A Project Report

On

Automate Diagnosis & Consultants (ADC)

submitted for partial fulfillment of the requirements.

for the award of the degree of

Bachelor of Technology

in

Computer Science

Submitted by

Ankit Yadav (1900290120016)
Ayush Srivastava (1900290120029)

Under the Supervision of

Dr. Harsh Khatter
Assistant Professor
Department of Computer Science

KIET Group of Institutions, Ghaziabad
Dr. A.P.J. Abdul Kalam Technical University
2022-2023

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written byanother person nor material that to a substantial extent has been accepted for theaward of any other degree or diploma of the university or other institute of higherlearning, except where due acknowledgment has been made in the text.

Signature of Students
Name: Ankit Yadav
Roll No.: 1900290120016

Name: Ayush Srivastava Roll
No.: 1900290120029

Date: 11/05/2023

CERTIFICATE

This is to certify that Project Report entitled "**Automate Diagnosis & Consultant"** which is submitted by **Ankit Yadav and Ayush Srivastava** in partial fulfillment of the requirement for the award of degree B. Tech. in the Department of Computer Science of Dr A.P.J.Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date: 11/05/2023

Supervisor Signature
Dr. Harsh Khatter
Assistant Professor
Department of Computer Science

ACKNOWLEDGEMENT

It gives us great pleasure to present the synopsis of the B. Tech Mini Project undertaken during B.Tech. Fourth Year. We owe a special debt of gratitude to Dr. Harsh Khatter, Assistant Professor, Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness, and perseverance have been a constant source of inspiration for us. It is only his/her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Shrivastava, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the department's faculty members for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:
Name: Ankit Yadav
Roll No: 1900290120016

Name: Ayush Srivastava
Roll No.:1900290120029

Date: 11/05/2023

ABSTRACT

A sizable segment of the world's population lacks access to quality healthcare. The success of healthcare ultimately depends on the doctor's skill. In this study, we investigate if this knowledge may be represented as an information corpus, oras data that has been retrieved using data mining methods, particularly the Machine Learning & Deep Learning Model, to make a diagnosis. When the medical diagnosis is made widely available, coverage increases and life quality improves. To determine whether inferences about the causes of various diseases can be made from the data, this project provides an overview of machine learningapproaches used in the classification of various diseases. We outline a few of ourfindings from the trials we ran before offering some suggestions for the future.

The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures, monitoring systems are designed to lower health care expenditures. Using the data mining modeling technique, the integration of clinical decision support with computer-based patient records coulddecrease medical errors, increase patient safety, stop unwelcome practice variance, and improve practice outcomes. Connecting patients and doctors through a user-friendly interface will make it easier for patients to use in emergency situations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Introduction:

The health problem is the gap between acceptable or desirable health status and the present status. In today's life, health problems occurred more than last 25-30 years ago because of modernization, and industrialization. Suddenly, changes in the environment directly affect a health condition. So, more health-related problems are occurring day by day and require daily checkups on health conditions. This project is based on real-time implementation as well as more informative and realistic it can be highly used in Medical Diagnosis and understandingof the different feasibilities of the model.

The growth of ML has facilitated the increased use of computers in medicine. There are three primary goals of data mining that tend to be prediction, description, and presentation. Predictioninvolves some variables or fields in the dataset to predict unknown or future values of other variables of interest. The description focuses on finding patterns describing the data that can be interpreted by humans, on the other hand Presentation plays an important role to be easily understandable to humans.

Health issues are more prevalent today than they were 25 to 30 years ago due to urbanization and industrialization. Unexpected changes in the environment have a direct impact on health. As a result, more health issues emerge daily, necessitating daily health condition checks. This project, which is based on real-time implementation and is more informative and realistic, can be very useful in helping to diagnose medical conditions.

The goal of creating automated systems is to decrease the amount of time and money spent on health care by lowering the number of doctor visits, hospital stays, and diagnostic testing procedures. The second main method is supervised learning, which uses machine learning techniques to learn a function from a collection of training data. Prediction and generalization are two key requirements for supervised learning algorithm performance. The trained function ought to be capable of accurately predicting the results for data that are not included in the training set. It should also serve as a model that generalizes to new data points and captures the underlying traits of the training data.

1.2 Problem Statement:

In such unprecedented times, with a fragile healthcare infrastructure, the medical facilities for people are scarce and even fewer for people with special needs. The healthcare sector got so preoccupied with Covid-19 cases that people with other ailments didn't even get a chance to avail medical assistance. So, we are here to present a model to mitigate the scarcity of facilities & the lack of medical management.

- Scarcity of facilities and the lack of medical management.
- There is no website for Automated Diseases Prediction System.
- Decentralized Patient Records are not available.

- No platform Integrates doctors and patients via disease prediction with relevant doctorsuggestions.

1.3 Field of Invention:

The technology related to the general field of healthcare and management, i.e., machine learning-based automated medical diagnosis to enhance healthcare infrastructure. The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures,monitoring systems are designed to lower healthcare expenditures.

1.4 Objectives:

The objective of developing an automated system is to changing the way to presenting anddetection of any diagnosis by

- Reduce health care costs by reducing physical hospital visits. hospitalizations, anddiagnostic testing procedures.
- Integrate doctors with patients' disease prediction with relevant doctor suggestions.
- Automated Diseases Prediction System to make facilities available easily.

1.5 Scope:

Setting up an alarm alert system & adding more diseases to our model increases our feasibility& compatibility with the market.
Adding more body parts models that increase people's knowledge that would lead to trainingpurposes for the nurses.

1.6 Advantages:

- Integration of clinical decision support with computer-based patient records could reduce medical errors and enhance patient safety.
- Unwanted practice variance is eliminated, and practice outcomes are improved, both ofwhich can greatly raise the standard of clinical judgment.
- Reduce disease identification time and start treatment expeditiously.
- Reduce the cost of medical tests & by providing initial diagnostics in time which reduceshuge amounts of the burden on the patient's family.
- Decentralized Patient Records Led to data being more accessible & makes the fastdiagnosis of patients.
- Automated Diseases Prediction System leads to predicting disease based on symptoms & recommending doctors & also book Path Labs on it.
- Integrate doctors and patients via an interactive and user-friendly interface so thatPatients can easily utilize it in emergency situations.
- We also research pharmaceutical equipment & Medicines so that patients can get puremedicines from anywhere.

CHAPTER 2 REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

2.1 Feasibility Study (Technical, Economical, Operational)

The proposed project aims to develop an automated medical diagnosis system using machine learning techniques that will reduce healthcare costs by lowering physical hospital visits, hospitalizations, and diagnostic testing procedures. It will also integrate doctors with patients' disease predictions and suggestions, make facilities easily available, and reduce disease identification time.

The feasibility of the project can be evaluated in terms of technical feasibility, economic feasibility, legal feasibility, and operational feasibility.

Technical Feasibility:
The project requires the development of an automated medical diagnosis system using machine learning algorithms, which is technically feasible. There are several tools and frameworks available for developing machine learning models, such as Python's Scikit-learn, TensorFlow, Keras, etc.

Economic Feasibility:
The project's economic feasibility depends on several factors such as development cost, operating cost, and revenue generation potential. The project's development cost includes hardware and software infrastructure, development team salaries, and training costs. The operating cost includes maintenance costs, server costs, and cloud services costs. Revenue generation potential could come from subscription-based models, advertisements, or commission-based fees on doctor consultations or pathology tests booked through the platform.

Legal Feasibility:
The project should comply with local regulations and laws related to healthcare, patient privacy, and data protection. It should also adhere to ethical guidelines related to patient data collection, storage, and usage.

Operational Feasibility:
The project's operational feasibility depends on the user's ability to use the platform easily and the availability of healthcare professionals to respond to patient queries and concerns. It also requires seamless integration with existing healthcare systems to enable access to patient records and avoidduplication of data.

Overall, the project appears to be technically feasible, economically feasible, and legally feasible. However, the operational feasibility may require further testing and development to ensure that it is user-friendly and compatible with existing healthcare systems.

2.2 Software Requirement Specification Document

Introduction

The purpose of this SRS document is to specify the software requirements for an automated diagnosis and consultant web application. This document will outline the data, functional, performance, maintainability, and security requirements for the application.

Data Requirements

The following data requirements must be met by the application:

- The ability to collect and store patient health data, including medical history, symptoms,and test results
- The ability to integrate with existing electronic health record (EHR) systems or other datasources
- The ability to protect patient data and comply with HIPAA regulations

Functional Requirements

The following functional requirements must be met by the application:

- The ability to perform automated diagnosis and consultation based on patient data andmachine learning algorithms
- The ability to provide personalized treatment recommendations based on the diagnosis
- The ability to integrate with healthcare providers and allow for secure communicationbetween providers and patients
- The ability to support multiple languages and accommodate patients from diverse backgrounds

Performance Requirements

The following performance requirements must be met by the application:

- The ability to handle a high volume of patient data and requests
- The ability to provide accurate diagnosis and consultation results within a reasonable timeframe
- The ability to maintain high availability and minimize downtime

Maintainability Requirements

The following maintainability requirements must be met by the application:

- The ability to quickly identify and resolve software bugs or issues
- The ability to perform software updates and upgrades with minimal disruption to users
- The ability to provide documentation and training for healthcare providers and patients

Security Requirements

The following security requirements must be met by the application:

- The ability to protect patient data and comply with HIPAA regulations
- The ability to prevent unauthorized access to the application and patient data
- The ability to implement secure authentication and communication protocols betweenproviders and patients

Conclusion
This SRS document has outlined the data, functional, performance, maintainability, and security requirements for an automated diagnosis and consultant web application. These requirements mustbe met in order to ensure the success and viability of the application.

