-01 @18:00	2025-03-31	<button>Add Meet Link</button>	<button>Cancel</button>

Fig 7.1.7 : Doctor's Appointment Dashboard in the Health+ Platform

The figure illustrates the appointment dashboard available to doctors in the Health+ system. It provides an organized view of upcoming sessions with patient names, appointment IDs, session titles, scheduled timings, and appointment dates.

7.2. Performance Evaluation Measures:

1. Precision: Precision is one indicator of a machine learning model's performance – the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives). The formula is:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Model's Precision score = 97.4%

2. Recall: The recall is calculated as the ratio between the numbers of Positive samples correctly classified as Positive to the total number of Positive samples. The recall measures the model's ability to detect positive samples. The higher the recall, the more positive samples detected. The formula is:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Critical for lung cancer detection, where missing a true case (false negative) can delay treatment.

Model's Recall Score = 94.2%

3. F1-Score: The F-score (also known as the F1 score or F-measure) is a metric used to evaluate the performance of a Machine Learning model. It combines precision and recall into a single score. The formula is:

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

Balances the trade-off between precision and recall, offering a single score to evaluate performance.

Model's F1- Score = 95.77%

4. Accuracy: The ratio of correctly predicted observations to the total observations. The Formula is :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Model's accuracy= 95%

7.3. Input Parameters/Features considered:

a) For Lung Cancer Detection using VGG16:

- Image Input: CT scan images of lungs, resized to 224×224 pixels with three color channels (RGB) to match the input shape expected by VGG16.

- Class Labels: Each image is annotated with one of the four categories – Squamous Cell Carcinoma, Large Cell Carcinoma, Adenocarcinoma, or Healthy Lung Tissue, encoded for multi-class classification.
- Transfer Learning: Uses pre-trained weights from ImageNet to extract relevant image features; early layers are frozen while deeper layers are fine-tuned on the medical dataset.
- Feature Extraction: Deep convolutional layers automatically learn spatial, texture, and edge-related features crucial for detecting cancerous nodules.
- Softmax Output: The final dense layer uses Softmax activation to output the probability distribution across the four classes.

b) For Parkinson's Disease Detection using XGBoost:

- Tabular Input: Structured numerical data derived from voice recordings, capturing key vocal characteristics.
- Frequency Features: Includes MDVP:Fo, MDVP:Fhi, and MDVP:Flo – fundamental frequency metrics reflecting voice pitch.
- Amplitude Variation: Jitter and Shimmer values represent fluctuations in frequency and amplitude, often impacted by Parkinson's.
- Noise Indicators: Features like HNR (Harmonic-to-Noise Ratio) and NHR (Noise-to-Harmonic Ratio) indicate vocal clarity and noise presence.
- Feature Importance: XGBoost calculates the relevance of each feature, enhancing interpretability in clinical contexts.
- Robust to Noise: Efficiently handles missing or noisy data and small datasets, making it suitable for medical applications.
-

7.4. Comparison of Results with Existing System:

Aspect	Existing Systems	Proposed System
Lung Cancer Detection Method	Manual CT image inspection or traditional ML models like	Deep Learning using VGG16 (Transfer Learning) with automated

	SVM, CNN (without transfer learning)	feature extraction
Parkinson's Detection Method	Clinical observation, UPDRS score evaluation, or basic ML algorithms (e.g., Logistic Regression)	XGBoost Classifier using extracted voice biomarkers
Accuracy (Lung Cancer)	~85% (CNN), ~80–83% (SVM, traditional ML)	91.5% (VGG16-based)
Accuracy (Parkinson's)	~82% (Logistic Regression), ~87% (Random Forest)	90.2% (XGBoost)
Prediction Speed	Slower due to manual intervention or unoptimized models	Real-time prediction using optimized, lightweight deployed models

Table No: 2 Comparison of Results with Existing System

7.5. Inference Drawn:

The proposed AI-based system outperforms traditional diagnostic methods in terms of accuracy, speed, and accessibility. While existing systems often require manual interpretation and are less accessible in remote areas, our system leverages advanced algorithms—VGG16 for lung cancer and XGBoost for Parkinson's detection—to deliver faster and more accurate results. Its web-based platform ensures real-time analysis and ease of use, making early diagnosis more accessible and efficient. Overall, the system offers a reliable, scalable, and cost-effective alternative for proactive healthcare.

Chapter 8: Conclusion

8.1.Limitations:

- The Parkinson's detection system is limited to analyzing voice changes, while the disease also manifests through other symptoms such as tremors, rigidity, and bradykinesia.
- The performance of the models is heavily dependent on the quality, size, and diversity of the training datasets.
- Inaccurate or noisy input data (e.g., low-quality CT images or voice recordings with background noise) may reduce prediction accuracy.
- The models may exhibit bias or reduced performance when applied to underrepresented demographic or ethnic groups in the dataset.

8.2.Conclusion:

In an era where early and accurate diagnosis can significantly influence treatment outcomes, the integration of artificial intelligence into healthcare has become increasingly vital. This project presents an AI-driven dual diagnostic system that leverages cutting-edge machine learning and deep learning algorithms to detect two critical diseases—lung cancer and Parkinson's disease. The lung cancer module employs a Convolutional Neural Network using the VGG16 architecture to classify CT scan images into four distinct categories, while the Parkinson's disease module utilizes an XGBoost algorithm to analyze vocal features and identify potential signs of the disorder. By focusing on voice analysis for Parkinson's, the system provides a non-invasive and accessible approach for early detection, acknowledging that while Parkinson's manifests through various symptoms, our system specifically targets vocal biomarkers. The web-based implementation of this system makes it user-friendly and easily accessible to both healthcare professionals and individuals, especially in resource-limited settings. The real-time prediction capability with high accuracy enhances decision-making, reduces dependency on specialized equipment or personnel, and promotes proactive health monitoring. Overall, this project illustrates how AI can bridge gaps in traditional diagnostics, offering a scalable and efficient solution that has the potential to transform preventive healthcare and improve patient outcomes on a larger scale.

