**Abstract**

Heart disease is one of the major problems worldwide today. Effective early detection methods are needed for timely interventions. This project presents the comprehensive strategy and techniques to predict heart disease through the detailed analysis of electrocardiogram signals, and ECG image reports. Developing an advanced machine learning model by combining demographic information with ECG signal data, exploring ECG signal processing, and converting ECG image-based reports to signals. The project aims to make disease predictions more accurate and develop machine learning models to predict heart diseases accurately. The Project mainly focused on combining demographic information like age, gender, and medical history with their heart rhythm recordings to improve prediction. The project also works with image-based ECG signals and establishes a separate pipeline to convert ECG image-based reports into data that computers can understand. Once trained, these machine learning models can be deployed on the cloud service, and then seamlessly integrated into user friendly web interface. The web interface makes it easy to upload bulk ECG data and provides essential features like user authentication, keep track of patient information, and sees past predictions. The inclusion of a medical chatbot within web interface assists users to interpret results and offer valuable prescriptions, medical guidance, and diet plans. The mobile app also helps users to upload their image-based reports and see the predictions. In essence, This project represents a significant step forward in advancing heart disease prediction methodologies and offering scalable and user-friendly solutions.

# Introduction

Early detection of heart disease is important and beneficial for heart patients to get help before any serious circumstances. ECG signal data provide useful insights into heart health. These insights can be used for the prediction of heart diseases in the early stages [1]. CardioGraph Pro is specially trained to predict diseases on such data using machine learning techniques to achieve better health treatment of heart patients [2].

This project mainly focusses to establish a robust data pipeline that have potential to integrate demographic data and Electrocardiogram signals together [3]. The system will take raw data that have both type of information and turn into that format which can be easy to use for analysis and prediction. The project will use the PhysioNet's Database (WFDB) for ECG signals data [4]. Then that data converts into CSV files for further processing.

After organising data and preprocessing it, the project will use machine learning techniques to give predictions about heart diseases based on ECG signals [5]. The model is tested thoroughly to make sure that it will predict accurately and efficiently [6]. The project will also explore image-based ECG signal processing, developing a separate pipeline to convert ECG images into analysable signals [7]. These image-based signals will undergo the same predictive modelling process as conventional ECG data, enriching the predictive capabilities of the models.

Upon successful training and validation, the trained machine learning model will be deployed on cloud service for scalability and accessibility. And it will relate to a website to use. This way the doctor and the regular person can upload their heart rhythm data to get predictions [8]. The website will even have a chatbot to help explain the result nicely This project also consists of a mobile app for those who can simply upload their ECG image to predict results quickly [9].

# Success Criterion

* Achieve a minimum accuracy threshold (e.g., 85% or higher) in predicting heart disease when using both ECG signals and demographic data.
* Compare this with models using ECG signals alone to determine if incorporating demographic information improves performance.
* The model on the cloud can handle a lot of ECG tests being uploaded at once without having any issues.
* The user finds the interface intuitive, and effective and should find it easy to use and understand the results, prescriptions, and diet plans that it gives to them.
* The chatbot explains the result correctly and gives advice based on medical guidelines.
* The system may be able to find heart disease earlier than normal tests doctors use now.
* Compare when the system predicts heart disease versus when a doctor normally finds it in the same group of patients. Track if early prediction leads to better outcomes.

# Related Work

[10] This study investigates the analysis of ECG signals for the purpose of detecting cardiovascular illness using a convolutional neural network (CNN). This research paper presents a Deep Neural Network model that was trained over the Two million labelled tests that were gathered as part of the Clinical Outcomes in Digital Electro cardiology study and assessed by the Telehealth Network of Minas Gerais.  With F1 scores is above 80% and specificity almost over 99% for the diagnosis of atrial fibrillation or normal rhythm.