2.3 Prototype Model To Be Used in Automate diagnosis & consultantWeb Application

The Prototype model is an iterative approach to software development that involves building a working prototype of the application to gather feedback and refine the requirements before developing the final product. Here's an example of how the Prototype model could be implementedfor an automated diagnosis and consultant web application:

1-Requirements Gathering:
The first step in the Prototype model is to gather and analyze the requirements for the web application. This involves identifying the needs of the healthcare providers and patients, the data sources to be integrated, and the features and functionalities of the application.

2-Initial Prototype Development:
The next step is to develop an initial prototype of the web application that includes the core features and functionalities. This prototype should be a working model of the application that can be tested and evaluated by the stakeholders.

3-Evaluation and Feedback:
Once the initial prototype is developed, it should be evaluated and tested by the stakeholders, including the healthcare providers and patients. Feedback should be gathered from the stakeholders regarding the prototype's functionality, usability, and user interface.

4-Prototype Refinement:
Based on the feedback received, the prototype should be refined and updated to address any issuesor concerns raised by the stakeholders. This may involve adding new features, modifying existing ones, or changing the user interface.

Repeat Steps 3 and 4:
The evaluation and feedback loop should be repeated until the stakeholders are satisfied with the prototype's functionality and usability. The prototype should be refined and updated based on the feedback received until it meets all the stakeholders' needs.

Final Product Development:
Once the stakeholders are satisfied with the prototype, the final product can be developed based on the refined prototype's specifications. The final product should include all the features and functionalities of the prototype and should be thoroughly tested before deployment.

In the case of an automated diagnosis and consultant web application, the Prototype model could be implemented by building an initial prototype that includes the core features such as patient data collection, machine learning algorithms for diagnosis, and personalized treatment recommendations. The prototype could then be evaluated and tested by healthcare providers and patients, and feedback could be gathered to refine and improve the prototype's functionality and usability. The prototype refinement process could involve adding new features such as secure communication between providers and patients, accommodating patients from diverse backgrounds, and integrating with electronic health record (EHR) systems. Once the stakeholders are satisfied with the prototype, the final product could be developed based on the refinedprototype's specifications, and thoroughly tested before deployment

CHAPTER 3 SYSTEM DESIGN

3.1 Detail Design

System design for an automated diagnosis and consultant web application involves defining the architecture, components, and technical specifications of the system. Here are the key components of the system design for the web application:

Architecture:
The system architecture defines the overall structure of the web application and how its components interact with each other. The web application could be designed using a client-server architecture, where the client sends requests to the server, and the server processes therequests and sends back the responses.

User Interface:
The user interface is the graphical interface that allows users to interact with the web application. The user interface should be designed to be intuitive and user-friendly, and it should provide easy access to all the features and functionalities of the web application.

Database:
The database is the backbone of the web application, and it stores all the patient data, diagnosis, and treatment recommendations. The database should be designed to be scalable,secure, and easily accessible to authorized users.

Machine Learning Algorithms:
Machine learning algorithms are the heart of the web application, and they are responsible for analyzing patient data and making accurate diagnoses and treatment recommendations. The machine learning algorithms should be trained on large datasets andshould be capable of handling complex and diverse patient data.

Communication:
The web application should provide secure and reliable communication channels between healthcare providers and patients. This could involve integrating video conferencing, messaging, and other communication tools into the web application.

Security:
The web application should be designed with security in mind, and it should be equipped withrobust security measures to protect patient data and prevent unauthorized access. This could involve implementing encryption, user authentication, and access control mechanisms.

Performance:
The web application should be designed to be highly performant, with fast response times andminimal downtime. This could involve optimizing database queries, using caching mechanisms, and scaling the application horizontally.

Integration:
The web application should be designed to integrate with other healthcare systems such as electronic health record (EHR) systems, billing systems, and pharmacy systems.

In summary, the system design for an automated diagnosis and consultant web application involves defining the architecture, user interface, database, machine learning algorithms, communication channels, security measures, performance optimization, and integration withother healthcare systems. The design should be scalable, secure, reliable, and user-friendly, with a focus on providing accurate and personalized diagnosis and treatment recommendations to patients.

3.2 UML Diagrams

Various embodiments of the invention are disclosed in the following detailed description & accompanying drawings:

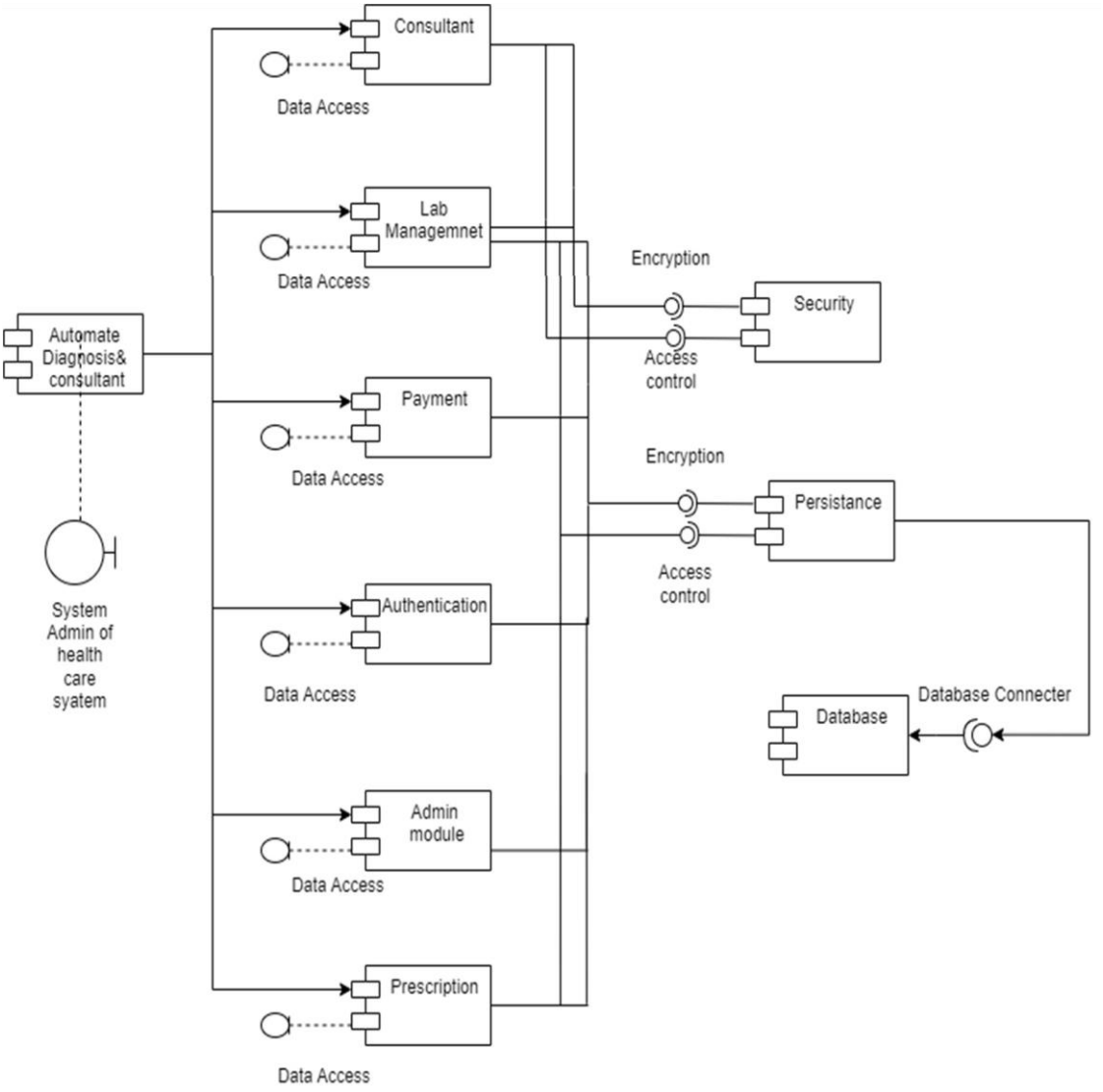

Fig 3.1: Component Diagram

The diagnostic component: This component would be responsible for analysing patient data, such as symptoms and medical history, and making a diagnosis based on that data. It may include machine learning algorithms, databases of medical knowledge, and other tools to help make accurate diagnoses.

The consultation component: This component would be responsible for providing personalized recommendations to the patient based on their diagnosis. It may include resources such as treatment plans, medication recommendations, and other advice to help the patient manage their condition.

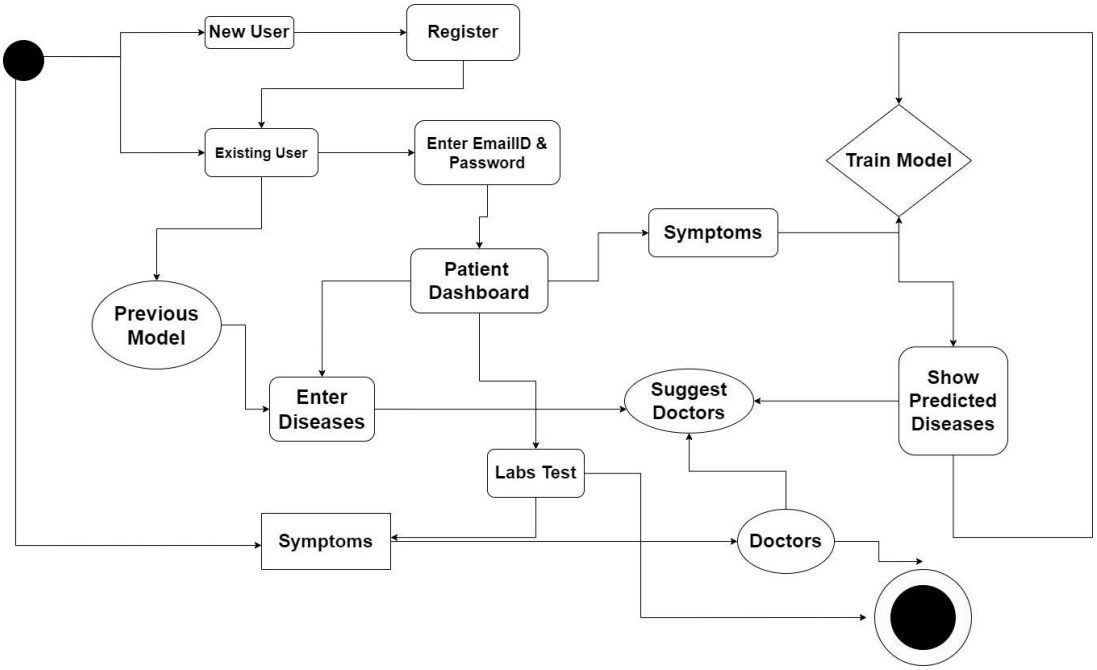

Fig 3.2: Activity Diagram (Path Labs module).