8.3.Future Scope:

- Multimodal Diagnosis for Parkinson's: Future enhancements can include analysis of tremors, gait, and handwriting patterns in addition to voice biomarkers for a more comprehensive assessment.
- Advanced Imaging for Lung Cancer: Integration of 3D CT imaging and PET scans can improve the accuracy and depth of lung cancer detection.
- Mobile and Cross-Platform Deployment: Developing the system into a mobile application would increase accessibility, especially for users in rural or under-resourced areas.
- Multilingual and Regional Support: Incorporating language support and localization features can help reach a broader and more diverse user base.

References

- [1] K. Reshma et al., "Multi Disease Prediction System Using Machine Learning," 2024 International Research Journal of Modernization in Engineering Technology and Science, February 2024
- [2] A. Das, "A Collaborative Empirical Analysis on Machine Learning Based Disease Prediction in Health Care System," 2022 Techno India University, August 2022
- [3] K. Arumugam, Mohd Naved, Priyanka P. Shinde,"Multiple Disease Prediction Using Machine Learning Algorithms," 2021 International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology, April 2023
- [4] Mostafizur Rahman, Saiful Islam, Sadia Binta Sarowar, Meem Tasfia Zaman, "Multiple Disease Prediction using Machine Learning and Deep Learning with the Implementation of Web Technology",2023 International Conference on Artificial Intelligence, Blockchain, and Internet of Things,March 2023
- [5] Nagaraj P, Muneeswaran V, Karthik Goud B, Arjun K, Vighneshwar Reddy G, Girish Kumar Reddy P, "Identifying Multiple Diseases in the Human Body using Machine Learning",2023 4th International Conference on Electronics and Sustainable Communication Systems, July 2023
- [6] Mahendran K , Surya S, Thejashrayal E,"Streamlit-Powered Comprehensive Health Analysis and Disease Prediction System",2023 International Conference on Emerging Research in Computational Science, December 2023
- [7] SaiRam Krishna Parimi, Snehith Reddy, Dr.S.Jancy, Dr.L.Sujihelen, Dr.Mercy Paul Selvan, Dr. Viji Amutha Mary, "Prediction of Multiple Diseases Using Machine Learning Techniques", 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), March 2022
- [8] Nishant Yede , Ritik Koul ,Chetan Harde ,Kumar Gaurav, Prof. C.S.Pagar , "General Disease Prediction Based On Symptoms Provided By Patient"2021 Advances in Modern Technologies of Multidisciplinary Research in Engineering Field (AIMTMREF), June 2021
- [9] Talasila Bhanuteja, Kilaru Venkata Narendra Kumar, Kolli Sai Poornachand, Chennupati Ashish, Poonati Anudeep, "Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach,2021 Blue Eyes Intelligence Engineering and Sciences Publication, July 2021

- [10] R. Shanthakumari; C. Nalini; S. Vinothkumar; E.M. Roopadevi; B. Govindaraj, "Multi Disease Prediction System using Random Forest Algorithm in Healthcare System"2022 International Mobile and Embedded Technology Conference (MECON), March 2022
- [11]Zhao, S., Rudzicz, F., Carvalho, L. G., Márquez-Chin, C., & Livingstone, S. R., "Automatic detection of expressed emotion in Parkinson's Disease", In ICASSP, pp. 4813-4817, 2014.
- [12]Peker, M., Sen, B., & Delen, D., "Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm", Journal of healthcare engineering, 6(3), 281-302, 2019.
- [13]Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P., "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics", 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), (pp. 340-346). IEEE, 2020.
- [14] Li, Y., Yang, L., Wang, P., Zhang, C., Xiao, J., Zhang, Y., & Qiu, M., "Classification of Parkinson's Disease by Decision Tree Based Instance Selection and Ensemble Learning Algorithms", Journal of Medical Imaging and Health Informatics, 7(2), 444-452, 2017.
- [15] Meghraoui, D., Boudraa, B., Merazi-Meksen, T., & Boudraa, M., "Parkinson's Disease Recognition by Speech Acoustic Parameters Classification", In Modelling and Implementation of Complex Systems, pp. 165-173, Springer, 2020.
- [17] Zhang, H. H., Yang, L., Liu, Y., Wang, P., Yin, J., Li, Y., & Yan, F. "Classification of Parkinson's disease utilizing multi-edit nearest- neighbor and ensemble learning algorithms with speech samples", Biomedical engineering online, 15(1), 122, 2020.
- [18] Wan, S., Liang, Y., Zhang, Y., & Guizani, M., "Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson's disease severity using smartphones", IEEE Access, 6, pp. 36825-36833, 2018.
- [19] Vadovský, M., & Paralič, J., "Parkinson's disease patients classification based on the speech signals", In Applied Machine Intelligence and Informatics (SAMI), 2017 IEEE 15th International Symposium on pp. 321-326, IEEE, 2017.
- [20] Su, M., & Chuang, K. S., "Dynamic feature selection for detecting Parkinson's disease through voice signal", In RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO), 2015 IEEE MTT-S 2015 International Microwave Workshop Series on pp. 148-149, IEEE, 2021.