[11] This study investigates the classification of Electrocardiogram signals using a Radial Basis Function (RBF) Neural Network for the detection of cardiac disease. The data used in this study report came from a PTB-XL database. The researcher proposed the three neural network architecture approaches the first one based on the convolution network the second on the SincNet and the third was mix of first type with additional entropy-based features. Comparably, the network built on top of SincNet recognized two classes with 85.8% ACC and five classes with 73.0% ACC. 89.82% ACC was attained by the network that based on the convolutional network with entropy features when two classes were recognized, and 76.5% ACC when five classes were recognized.

[12] This paper tried a hybrid deep learning model for heart disease prediction using ECG signals. It combines conventional ECG beat extraction with a Long Short-Term Memory (LSTM) network. This article's goal was to present a novel approach to categorizing heart disease from Electrocardiogram data. They developed a system that maintains critical heart wave data while enabling automation. CNN's automatic feature learning produces an accurate representation of heart function through adaptive heartbeat segmentation. It contributes to a decrease in incorrect classifications. The FSN approach provides a more accurate and dependable feature set for the classification of cardiac diseases since the features vector is created using Convolution neural network and QRS complex features. The experimental findings demonstrate that the proposed model outperforms the previous deep learning-based techniques. To determine how dependable the performance of the suggested model is, we advise examining further data sets in the future.

[13] In this study, they provide a novel approach that combines deep learning modelling, signal processing techniques, and cardiology to predict cardiac illnesses using ECG signals. To turn 1-D signals into 2-D images, they used wavelet transformation, which enables deep learning models.to examine the properties of the signal's various frequency bands concurrently. Their system is at the cutting edge and is quick and easy to use. Owing to their system's outstanding performance in the four ECG signal classification challenges, we anticipate that this approach will be applicable to increasingly challenging tasks. We'll test this with bigger datasets that have more detailed annotations. Furthermore, our method performs differently because different wavelets can extract different frequency characteristics utilizing the wavelet transform. We'll try to apply this model on other wavelets to process signals.

[14] This project aims to create algorithmic models for the analysis of ECG tracings to forecast cardiovascular illnesses. This project helps save lives and improve health care quality while losing cost. Improving medical care and saving lives are the immediate results of this study, which is particularly relevant as global health care and insurance prices rise. To optimize deep-learning parameters, we carried out several tests. Two different machine learning algorithms (VGG16 and MobileNetV2) tested, and both were performed very accurate to gives 95% result. When the model is put on Raspberry Pi, they are still accurate but slightly less do 94% for MobileNetV2 and 90 % for VGG16he main goal of this project is to improve how we can keep track of the health information in real time using smart devices.

# Project Rationale

Heart disease poses a critical global health threat, necessitating advanced methods for early detection and intervention. Traditional diagnostics often lead to delays in diagnosis, limiting treatment options and impacting patient outcomes. This project aims to develop a comprehensive heart disease prediction system leveraging electrocardiogram (ECG) signals and demographic data. This combination has the potential to improve prediction accuracy compared to existing methods. The development of image-based ECG processing pipelines will broaden input sources and enhance accessibility for diverse users. The motivation behind this process is to develop a more accurate and user-accessible solution.

Through conducting this research and developing this solution, our main learning will be a depth understanding of how to train test and deploy machine learning models along with the integration process for web and mobile app. One of our main learnings will be understanding data preprocessing, data pipeline designing and working with the imaging data. Also, we will learn how to deploy and maintain a machine learning model to keep it working for the user on our website and mobile app.

This project is driven by the need for accessible and user-friendly prediction solutions. A cloud-based machine learning model that integrates with a web interface and a dedicated mobile application will prioritize widespread access. The integration of a medical chatbot will empower users by offering real-time interpretation of results and guidance on the next steps. Ultimately, this project aims to introduce early detection tools for heart disease management. The potential impact includes improving patient outcomes and helping them to seek better treatment early.

## Aims and Objectives

The primary aim of this project is to develop an advanced system for heart disease prediction that integrates ECG signals analysis and image ECG image-based report processing and provides a user-friendly interface to get accurate predictions about heart diseases.