The activity diagram starts with the "Start" node, which is followed by the "Enter patient and test details" node. In this step, the user would enter information such as the patient's name, the type of test being conducted, and any other relevant details.
The next step is "Collect sample and conduct test", which involves collecting a sample from the patient and performing the necessary lab tests. Once the test results are available, the user records them and uploads them to the system, as shown in the "Record test results and upload" node. Finally, the activity diagram ends with the "End" node, indicating that the process is complete.

Throughout the activity diagram, there may be decision points or branching paths depending on the specific requirements of the system. For example, there may be additional steps required for certain types of tests or for patients with specific medical conditions. These decision points could be represented in the activity diagram using diamond-shaped nodes with multiple outgoing arrows.

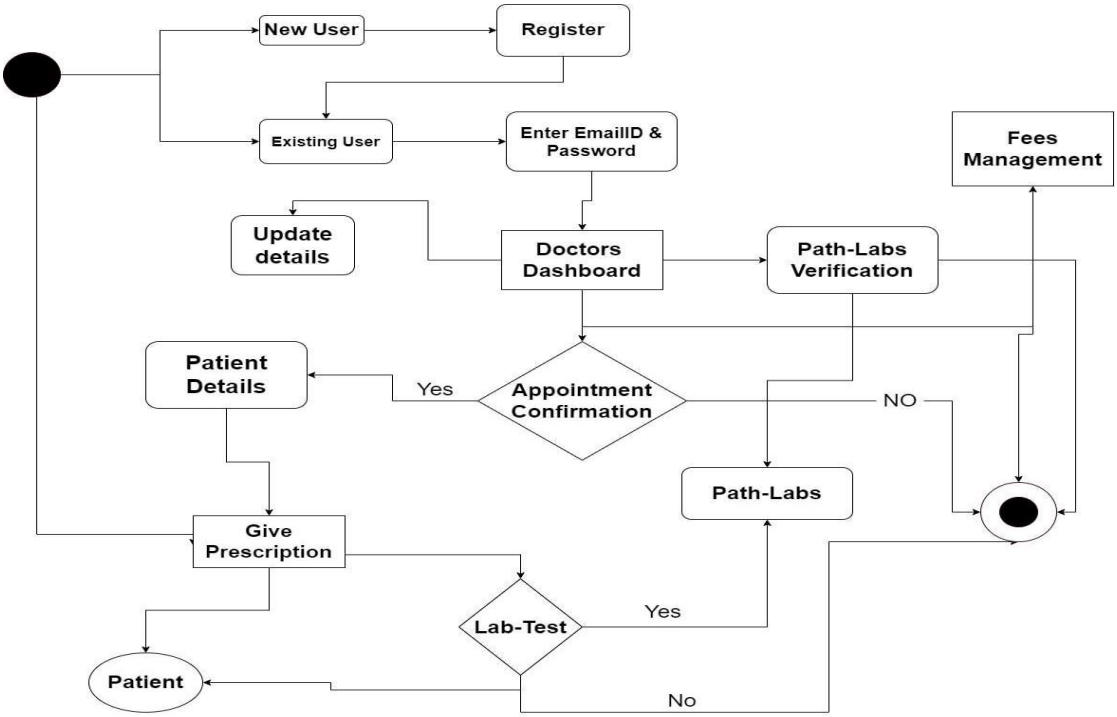

Fig 3.3: Activity Diagram (Doctors module).

The activity diagram starts with the "Start" node, which is followed by the "Retrieve patient symptoms/history" node. In this step, the doctor retrieves information about the patient's medical history, symptoms, and any other relevant information.

The next step is "Perform Diagnosis", which involves using the patient information to make a diagnosis. This step may include using machine learning algorithms, databases of medical knowledge, or other tools to help make an accurate diagnosis.

The following step is "Check for Previous Diagnoses", which involves checking the patient's medical history to see if they have any previous diagnoses or medical conditions that could impact their current diagnosis and treatment.

Once the diagnosis is complete, the doctor prescribes medication, as shown in the "Prescribe Medication" node. This step may involve selecting appropriate medications based on the patient's condition, checking for potential drug interactions, and providing instructions for taking the medication.

Finally, the activity diagram ends with the "End" node, indicating that the process is complete.

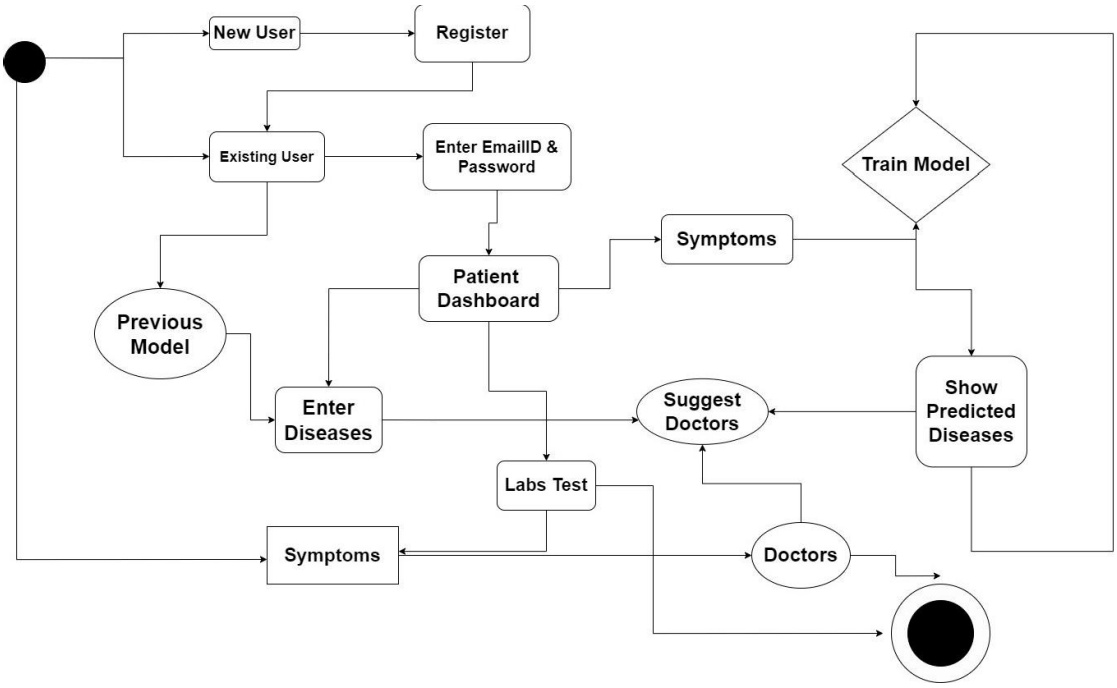

Fig 3.4: Activity Diagram (Patients module).

The activity diagram starts with the "Start" node, which is followed by the "Login to patient account" node. In this step, the patient logs in to their account in the automated diagnosis system.

The next step is "View medical history and records", which allows the patient to access their medical history and records stored in the system.

The following step is "Schedule an appointment", which allows the patient to schedule an appointment with a healthcare provider. This step may involve selecting a provider, choosing a date and time for the appointment, and providing any necessary information about the reason for the appointment.

Once the appointment is scheduled, the patient can complete a patient assessment, as shown in the "Complete patient assessment" node. This step may involve answering questions about their medical history, current symptoms, and other relevant information.

The healthcare provider uses the information provided by the patient to make a diagnosis and provide advice, as shown in the "Receive a diagnosis and advice" node.

Finally, the activity diagram ends with the "End" node, indicating that the process is complete.

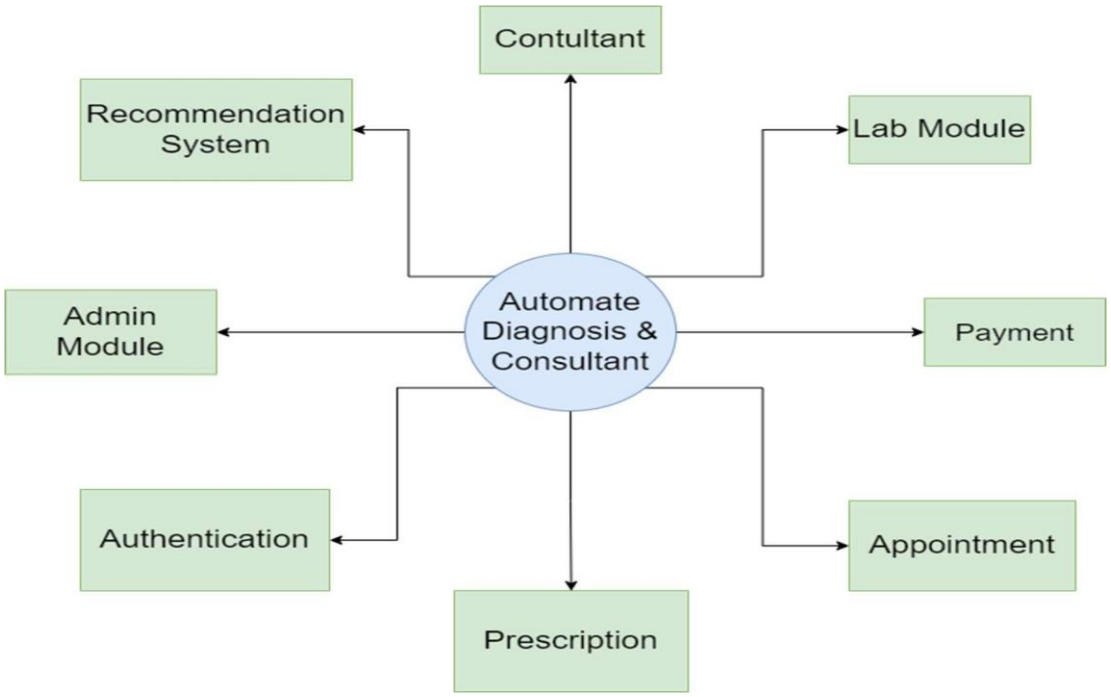

Fig 3.5: Data Flow Diagram (Level -0)

The level 0 DFD has three main components:

External Entities: These are the individuals or groups who interact with the system. In this case, the external entities are patients, healthcare providers, and lab workers. They provide input to the system and receive output from it.