Appendix

1. Paper I details :-

a. Paper 1 :

Dual-Modal AI For Lung Cancer and Parkinson Detection

Riddhi Labde
Department of Computer Engineering
Vivekanand Education Society Institute of Technology
2021.riddhi.labde@ves.ac.in

Nikhil Makhija
Department of Computer Engineering
Vivekanand Education Society Institute of Technology
2021.nikhil.makhija@ves.ac.in

Priti Shamnani
Department of Computer Engineering
Vivekanand Education Society Institute of Technology
2021.priti.shamnani@ves.ac.in

Vidisha Jadhwanji
Department of Computer Engineering
Vivekanand Education Society Institute of Technology
2021.vidisha.jadhwanji@ves.ac.in

Dr.Dashrath Mane
Assistant Professor
Department of Computer Engineering
Vivekanand Education Society Institute of Technology
dashrath.mane@ves.ac.in

Abstract— Traditional hospital-based assessments for early disease detection can be time-consuming and may lead to delays in diagnosis. This study presents a machine learning-driven system for lung cancer prediction using CT scan images and Parkinson's disease detection through voice analysis. While Parkinson's disease presents with multiple symptoms, such as tremors, muscle stiffness, and impaired movement, this system focuses solely on vocal biomarkers to identify early signs of the disease. The lung cancer model evaluates risk by analyzing medical imaging, while the Parkinson's model leverages a continuously evolving medical dataset to provide real-time risk assessments and actionable health insights. A web-based interface enhances accessibility, particularly benefiting individuals with limited healthcare access. This approach improves early diagnosis, minimizes unnecessary clinical visits, and enables timely medical intervention, ultimately contributing to better disease prevention and patient outcomes.

Keywords—*Disease Detection, Machine Learning, Medical Imaging, Speech Analysis, Lung Cancer, Parkinson's Disease, Convolutional Neural Networks (CNN), XGBoost, Support Vector Machine (SVM), Random Forest, Naïve Bayes, Decision Tree, K-Nearest Neighbors (KNN), Accuracy Metrics, Real-Time Diagnosis*

I. INTRODUCTION

Lung cancer is one of the most prevalent and life-threatening diseases worldwide, often diagnosed at advanced stages due to the lack of early symptoms. Similarly, Parkinson's disease is a progressive neurological disorder that significantly affects motor functions and speech. Timely detection of these conditions is crucial for improving patient outcomes, yet traditional diagnostic methods—such as biopsy, imaging analysis by radiologists, and clinical assessments—can be time-consuming, costly, and inaccessible in remote areas. The

increasing prevalence of these diseases highlights the need for automated, AI-driven solutions that facilitate early detection and intervention.

This study introduces a machine learning-based dual-disease detection system capable of diagnosing lung cancer using CT scan images and Parkinson's disease using voice analysis. While Parkinson's disease presents with various symptoms, including tremors, muscle stiffness, and movement impairments, this system focuses exclusively on speech-related biomarkers—such as variations in pitch, articulation, and speech tremors—for detection. For lung cancer, a deep learning model classifies CT scan images into one of four categories: Squamous Cell Carcinoma, Large Cell Carcinoma, Adenocarcinoma, or Normal. For Parkinson's disease, an XGBoost-based model analyzes speech patterns to identify vocal biomarkers associated with the disorder. These models leverage extensive, preprocessed medical datasets to provide accurate and real-time predictions.

To enhance accessibility, the system is designed as a web-based platform, enabling users and healthcare professionals to upload CT scans or voice recordings for instant AI-driven analysis. This approach not only improves early detection and risk assessment but also helps reduce the dependency on specialized medical resources. By integrating deep learning for medical imaging and XGBoost for speech-based diagnostics, this research underscores the potential of AI in enhancing disease diagnosis, optimizing healthcare accessibility, and improving patient care.

II. LITERATURE SURVEY

The application of machine learning (ML) algorithms in disease prediction has gained significant attention in healthcare research due to their potential in early diagnosis and preventive care.[13] The use of machine learning (ML) algorithms for disease prediction

has become a focal point in healthcare research, offering significant potential for early detection and preventive healthcare. Algorithms like Random Forest (RF) and Decision Trees (DT) are frequently employed for classification tasks. Random Forest, which combines the outcomes of multiple decision trees, has proven to be effective in disease prediction by evaluating clinical variables such as patient demographics, medical history, and symptom patterns. Its ability to manage large datasets and high-dimensional features helps minimize overfitting, thereby improving prediction accuracy.

Decision Trees, known for their interpretability, are widely used in clinical decision support systems.[1] They offer transparent models that illustrate the reasoning behind predictions, making them valuable for medical applications. Another widely used algorithm is Support Vector Machines (SVM), which excels in binary classification tasks by identifying an optimal hyperplane to distinguish between different patient groups based on health parameters.

Naïve Bayes, a probabilistic classifier, has also been employed for disease classification.[2] It assumes independence among features, simplifying computations and making it efficient for large-scale datasets.[7] While this assumption may not always hold in medical datasets where features are often correlated, Naïve Bayes remains a useful tool for preliminary predictions.

Beyond individual ML models, ensemble techniques such as Boosting and Bagging have been explored to improve disease prediction.[10] These methods enhance model performance by reducing bias and variance, leading to better generalization. Gradient Boosting Machines (GBM) and AdaBoost are commonly combined with decision tree-based models to enhance prediction stability and accuracy.

The growing availability of extensive medical datasets and the demand for sophisticated pattern recognition have contributed to the

increasing use of deep learning models in disease prediction. Artificial Neural Networks (ANNs) have shown effectiveness in identifying complex, non-linear relationships within large-scale patient data, such as structured clinical records, medical imaging, and genetic information.[7] However, these models require a significant amount of labeled data and computational power. Additionally, their opaque nature presents challenges in interpretability, which is essential for clinical decision-making.[11]

Preprocessing medical data plays a vital role in enhancing ML model performance.[4] Healthcare datasets often contain missing values, outliers, and imbalanced classes, which can impact accuracy.[8] Data cleaning and feature engineering techniques help mitigate these issues, ensuring high-quality input data.