Making it helpful for the doctors to get predicted disease labels along with ECG reports for several patients in a very sophisticated and rebuts way. It also aims to help a patient having any ECG report to get know about his disease, get prescriptions, and diet plans accordingly also integrates a chat bot to ask questions about the disease or answer any concern of the patient.

### Objectives

* Design and construct the data pipeline that combines ECG signals and demographic data from the PhysioNet WFDB database.
* Make the pipeline that converts ECG images into the data (signals) that the model can understand.
* Train test and optimize the machine learning model to accurately and precisely predict heart disease based on ECG signals and image-based reports.
* Training a model to give prescriptions based on the predicted disease and have diet plan recommendations.
* Deploy the optimized prediction model on a cloud platform to ensure scalability and accessibility.
* Build the website and app where the user can upload their heart data, safely log in manage their record, and see medical history.
* Include the medical chatbot (only for the website). that can explain the result answer the question and give them basic guidelines.

## Scope of the Project

The scope of this project includes:

* Designing a data pipeline and developing a preprocessing module to acquire data and convert it into a suitable format for applying machine learning techniques.
* Develop a machine learning based heart disease prediction system to accurately predict disease from ECG signal data or ECG image-based reports.
* A fully functional cloud powered application solution accessible for doctors and patients to process ECGs and get predictions, prescriptions, and assistance from the integrated medical chatbot fine-tuned for cardio related diseases.

# Proposed Methodology and Architecture

In this project a systematic methodology will be used leveraging advanced machine learning techniques including data collection, preparation, feature extraction, model training, and validation. The architectural design consists of key components such as data pipeline, machine learning models, cloud services integration and user interface development.

## Methodology

### Data Acquisition and Preprocessing

The Project begins by collecting the ECG and demographic data from the PhysioNet WFDB database. The system cleans and organizes the collected data. This includes steps like noise filtering, single segmentation by dividing the signals into smaller parts and then feature extraction (finding the most important insights within the data)

### Data Integration

This step represents the integration of ECG signal and demographic data. Then combined them into a format suitable for a machine learning model that easily understands and uses either training or making predictions.

### Image Processing

The system also allows image-based report input. These image-based reports will be converted to signals form through some image processing techniques to be compatible with the data collected in the form of signals from Waveform Database.

### Machine Learning Model

The system has two phases training and testing. In the training phase, it uses the combined dataset to learn the machine learning model to find the pattern linked to heart disease. This involves splitting the data and practising with some of it. And testing itself on the rest. In the prediction phase, the system uses the trained model to analyse the new person's data and predict if they have a risk of heart disease

### Results and User Interface

The system takes the predictions made by the machine learning model and prepares them to show the user on web and mobile devices. Users can see past predictions and enter the new data to get another prediction. Finally, the result is shown and the chatbot for a better understanding of the result and medical guidance.