System: This is the main component of the system. It includes the automated diagnosis and consultant software, as well as any hardware or other tools needed to operate the system.

Outputs: These are the results generated by the system. They include diagnoses, consultations, lab results, and medication recommendations.

The arrows in the diagram show the flow of data between the components of the system. External entities provide input to the system, which is processed by the automated diagnosis and consultant software. The output from the system is then delivered to the external entities, who use it to make healthcare decisions.

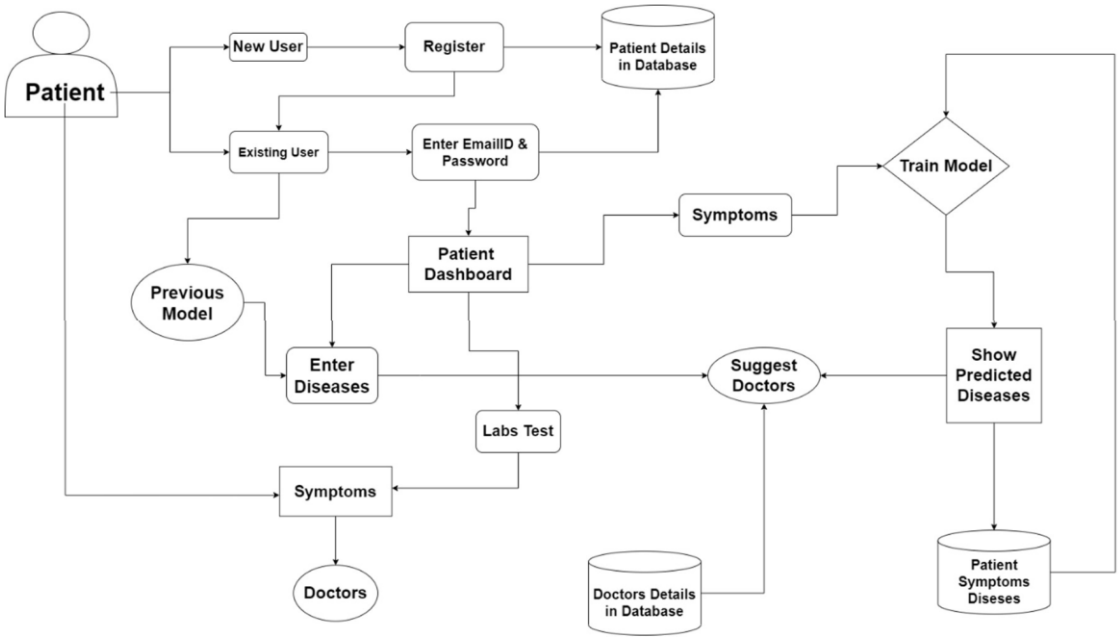

Fig 3.6: Data Flow Diagram (Level -1 Patient Module)

The Level 1 DFD of Patient Module has three main components:

External Entity: The patient, who interacts with the system.

Processes: These are the specific processes involved in the patient's interactions with the system. They include logging in to the system, viewing their medical records, scheduling an appointment with a healthcare provider, completing a patient assessment, and receiving a diagnosis.

Data Stores: The patient records database, which stores the patient's medical records and other relevant information.

The arrows in the diagram show the flow of data between the components of the system. The patient provides input to the system through various processes, and the system uses the data stores to store and retrieve information as needed.

Overall, the level 1 DFD provides a more detailed view of the patient module in the automated diagnosis and consultant system. Further levels of the DFD would break down the specific components of each process and provide more detail about the specific interactions involved.

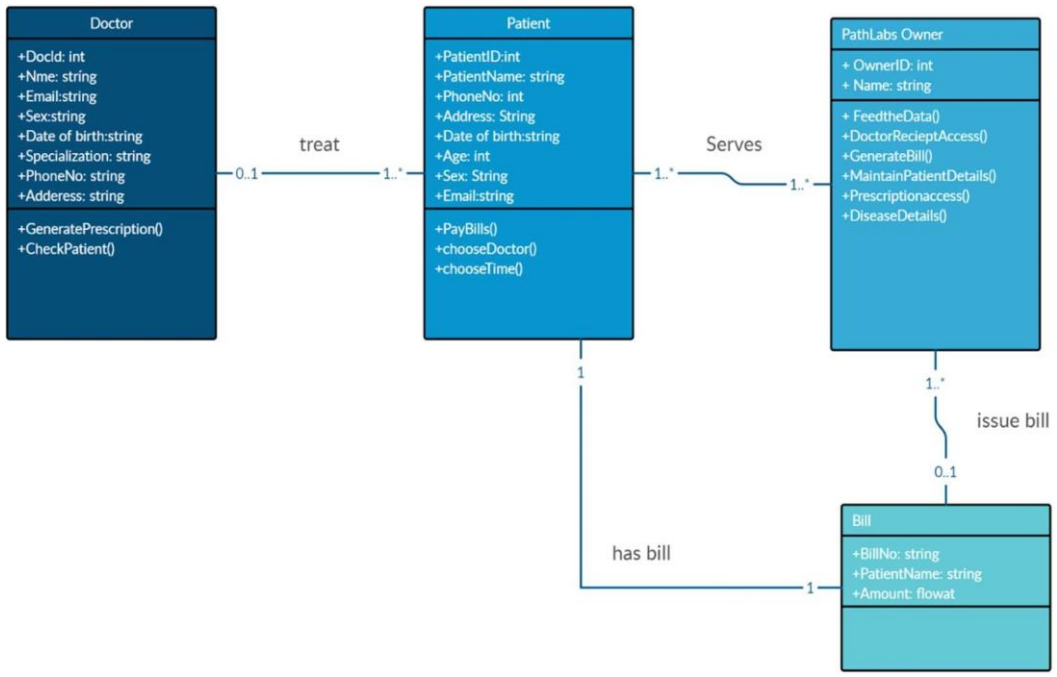

Fig 3.7: Entity Relationship Diagram

The ERD has four main entities:

Patient: Stores information about the patient, including their name, date of birth, and gender.

Visit: Represents a patient visit to a healthcare provider, and includes information about the visit date, symptoms, and associated diagnosis.

Diagnosis: Stores information about diagnoses, including the diagnosis name and description.

Healthcare Provider: Stores information about healthcare providers, including their name and specialty.

There is also an entity for Lab Results, which stores information about laboratory test results associated with a particular visit.

The arrows in the ERD show the relationships between the entities. For example, the Visit entity has a foreign key that references the Patient entity, indicating that each visit is associated with a specific patient. Similarly, the Diagnosis entity has a foreign key that references the Visit entity, indicating that each diagnosis is associated with a specific patient visit.

Overall, the ERD provides a visual representation of the relationships between the entities in the automated diagnosis and consultant system, which is useful for understanding how patient data and healthcare information are stored and managed.