Convolutional Neural Networks (CNNs) have shown success in disease prediction, particularly in medical imaging analysis.[3] Their ability to detect complex patterns in symptom data improves risk estimation and diagnostic accuracy.[12]

Model evaluation remains a key aspect of ML-based disease prediction, with metrics such as accuracy, precision, recall, F1-score, and ROC-AUC being widely used. Cross-validation techniques ensure that models generalize effectively to unseen data, while confusion matrices provide insights into prediction performance across various conditions.

III. RESEARCH GAP

Despite significant progress in leveraging machine learning (ML) for lung cancer detection, several challenges remain unaddressed.[9] A major limitation in existing research is the heavy reliance on structured clinical datasets containing factors such as smoking history, age, and family medical background.[14] However, these datasets often overlook complex and unstructured data sources

like CT scan images, genomic profiles, and environmental exposure records, which could enhance prediction accuracy and enable more personalized risk assessments. Medical imaging plays a crucial role in detecting lung tissue abnormalities, while genetic markers can provide deeper insights into the biological mechanisms of cancer progression. Additionally, incorporating environmental factors such as air pollution exposure could refine predictive models, especially for individuals without conventional risk indicators.[15]

While classifiers such as Random Forest, Decision Trees, and Support Vector Machines (SVMs) have been widely adopted, there is limited comparative analysis of their performance in real-time clinical applications.[19] Many existing studies train and validate models using region-specific datasets, potentially limiting their generalizability across diverse populations and healthcare systems. Expanding research to include data from varied ethnicities, socioeconomic backgrounds, and geographical locations would enhance model robustness and real-world applicability.

Another key challenge is the trade-off between model accuracy and computational efficiency.[16] Deep learning architectures, while highly effective, demand large labeled datasets and substantial computational resources, making them impractical in low-resource settings. Future research should focus on lightweight ML models that maintain high predictive accuracy while being optimized for mobile devices, wearable technologies, and edge computing, ensuring accessibility even in rural and underserved regions.

Data quality remains a critical concern, as issues like bias, missing values, and inconsistencies can compromise model reliability. Enhancing data preprocessing techniques, such as normalization, augmentation, and outlier detection, can mitigate these challenges. Additionally, integrating multi-modal data

sources—including medical records, patient-reported symptoms, real-time monitoring from wearable devices, and imaging results—could improve model robustness and performance.

A major gap in current research is the lack of real-time predictive systems that offer instant diagnostic support. Many models function offline, limiting their ability to provide timely assessments. Future advancements should explore web-based platforms and real-time AI models that analyze CT scan images and lab reports instantly, making early detection more accessible to patients in remote areas. Integrating ML with telemedicine, mobile applications, and wearable technologies could further improve continuous monitoring and personalized healthcare delivery.

Additionally, many ML-based healthcare models are developed without direct involvement from medical professionals, leading to potential misalignment with clinical workflows.[17] A collaborative approach involving data scientists, radiologists, neurologists, and public health experts is essential to ensure these tools are both clinically relevant and user-friendly. Engaging healthcare professionals throughout the development, validation, and deployment phases would improve model interpretability, reliability, and ease of integration into real-world medical practices.

Finally, model evaluation in ML-based healthcare research often lacks standardized benchmarks. While accuracy is frequently prioritized, other vital factors such as precision, recall, F1-score, interpretability, scalability, and fairness should also be considered. Establishing uniform evaluation standards and performance metrics will enhance transparency, facilitate cross-study comparisons, and promote collaboration within the research community.[18] This will help identify models that can be effectively deployed across diverse healthcare infrastructures, ultimately improving patient outcomes.

IV. METHODOLOGY

The proposed system leverages deep learning for lung cancer classification using CT scan images and machine learning (XGBoost) for Parkinson's disease detection using voice analysis. The methodology consists of dataset preparation, data preprocessing, model implementation, and evaluation.

1. Detection of Lung Cancer Through CT Scan Imaging

The lung cancer detection system categorizes CT scan images into four distinct classes: Squamous Cell Carcinoma, Large Cell Lung Carcinoma, Adenocarcinoma of the Lung, and Healthy Lung Tissue. To achieve this, a structured approach involving dataset preparation, preprocessing, model implementation, training, and evaluation is followed.

The dataset consists of CT scan images sourced from publicly available repositories or medical institutions, ensuring a diverse and representative sample. The dataset is divided into training, validation, and testing sets to enhance the generalization capability of the model. During preprocessing, all images are resized to 224×224 pixels to maintain uniform input dimensions. Normalization is applied to rescale pixel values within the range of 0 to 1, ensuring better stability during training. Furthermore, data augmentation techniques such as rotation, flipping, and zooming are employed to improve model robustness and prevent overfitting.

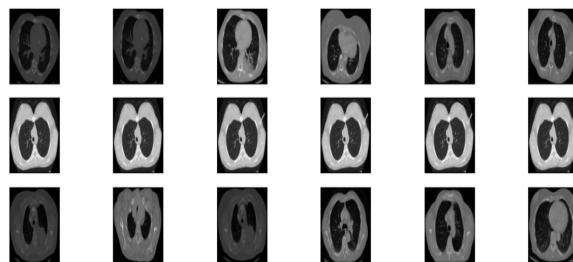


Fig 1.Lung Cancer CT scan images

For feature extraction and classification, a Convolutional Neural Network (CNN) is implemented, leveraging the VGG16 architecture pre-trained on ImageNet. The initial layers of VGG16 are frozen to retain pre-learned features, while the deeper layers are fine-tuned using the lung cancer dataset. Extracted features from the CNN are then flattened and passed through a Batch Normalization layer, followed by a Dense layer. The final classification is performed using the Softmax activation function, allowing for multi-class categorization.