## Architecture

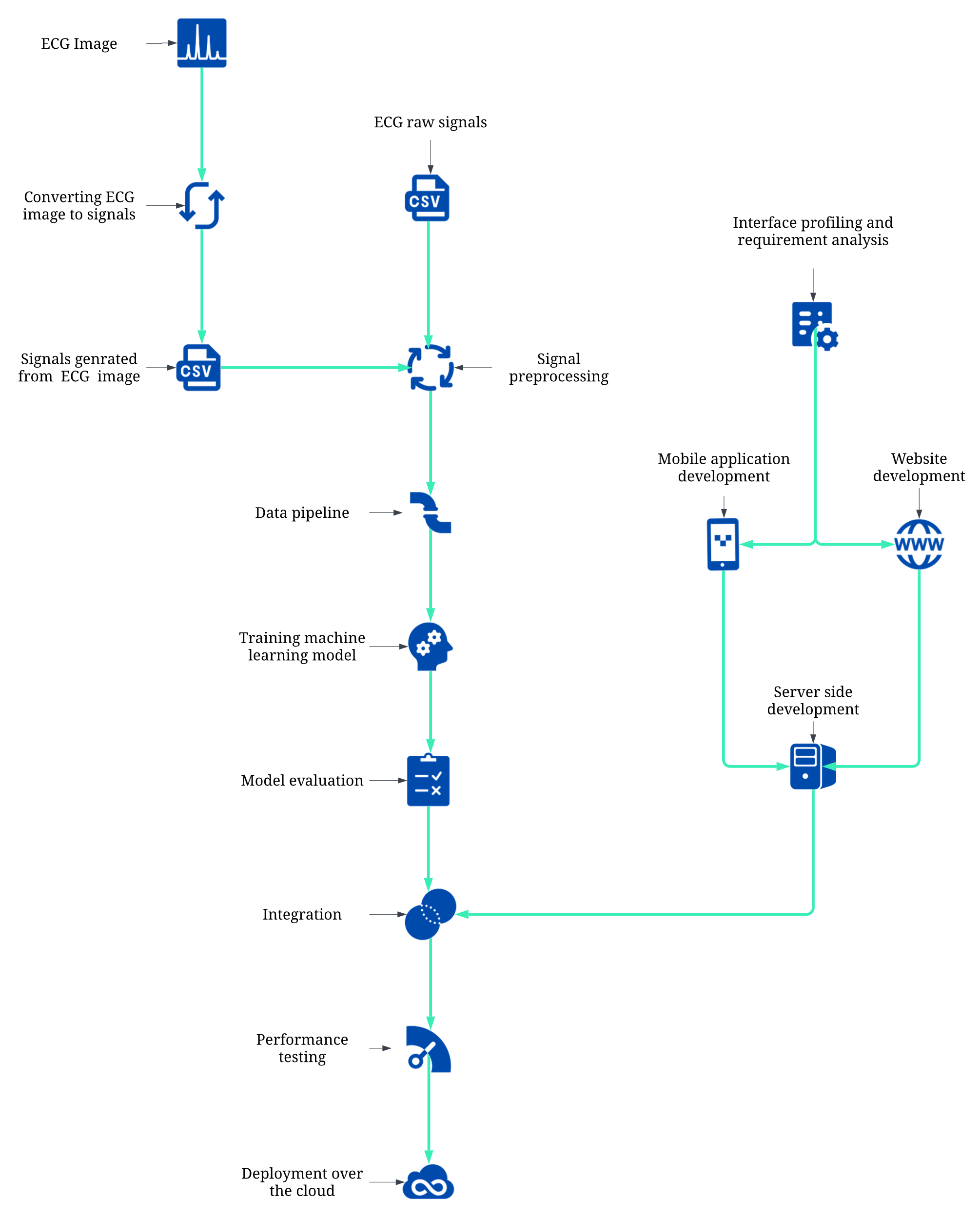


Figure 1: Architecture

## Application Working Flowchart

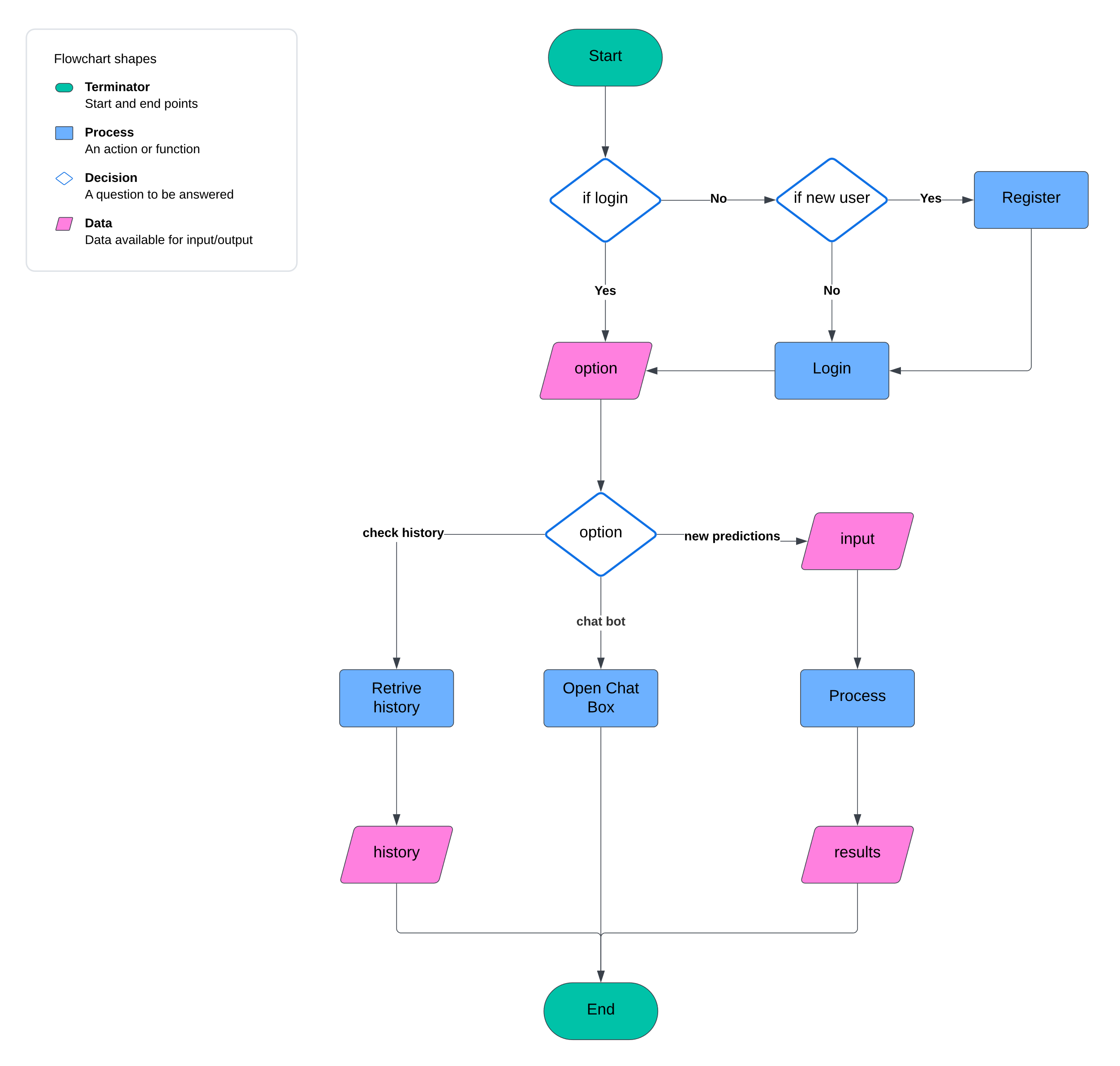


Figure 2: Application Working Flowchart

# Individual Tasks

Table 1: Individual Tasks

|  |  |  |
| --- | --- | --- |
| **Team Members** | **Activity** | **Tentative Date** |
| Member 1, Member 2, Member 3 | Project Planning and Research | 11 March 2024 – 07 April 2024 |
| Member 2, Member 3 | Literature Review | 15 March 2024 – 07 April 2024 |
| Member 1, Member 3 | Dataset Collection | 20 March 2024 – 29 March 2024 |
| Member 3, Member 1 | Data Preprocessing and Pipeline Designing | 01 April 2024 – 12 April 2024 |
| Member 1 | Interface Profiling and Requirement Analysis | 11 March 2024 – 12 April 2024 |
| Member 1, Member 2, Member 3 | Model Development and Training | 15 April 2024 – 10 May 2024 |
| Member 3 | Website Development | 15 April 2024 – 10 May 2024 |
| Member 1, Member 2 | Model Evaluation and Validation | 13 May 2024 – 24 May 2024 |
| Member 1, Member 2, Member 3 | Integration and Testing | 13 May 2024 – 24 May 2024 |
| Member 1 | Mobile App Development | 24 June 2024 – 09 Aug 2024 |
| Member 3, Member 1 | Model Optimization and Enhancement | 24 July 2024 – 06 Sep 2024 |
| Member 1, Member 2, Member 3 | Interface Refinements | 11 Aug 2024 – 06 Sep 2024 |
| Member 3, Member 1 | Backend Development and Integration | 11 Aug 2024 - 09 Nov 2024 |
| Member 1, Member 3 | Deployment Process | 11 Nov 2024 – 30 Dec 2024 |
| Member 1, Member 2, Member 3 | Documentation and Reporting | 11 March 2024 – 30 Dec 2024 |