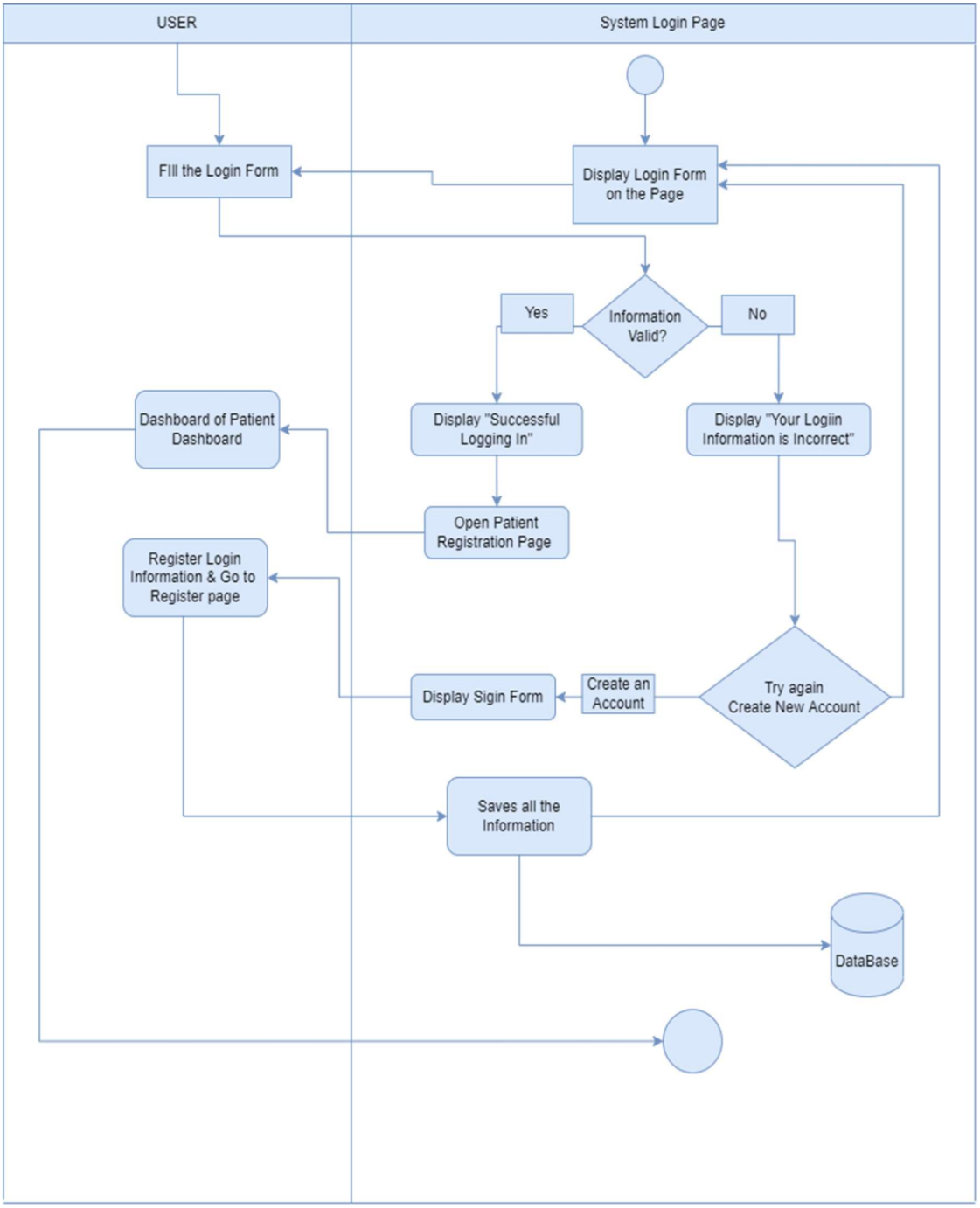

Fig 3.8: Login Module

The login module is a critical component of the automated diagnosis and consultant system, as it ensures that only authorized users are granted access to the system. To ensure the security of the system, it is important to use strong password policies, employ multi-factor authentication mechanisms, and periodically review and update user account information in the user database.

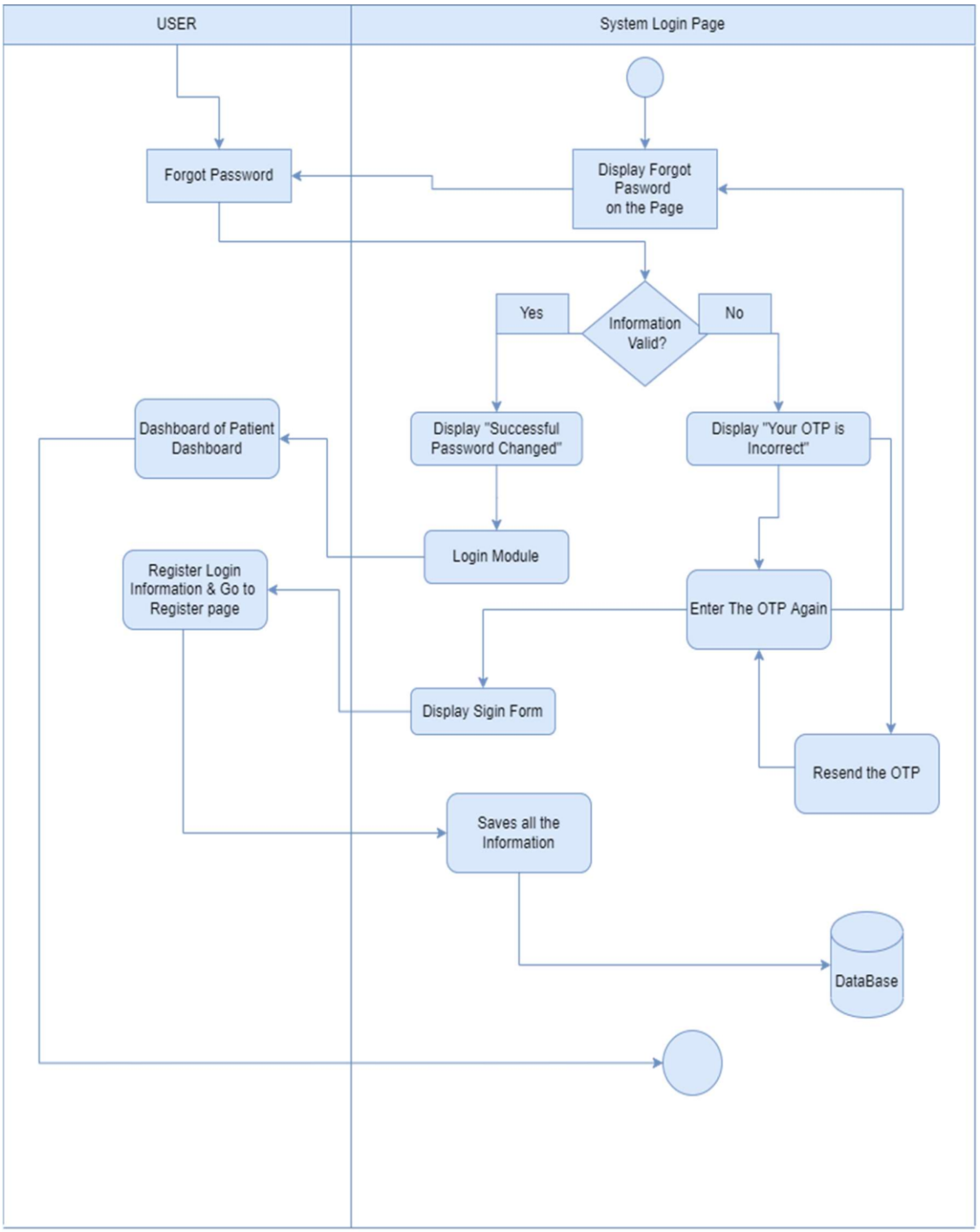

Fig 3.9: Forgot Password Module

The "Forgot password" module is an important feature of the automated diagnosis and consultant system, as it helps users recover their password in case, they forget it or if their account is compromised. To ensure the security of the system, it is important to use secure password reset links, enforce password policies, and monitor password reset activity for suspicious behavior

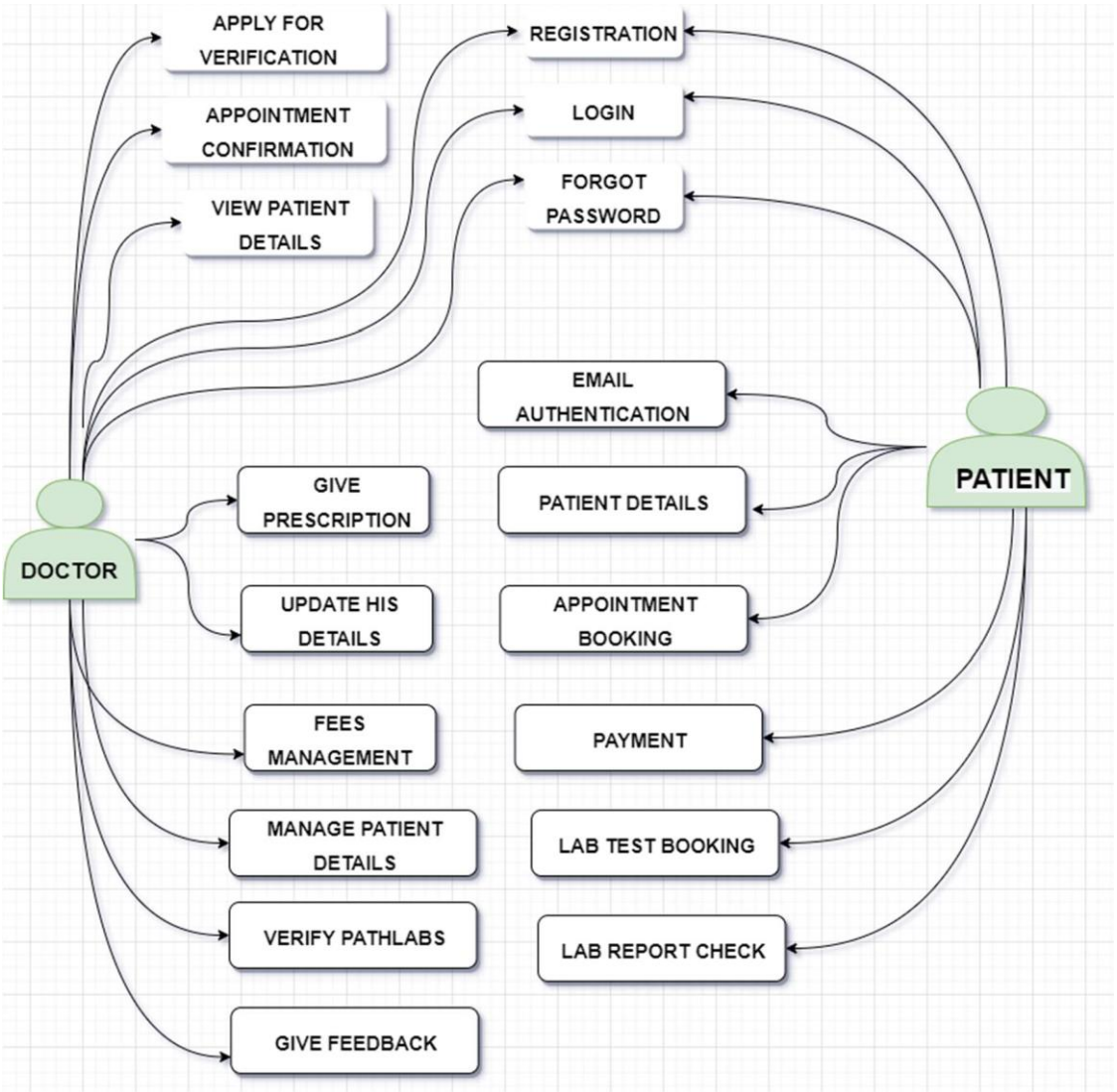

Fig 3.10: Use Case Diagram of doctor & patient Module

In the doctor module, there are three use cases:

View Patient: This use case allows the doctor to view the details of a patient, such as their medical history, test results, and other relevant information.
Add Diagnosis: This use case allows the doctor to add a diagnosis for a patient based on the patient's symptoms and test results.
View Reports: This use case allows the doctor to view the reports generated by the system based on the patient's diagnosis and test results.