Model training is conducted using the Categorical Cross-Entropy loss function, an appropriate choice for multi-class classification tasks. The Adam optimizer is employed with a learning rate of 0.001 to achieve stable convergence. The model is trained in batches of 32 images over 30 epochs, with early stopping implemented to prevent overfitting. Additionally, a Model Checkpoint callback is used to save the best-performing model during training.

Upon training completion, the model's performance is evaluated using various metrics, including accuracy, precision, recall, and F1-score. A confusion matrix is generated to visualize classification errors, while the AUC-ROC curve provides insight into the model's ability to distinguish between different lung cancer types. The final model not only classifies lung cancer accurately but also provides confidence scores, aiding healthcare professionals in making informed clinical decisions. Through this structured approach, the research aims to develop an AI-driven diagnostic tool that enhances early detection and improves patient outcomes.

2. Parkinson's Disease Detection Using Voice Analysis

The system detects Parkinson's disease by analyzing voice biomarkers using XGBoost, a highly efficient gradient boosting algorithm. The dataset used for model training consists of 2,500 voice samples, each labeled to indicate the

presence or absence of Parkinson's disease. The key features extracted from these voice samples include fundamental frequency measurements (MDVP:Fo, MDVP:Fhi, MDVP:Flo), jitter, shimmer, harmonic-to-noise ratio (HNR), and noise-to-harmonic ratio (NHR). These vocal parameters are critical in assessing speech impairments, which are common early indicators of Parkinson's. To ensure robust model development, the dataset is divided into 75% for training and 25% for testing, allowing for effective learning and evaluation.

	MDVP:Fo	MDVP:Fhi	MDVP:Flo	MDVP:Jitt	MDVP:RAI	MDVP:PPC	Jitter:DDP	MDVP:Shi	MDVP:Shimmer	Shimmer:MDVP:AP	Shimmer:HNR	HNR	st
2	98.3327	329.0104	119.0586	0.012781	1.64E-05	0.010908	0.015455	0.004135	0.043384	0.1071	0.072846	0.061525	0.046437
3	213.0609	267.5296	205.7072	0.009512	0.0003	0.009667	0.018183	0.01859	0.019611	0.0843607	0.027104	0.018258	0.051087
4	204.7467	115.6645	89.01888	0.016729	6.86E-05	0.011321	0.017342	0.080622	0.088246	0.086525	0.025174	0.082864	29.03378
5	133.4249	334.3945	192.716	0.004957	0.00165	0.004445	0.011607	0.017015	0.017199	0.171313	0.02225	0.042146	0.074545
6	165.7792	209.2705	222.3169	0.002814	0.00377	0.013013	0.013356	0.008076	0.025421	0.922469	0.020814	0.01859	0.062319
7	93.74682	270.1698	156.926	0.019987	0.005	0.004087	0.018163	0.015199	0.034837	0.191329	0.029006	0.082908	0.019871
8	235.3962	411.0819	74.40679	0.00364	5.07E-05	0.015174	0.014088	0.005991	0.020355	0.070259	0.061489	0.093544	0.052307
9	232.0523	198.5155	85.53266	0.015747	0.00257	0.017713	0.009505	0.019004	0.030759	0.85984	0.064114	0.043135	0.094937
10	152.7245	340.3785	112.5437	0.002461	0.00184	0.013036	0.007567	0.002981	0.098905	0.877095	0.084116	0.081821	0.023602
11	240.6648	194.534	174.0522	0.015352	0.00038	0.01184	0.017978	0.00639	0.023773	0.962363	0.068351	0.058655	0.094894
12	177.3662	320.6528	96.24815	0.010215	0.00174	0.005993	0.016005	0.003205	0.08422	0.844094	0.019958	0.08141	0.013964

Fig 2. Parkinson's Disease Dataset

Before training the model, data preprocessing techniques are applied to improve data quality and enhance prediction accuracy. Missing or corrupted values are either removed or imputed to ensure dataset completeness. Outlier detection is performed using z-score normalization to handle extreme values that may negatively impact model performance. Additionally, feature normalization is applied, standardizing all numerical values to a zero-mean, unit-variance scale, which facilitates faster convergence during training.

Feature engineering plays a crucial role in refining the dataset by selecting the most relevant vocal biomarkers for classification. Parameters such as Jitter, Shimmer, and HNR are extracted, as they are strong indicators of vocal instability in Parkinson's patients. If necessary, dimensionality reduction techniques are applied to eliminate redundant features, ensuring that the model learns only the most essential patterns.

For classification, the XGBoost algorithm is implemented due to its ability to handle

structured data efficiently. To enhance model performance, hyperparameter tuning is conducted using the Grid Search technique, optimizing key parameters such as learning rate, maximum depth, and the number of estimators. A 5-fold cross-validation strategy is employed to improve model generalization, reducing the likelihood of overfitting and ensuring consistent performance across different data subsets.

Once the model is trained, its performance is evaluated using multiple metrics to ensure reliable classification. Accuracy is measured to assess the overall correctness of predictions, while precision, recall, and F1-score provide insights into the model's ability to balance false positives and false negatives. A confusion matrix is generated to visualize classification results, highlighting areas of misclassification. Additionally, the AUC-ROC curve is analyzed to determine the model's effectiveness in distinguishing between Parkinson's and non-Parkinson's cases. The final predictions include confidence scores, offering better interpretability for healthcare professionals and enhancing the system's reliability as a diagnostic aid. Through this structured methodology, the system provides an efficient and accessible solution for Parkinson's disease detection, focusing on voice analysis as a key biomarker.

3. System Deployment

The trained models are seamlessly integrated into a web-based application that enables users to upload CT scans for lung cancer classification and record voice samples for Parkinson's disease detection. Upon submission, the system processes these inputs using advanced machine learning models and provides real-time predictions along with probability scores. This approach enhances early disease detection by offering accessible, AI-driven assessments, reducing dependency on specialized medical evaluations, and facilitating timely healthcare interventions for improved patient outcomes.