# Gantt Chart

Table 2: Gannt Chart

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | |  |  |  | | --- | --- | --- | | Semester 7 | Summer Break | Semester8 | |
| |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | |
| Project Planning and Research | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Literature Review | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Dataset Collection | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Data Preprocessing and Pipeline Designing | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Interface Profiling and Requirement Analysis | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Model Development and Training | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Website Development | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Model Evaluation and Validation | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Integration and Testing | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Mobile App Development | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Model Optimization and Enhancement | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Interface Refinements | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Backend Development and Integration | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Deployment Process | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |
| Documentation and Reporting | |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  | |

# Tools and Technologies

We have carefully selected different tools and technologies to make sure that we successfully achieve our goals and develop and implement a complete solution to our chosen problem.

## Technologies

### Machine Learning Model

* **Python:** Primary language for implementing ML models and web frameworks.
* **TensorFlow:** Building and training neural networks and deep learning models.
* **Keras:** Used with TensorFlow for rapid prototyping of high level ANNs.
* **scikit-learn:** Library for handling, classifications, and clustering problems.
* **OpenCV:** Library providing tools for image processing and manipulation.
* **NumPy:** Providing support to other required libraries.
* **Pandas:** EDA and data pipeline designing of feeding data to ML model.
* **Matplotlib:** Visualizing ML model’s performance and confusion matrix.



Figure 3: Technologies - Machine Learning Model

### Interface Design and Development

* **HTML 5:** Assisting in structuring our web pages.
* **CSS 3:** Style sheets to give a robust look to the user interfaces.
* **JavaScript:** Making applications interactive to enhance user experience.
* **Bootstrap 5:** Ensure responsive design and quick prototyping.
* **Flask:** Quickly develops a web server in Python dealing with ML models.
* **Pillow:** Processing images on the Python server for the ML model.
* **React Native:** Building a mobile app for engaging more users.
* **React:** Making a fully functional web application for the project.
* **Django:** Hosts and manages the complete server for web and mobile.



Figure 4: Technologies - Interface Design and Development

### Databases

* **MS SQL Server:** Storing our model training data after pipeline and preprocessing.
* **MongoDB:** Keeping user demographics and medical history for applications.



Figure 5: Technologies - Databases

## Tools

### Integrated Development Environments

* **Google Collab:** Cloud based notebooks for Model development and collaborations.
* **Visual Studio Code:** Lightweight and versatile IDE for app and web development.
* **Jupyter Notebook:** Interactive code writing environment for EDA and prototyping.
* **Anaconda:** Facilitates Python virtual environments and robust library management.



Figure 6: Tools - Integrated Development Environments

### Deployment

* **Microsoft Azure:** Facilitates ML model deployment and server less app hosting.
* **Docker:** Streamlines application management across different environments.



Figure 7: Tools - Deployment

### Others

* **Git and GitHub:** Enables code management, sharing and team to collaborate.
* **SSMS:** SQL Server Management Studio to manage SQL database server.
* **MongoDB Compass:** Efficiently managing MongoDB from the desktop.



Figure 8: Tools - Others