In the patient module, there are two use cases:

Book Appointment: This use case allows the patient to book an appointment with a doctor based on their availability and specialty.
View Reports: This use case allows the patient to view their reports generated by the system based on their diagnosis and test results

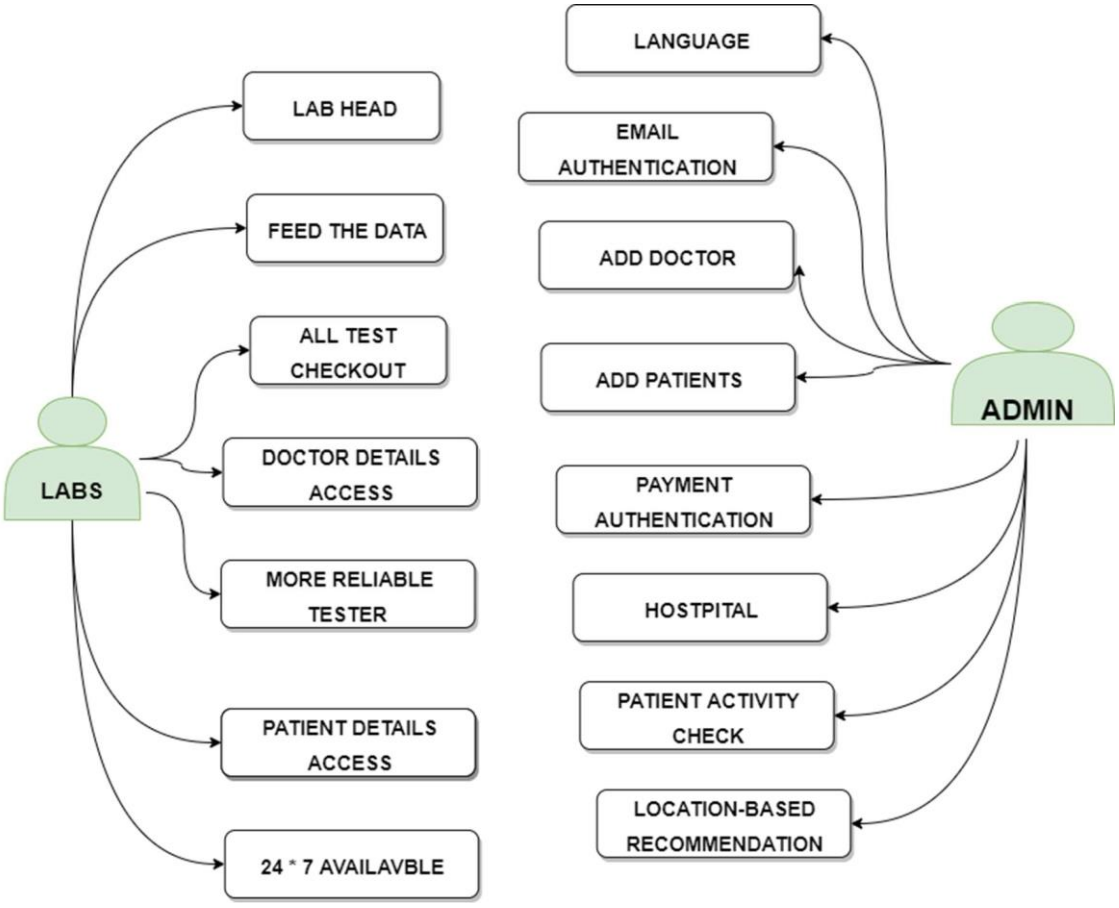

Fig 3.11: Use Case Diagram of Path labs & Admin Module

In the admin module, there are two use cases:

Manage Users: This use case allows the admin to manage the user accounts in the system, including creating new accounts, modifying existing accounts, and deleting accounts if necessary.

Manage Appointments: This use case allows the admin to manage the appointments scheduled in the system, including viewing, editing, and cancelling appointments.

In the path labs module, there are two use cases:

Upload Test Results: This use case allows the path labs to upload the test results of a patient into the system, which can be accessed by the doctor or patient.

View Test Results: This use case allows the path labs to view the test results that have been uploaded by the lab and are associated with a specific patient.

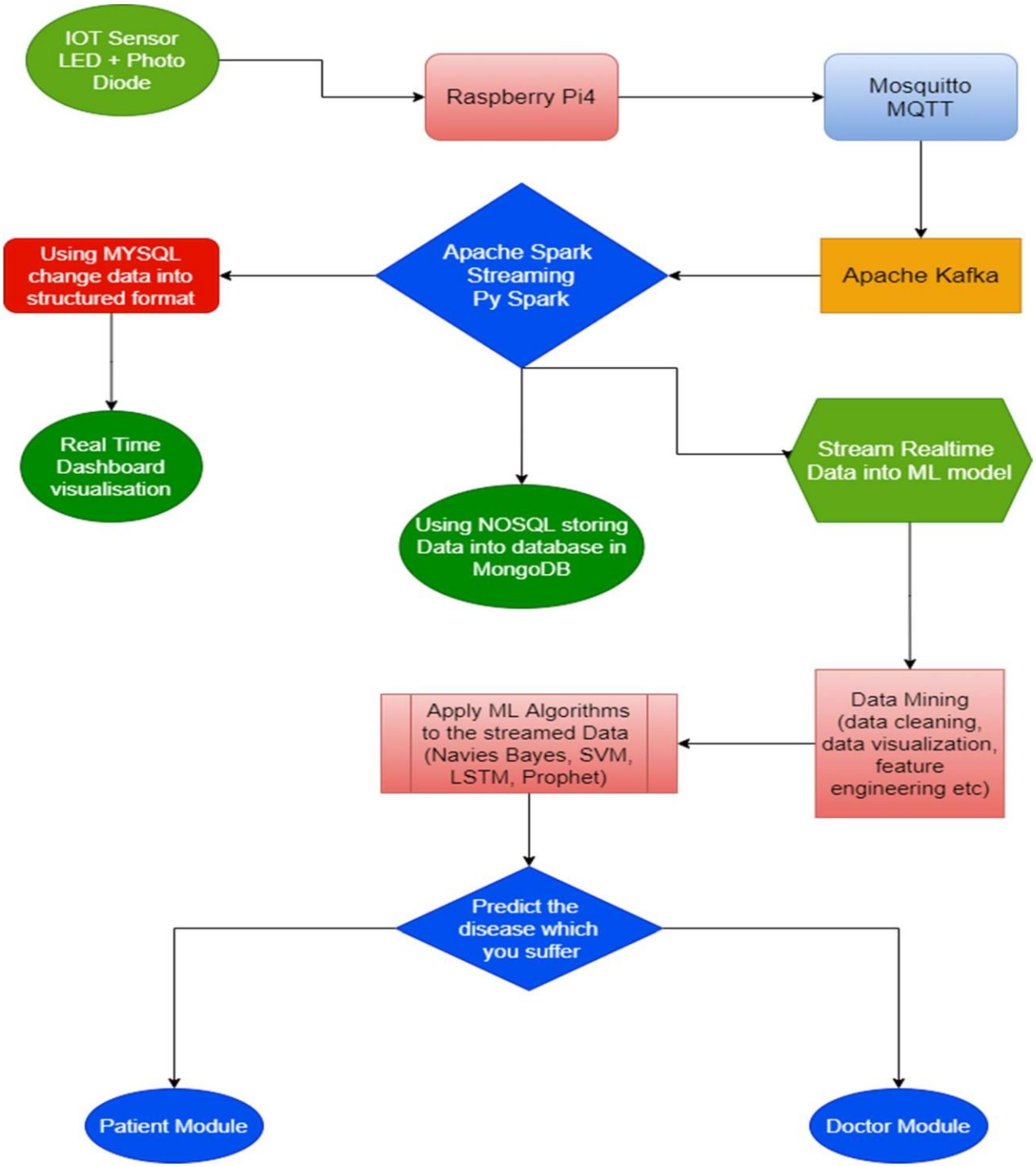

Fig 3.12: Data Streaming Model

The data streaming module of Automate diagnosis & consultant involves the real-time streaming of data from various sources to the system for analysis and processing. This module plays a critical role in the system's ability to provide quick and accurate diagnoses and recommendations. Here are some of the key components and processes involved in the data streaming module:

Data Sources: The data streaming module must be able to receive data from various sources, including medical devices, electronic health records, wearables, and other sources. These sources may provide data in different formats and require different protocols for communication.

Data Preprocessing: Once the data is received, it must be preprocessed to ensure that it is in the correct format for analysis. This may involve cleaning the data, removing duplicates, and ensuring that the data is consistent across different sources.

Data Integration: The data streaming module must be able to integrate data from multiple sources to provide a complete picture of the patient's health status. This may involve merging data from different devices or systems and ensuring that the data is synchronized and up-to-date.

Real-Time Streaming: The data streaming module must be able to receive data in real-time and process it quickly to provide immediate insights and recommendations. This requires the use of streaming technologies such as Apache Kafka, Apache Flink, or AWS Kinesis.