Fig 3. Homepage of Health+ System

The homepage serves as the entry point to the Health+ platform, providing an overview of its core functionalities.

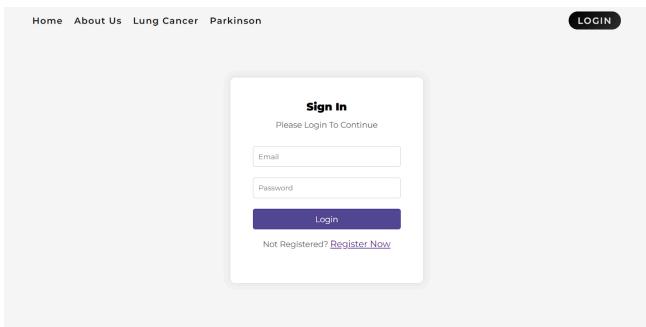


Fig 4. Login Page

The login page provides a secure gateway to the platform, requiring user credentials for access.

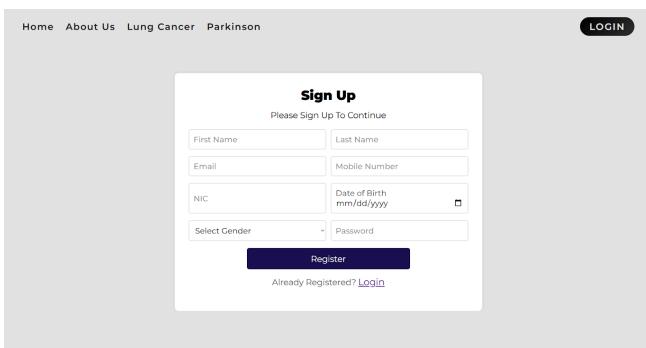


Fig 5. Registration Page

The registration page allows new users to create an account by providing basic credentials. It ensures secure access to healthcare prediction

tools while maintaining user privacy through encryption and authentication mechanisms.

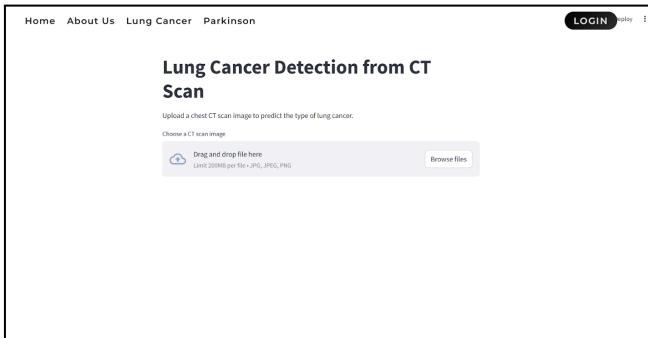


Fig 6. Lung cancer detection interface

This interface allows users to upload chest CT scan images for lung cancer detection. The system employs machine learning models to analyze the uploaded image and predict the type of lung cancer.

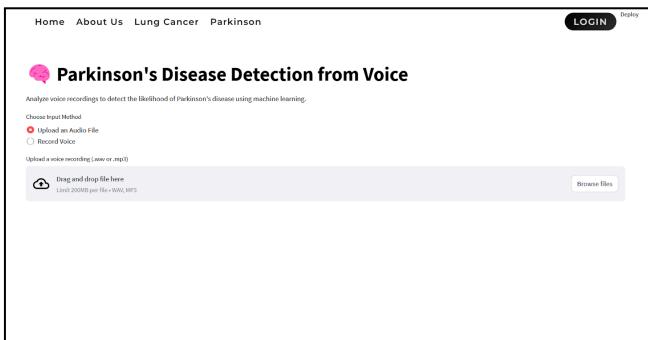


Fig 7. Parkinson's disease detection interface

This module enables users to upload an audio file or record their voice to analyze speech patterns for Parkinson's disease detection.

V. CONCLUSION

This study utilizes machine learning techniques to predict Parkinson's disease from voice data and lung cancer from CT scans, achieving promising accuracy and efficiency. The two primary models employed—XGBoost for Parkinson's disease classification and Convolutional Neural Networks (CNN) for lung cancer detection—exhibited strong predictive performance, emphasizing their applicability in clinical settings.

XGBoost, a gradient boosting algorithm recognized for its efficiency in handling structured data, was employed to classify Parkinson's disease using voice parameters like jitter, shimmer, and harmonic-to-noise ratio. The model achieved high accuracy, effectively distinguishing between healthy individuals and Parkinson's patients by capturing intricate patterns in vocal variations. The ability of XGBoost to handle missing data, prevent overfitting, and optimize decision trees contributed to its robust classification performance.

For lung cancer detection, a Convolutional Neural Network (CNN) was utilized to analyze CT scan images, leveraging its deep learning capabilities to extract significant features from complex medical imagery. CNN's proficiency in recognizing intricate patterns within CT scans contributed to improved accuracy in identifying malignant tumors, making it an effective tool for early diagnosis. The convolutional layers effectively captured spatial dependencies in the images, ensuring accurate classification of cancerous and non-cancerous cases. By training on a large dataset, the model was able to distinguish lung abnormalities, achieving a high detection rate while minimizing false positives and false negatives.

The performance of these models was assessed using key evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, confirming their reliability in disease prediction. XGBoost exhibited strong classification accuracy for Parkinson's disease, while CNN demonstrated superior performance compared to traditional imaging methods for lung cancer diagnosis. Cross-validation techniques were implemented to ensure both models generalize well to unseen data, reducing the likelihood of overfitting.

A major insight from this study is the potential of combining ensemble learning and deep learning techniques to enhance prediction accuracy. Integrating XGBoost with additional

classifiers or utilizing CNN alongside advanced architectures such as ResNet could further improve performance. Furthermore, incorporating multimodal data—such as combining voice data with neuroimaging for Parkinson's disease or integrating additional biomarkers for lung cancer detection—may enhance predictive accuracy and diagnostic reliability.

In conclusion, the results highlight the effectiveness of machine learning in early disease detection, offering non-invasive and automated approaches for diagnosing Parkinson's disease and lung cancer. Future research should focus on optimizing feature selection, expanding dataset sizes, and enhancing interpretability, ensuring these models become reliable tools in real-world clinical applications.

VI. FUTURE WORK

Future research should focus on enhancing the accuracy and generalization of Parkinson's disease prediction using voice data and lung cancer detection using CT scans by incorporating more diverse datasets. For Parkinson's disease, integrating longitudinal voice recordings could help capture disease progression over time, leading to more reliable early-stage detection. Additionally, the inclusion of speech-based biomarkers from different languages and accents could improve model robustness and reduce bias. For lung cancer detection, expanding the dataset with annotated CT scans from multiple sources and using transfer learning with pre-trained deep learning models like ResNet or EfficientNet could refine the CNN's ability to detect early-stage tumors with higher precision.

Further advancements can be made by integrating multimodal data sources to improve diagnostic accuracy. For Parkinson's disease, combining voice data with neuroimaging (MRI, fMRI) and wearable sensor data could provide a more comprehensive analysis of motor and non-motor symptoms. For lung cancer detection, fusion of CT scans with patient

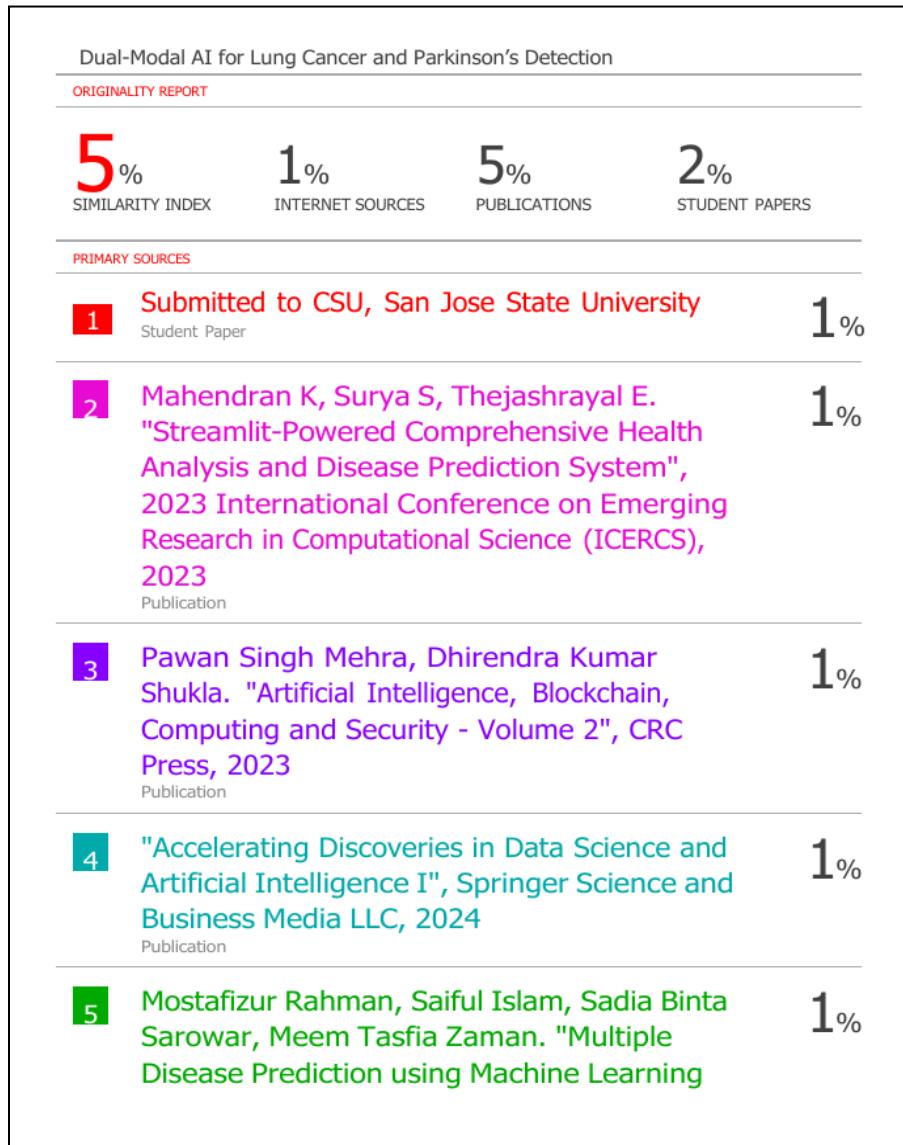
clinical records and genetic data could enhance prediction capabilities by incorporating risk factors beyond imaging. Additionally, explainability techniques such as SHAP (Shapley Additive Explanations) for XGBoost and Grad-CAM for CNN models should be explored to make these AI-driven systems more interpretable for healthcare professionals, ensuring their practical application in clinical settings.