Data Analysis: Once the data is received and preprocessed, it must be analyzed using machine learning algorithms to provide accurate diagnoses and recommendations. This may involve training machine learning models on large datasets to improve the accuracy of the analysis.

Visualization and Reporting: The results of the data analysis must be presented to healthcare professionals and patients in a way that is easy to understand. This may involve creating visualizations and reports that highlight key insights and trends.

Overall, the data streaming module is essential to the success of Automate diagnosis & consultant, as it enables the system to process and analyze large amounts of data in real-time to provide accurate diagnoses and recommendation.

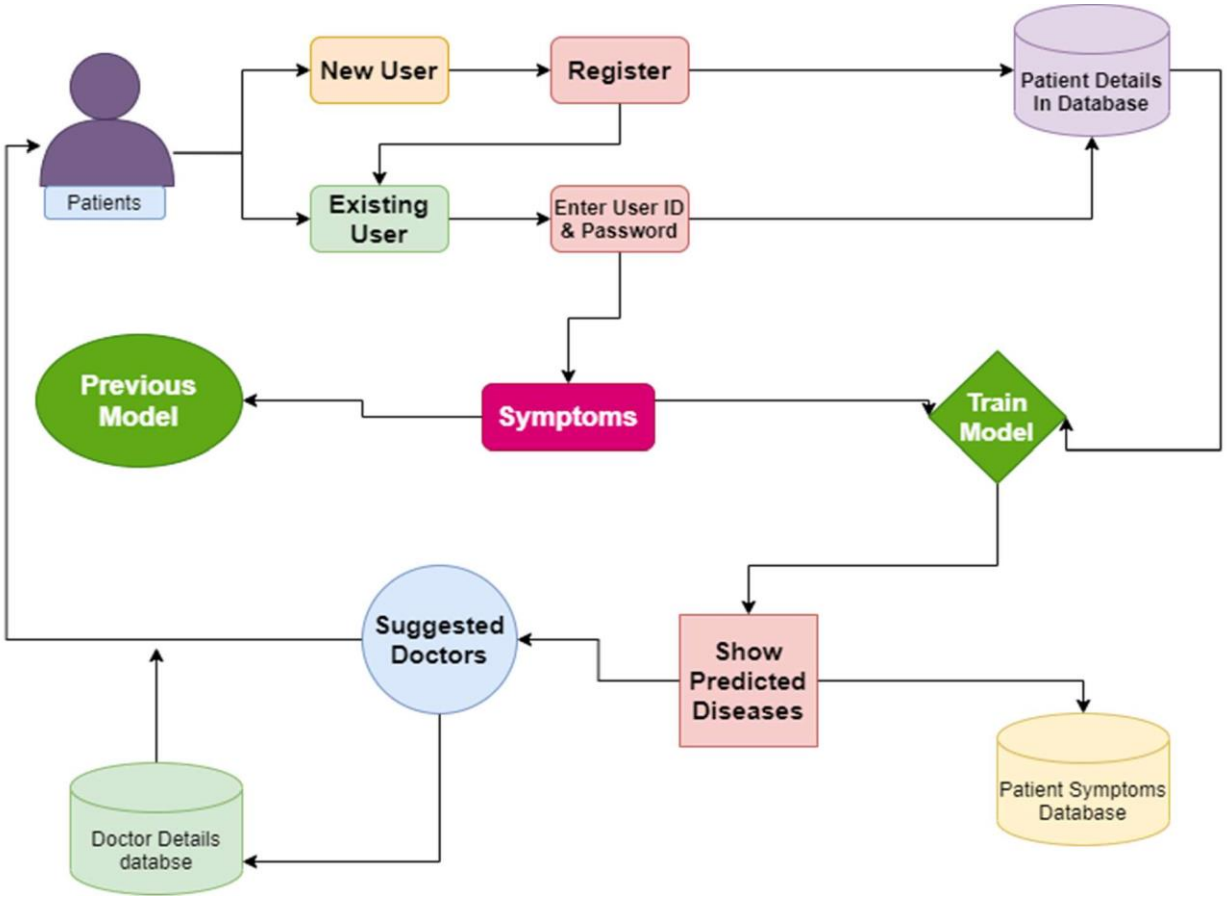

Fig 3.13: Use case model of the Frontend link with ML model.

The use case model shows the different use cases or functionalities that are available when the frontend is linked with an ML model.
Upload Image: This use case allows the user to upload an image of a medical scan or test for analysis.

Preprocess Image: This use case involves preprocessing the image to prepare it for analysis by the ML model. This may involve resizing, normalization, or other techniques.

Predict Disease: This use case involves using the ML model to analyze the preprocessed image and predict the presence of a particular disease or condition.

View Prediction Result: This use case allows the user to view the prediction result generated by the ML model, such as a diagnosis or probability of a specific disease.

These use cases are essential for the frontend with an ML model to function properly and provide the necessary services to the users.

CHAPTER 4 IMPLEMENTATION

The methodology's goal is to forecast an individual's risk of developing kidney, lung, and breastcancer, heart disease, and diabetes using a few questions and machine learning models in an end-to-end procedure. the system I used for my research has the following software and systemconfigurations: On an Intel(R) Core(TM) i5-2310M GPU @1650Ti with 8 GB RAM, Jupyter Notebook 5.5.0 and VS Code 1.73 are used to implement Python 3 and the Flask framework.

Figure 4.1 displays a block schematic of the fundamental procedures used for each machine-learning model. To transform the raw data into a form that can be used, data cleaning is done first. Data analysis is carried out after data cleansing to ascertain the significance of characteristics. Once a goal has been established, it is time to start gathering the data required for analysis. Your data team will be responsible with cleaning and sifting through the data onceit has been gathered from all the required sources. Because not all data is good data, data cleansing is crucial during the data analysis process. Data mining, which is referred to as "knowledge discovery inside databases," is one method. In order to predict what will probably happen next in the future, predictive analysis looks ahead to the future. These methods are a component of inferential statistics, which is the act of examining statistical data in order to make inferences about the connections between various sets of data. We have to utilize the Train/Test methodology to evaluate the model's performance. The train/test approach is a way to gauge how accurate your model is. Because you divide the data set into two sets, a training set and a testing set, the method is known as Train/Test. 20% for testing, 80% for training. Using the training set, you train the model. Utilizing the testing set, you test the model. Create the model by training it. To test a model is to determine its correctness. After confirming that our model is sound, we can now begin making fresh value predictions.

Integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decease unwanted practice variation and improve practice outcome which can help significantly improve the quality of clinical decisions using the data mining modelling technique & Integrate doctors and patients via interacting and user- friendly interface so that Patient can easily utilize in emergency situations. There are three primary goals of this model tend to be prediction, description, and presentation.

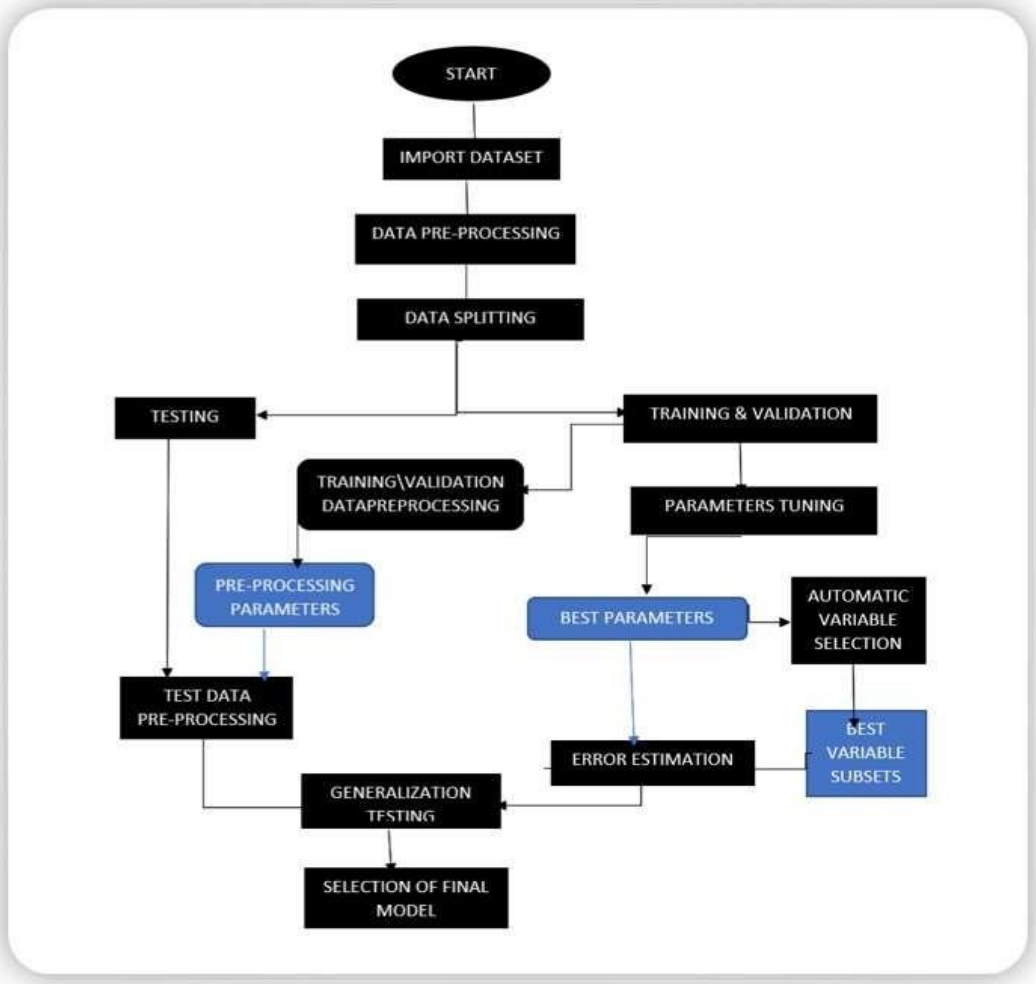

Fig 4.1: Machine Learning Algorithm's flow chart

4.1 Algorithm for the proposed system:

Random Forest Classifier Algorithm:
A group of several decision trees is called a random forest. Decision trees are employed as parallel estimators in the bagging technique, which is used to construct random forests. Whenused in a classification issue, the outcome is determined by the majority vote of the findings from each decision tree. In a regression, the mean value of something like the target values ina leaf node serves as the prediction. The mean value of the decision tree outcomes is considered by random forest regression.