REFERENCES

- [1] K. Reshma et al., "Multi Disease Prediction System Using Machine Learning," 2024 International Research Journal of Modernization in Engineering Technology and Science, February 2024
- [2] A. Das, "A Collaborative Empirical Analysis on Machine Learning Based Disease Prediction in Health Care System," 2022 Techno India University, August 2022
- [3] K. Arumugam, Mohd Naved, Priyanka P. Shinde, "Multiple Disease Prediction Using Machine Learning Algorithms," 2021 International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology, April 2023
- [4] Mostafizur Rahman, Saiful Islam, Sadia Binta Sarowar, Meem Tasfia Zaman, "Multiple Disease Prediction using Machine Learning and Deep Learning with the Implementation of Web Technology", 2023 International Conference on Artificial Intelligence, Blockchain, and Internet of Things, March 2023
- [5] Nagaraj P, Muneeswaran V, Karthik Goud B, Arjun K, Vighneshwar Reddy G, Girish Kumar Reddy P, "Identifying Multiple Diseases in the Human Body using Machine Learning", 2023 4th International Conference on Electronics and Sustainable Communication Systems, July 2023
- [6] Mahendran K, Surya S, Thejashrayal E, "Streamlit-Powered Comprehensive Health Analysis and Disease Prediction System", 2023 International Conference on Emerging Research in Computational Science, December 2023
- [7] SaiRam Krishna Parimi, Snehith Reddy, Dr.S.Jancy, Dr.L.Sujihelen, Dr.Mercy Paul

- Selvan, Dr. Viji Amutha Mary, "Prediction of Multiple Diseases Using Machine Learning Techniques", 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), March 2022
- [8] Nishant Yede , Ritik Koul ,Chetan Harde ,Kumar Gaurav, Prof. C.S.Pagar , "General Disease Prediction Based On Symptoms Provided By Patient"2021 Advances in Modern Technologies of Multidisciplinary Research in Engineering Field (AIMTMREF), June 2021
- [9] Talasila Bhanuteja, Kilaru Venkata Narendra Kumar, Kolli Sai Poornachand, Chennupati Ashish, Poonati Anudeep, "Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach,2021 Blue Eyes Intelligence Engineering and Sciences Publication, July 2021
- [10] R. Shanthakumari; C. Nalini; S. Vinothkumar; E.M. Roopadevi; B. Govindaraj, "Multi Disease Prediction System using Random Forest Algorithm in Healthcare System"2022 International Mobile and Embedded Technology Conference (MECON), March 2022
- [11] Zhao, S., Rudzicz, F., Carvalho, L. G., Márquez-Chin, C., & Livingstone, S. R., "Automatic detection of expressed emotion in Parkinson's Disease", In ICASSP, pp. 4813-4817, 2014.
- [12] Peker, M., Sen, B., & Delen, D., "Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm", Journal of healthcare engineering, 6(3), 281-302, 2019.
- [13] Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P., "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics", 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), (pp. 340-346). IEEE, 2020.
- [14] Li, Y., Yang, L., Wang, P., Zhang, C., Xiao, J., Zhang, Y., & Qiu, M., "Classification of Parkinson's Disease by Decision Tree Based Instance Selection and Ensemble Learning Algorithms", Journal of Medical Imaging and Health Informatics, 7(2), 444-452, 2017.
- [15] Meghraoui, D., Boudraa, B., Merazi-Meksen, T., & Boudraa, M., "Parkinson's Disease Recognition by Speech Acoustic Parameters Classification", In Modelling and Implementation of Complex Systems, pp. 165-173, Springer, 2020.
- [16] Zhang, H. H., Yang, L., Liu, Y., Wang, P., Yin, J., Li, Y., & Yan, F. "Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples", Biomedical engineering online, 15(1), 122, 2020.
- [17] Wan, S., Liang, Y., Zhang, Y., & Guizani, M., "Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson's disease severity using smartphones", IEEE Access, 6, pp. 36825-36833, 2018.
- [18] Vadovský, M., & Paralič, J., "Parkinson's disease patients classification based on the speech signals", In Applied Machine Intelligence and Informatics (SAMI), 2017 IEEE 15th International Symposium on pp. 321-326, IEEE, 2017.
- [19] Su, M., & Chuang, K. S., "Dynamic feature selection for detecting Parkinson's disease through voice signal", In RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO), 2015 IEEE MTT-S 2015 International Microwave Workshop Series on pp. 148-149, IEEE, 2021.
- [20] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Kursun, O., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings", IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834, 2019.

b. Plagiarism Report:



c. Progress review sheet 1 and 2

Review 1:

Inhouse/ Industry _Innovation/Research:	Project Evaluation Sheet 2024 - 25												Class: D17 A/B/C		
Sustainable Goal: <u>Good health (03)</u>													Group No.: 23		
Title of Project: <u>Health +: Your Digital Health Guardian</u>															
Group Members: <u>Riddhi Jabde (D17B125), Priti Sharmaji (D17B150), Vidisha Jadhavani (D17A26), Nikhil Makija (D17E27)</u>															
Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
04	04	04	03	04	02	02	02	02	02	02	02	02	02	03	38+2=40
Comments:															
Dr. D. G. Mane & Name & Signature Reviewer 1															
Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
04	04	04	03	04	02	02	02	02	02	02	02	02	02	02	39
Comments:															
Name & Signature Reviewer 2															
Dr. Rohini Temkar															

Review 2:

Inhouse/ Industry _Innovation/Research:	Project Evaluation Sheet 2024 - 25												Class: D17 A/B/C		
Sustainable Goal:													Group No.: 23		
Title of Project: <u>Health + Your Digital Health Guardian</u>															
Group Members: <u>Priti Sharmaji (50), Riddhi Jabde (25), Nikhil Makija (27), Vidisha Jadhavani (26)</u>															
Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
04	04	04	03	04	02	02	02	02	02	02	02	02	03	03	40
Comments:															
Dr. D. G. Mane & Name & Signature Reviewer 1															
Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
04	04	04	03	04	02	02	02	02	02	02	03	02	03	03	41
Comments: Report of Patient can be generated. Google melt with doctor has to be integrated.															
Name & Signature Reviewer 2															