Logistic Regression:
A set of independent variables is used to estimate discrete values (often binary values like 0/1) using logistic regression. By fitting data to a logit function, it aids in the prediction of anevent's likelihood. Logistic regression is another name for it. The following techniques are frequently used to enhance logistic regression models: incorporate interaction terms, removefeatures, regularize approaches, and employ a non-linear mode.

Decision Tree:
The Decision Tree method, a supervised learning technique used for issue classification, isone of the most widely used algorithms in machine learning today. Both continuous and

categorical dependent variables may be classified using it. Based on the most important characteristics/independent variables, this method splits the population into two or more homogenous groupings.

Step 1: Gathering and identifying the data that will be provided to the network as input is thefirst stage in the categorization of diabetes or heart disease using Logistic Regression, Decision Tree & Random Forest.
Step 2: The network receives a defined training dataset and the selected training method. After the training phase, the Logistic Regression, Decision Tree & Random Forest are alsoput to the test to get feedback on how well they categorize the condition.

Category	Algorithm	Function	Use
Basic Regression	Logistic Regression	linear_model.LogisticRegression()	The target variable is categorical.
Classification	Decision Tree	tree.DecisionTreeClassifier()	If/then/else. Non-contiguousdata. can also be regression.
Classification	Random Forest	ensemble.RandomForestClassifier()	Find the best split randomly.can also be regression.

Table 4.1

Step 3: The "training data set" is used to train the classifier, the "validation set" is used to fine-tune the parameters, and the "test data set" is used to evaluate the performance of the classifier. It's vital to keep in mind that only the training and/or validation set is available when the classifier is being trained. The test data set must not be used for classifier training.Only when the classifier is being tested will the test set be accessible.

Step 4: We may then create a confusion matrix, which indicates how effectively our model has been trained after this is complete. True Positives, True Negatives, False Positives, and False Negatives are the four parameters that make up a confusion matrix. In order to create a model that is more accurate, we would want to obtain more data for the true negatives and true positives. The number of classes has a direct impact on the size of the confusion matrix.

Step 5: The model creation process includes a step called model evaluation. Finding the model that best depicts our data and predicts how well the model will perform in the future is helpful. In order to enhance the model, we may adjust its hyper-parameters to attempt to raise its accuracy while also looking at the confusion matrix to try to increase the proportion of truepositives and true negatives.

4.2 Proposed System of Automate Diagnosis & Consultant:

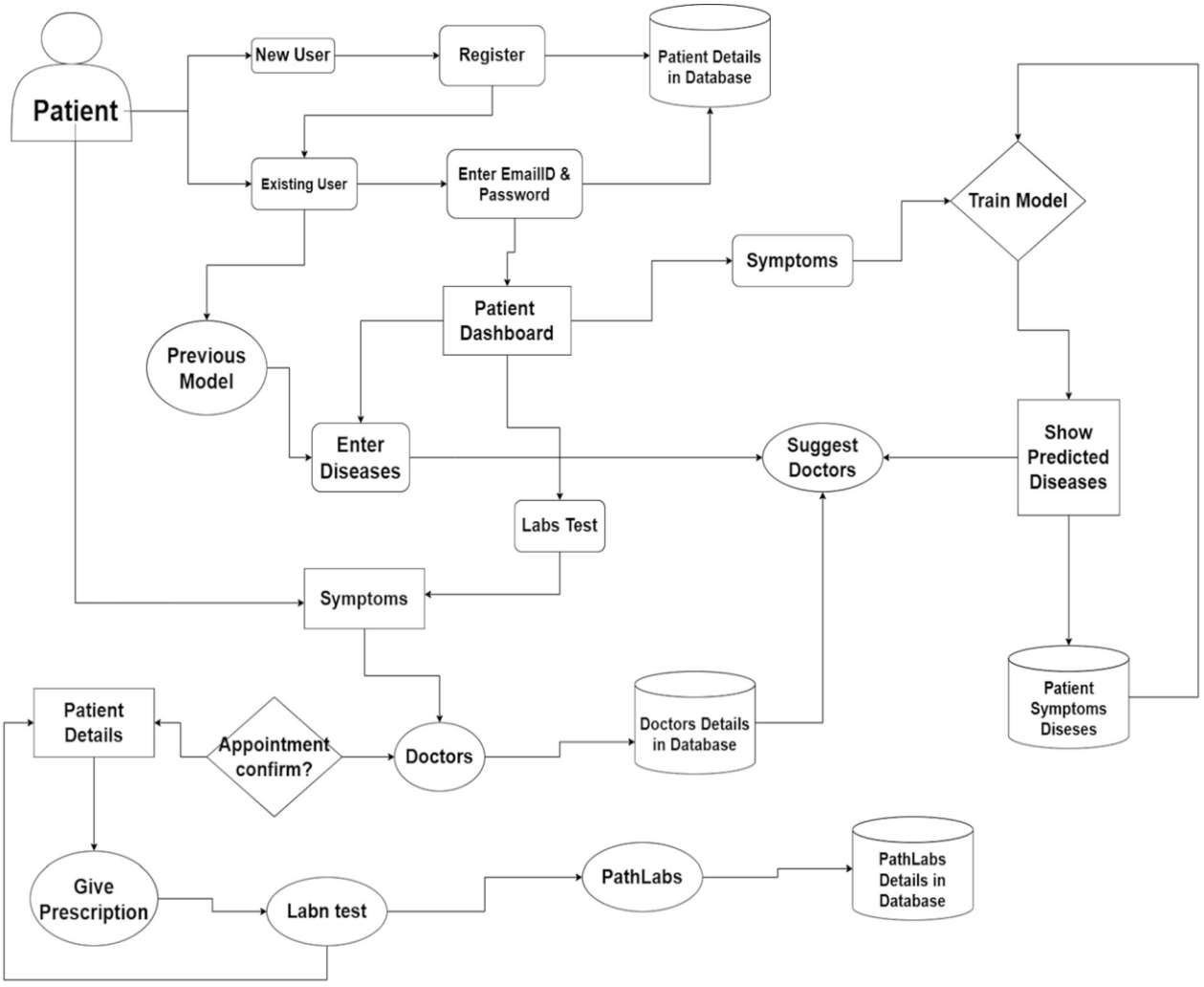

Fig 4.2: System flow Diagram

The dashboard will run the function that connects patients and doctors. The dashboard has beendesigned to get a better understanding of a patient's condition. The medical profession is being transformed by technological advancements, which is causing a rapid change in the medical business. Artificial intelligence (AI) is being used to diagnose illnesses, predict diseases, and develop treatments. In this article, it will be discussed how AI can support the management of level-three 911 calls and assist physicians in making better patient decisions. According to the doctors' available time slots, those who have appointments can show up for their appointments.To avoid any unneeded waiting times, this is done. Modern medical professionals are employing cutting-edge technology to provide more precise diagnoses and individualized treatment strategies. Doctors can now give patients extensive medical data, an in-depth report on their symptoms, and personalized care plans that can be seen on a mobile device thanks to artificial intelligence (AI). With the use of an app, the doctor may view the patient's information, write him a prescription, and suggest testing. This software

assists in quickly diagnosing the patient since it has all the information required to understand the ailment, its symptoms, and recommended treatments. The outdated paper lab slip is being replaced with a modern dashboard. Patients will be able to schedule an appointment at a time that works for them because these slips are now digital. As soon as the reports are finished being processedby the laboratory, they will also be automatically uploaded into the patient's database. Doctors use information from a variety of sources to provide the most precise diagnosis possible. With the more recent use of electronic medical records, a more complete image of patients is now possible, improving care.

- **Scalability:**
  - The primary objective of a SWOT analysis is to help organizations develop a full awareness of all the factors involved in making a business decision.
  - Strength: Small (4-5 employees) can change and adapt quickly & gives a unique advantage.
  - Weakness: Team members who are not much trained in both technical and professional engineering. No one has been through any such type of technical training programs.
  - Opportunity: Market trends & Economic trends will take us to new heights.
  - Threats: Political, environmental, economic regulations & Funding part is our main threat.

- **How Realistic to Achieve/Make?**
  - It can be easily implementable because in this pandemic period most of peoples are technically advanced & they are mostly relying on the internet.
  - It can be widely spread on different social media platforms, Application based platforms like google play store, Appstore etc.
  - It can be easily accessed & modified.

- **Any Threat/Risk/Problem that you can foresee?**

  Bad Design, Poor user Experience, Slopy implementation, feature creep, & lack ofquality control all contribute to product failure.

4.3 Implementation activity:

Disease diagnosis using Machine Learning & Deep Learning Models

- This project focused on many health issues from which not only India but the whole Worldis suffering. Visualize the dataset & using All ML & DL algorithms to diagnose Diseases.
- Dataset used: Kaggle Cardiovascular, diabetes, Kidney, Breast Cancer, malaria etc. dataset
- Project GitHub Link: https://github.com/ankiii07/Automate-Diagnosis

- **Economic Sustainability:**
    - We first go personally to many hospitals for our product launching & later with some remote areas so that they use our application & get more & more benefits from it from anywhere.
    - At Initial Stage we just put more & more efforts to setup in market not on generating revenue.
    - It is not for generating revenue but later after this product get popular and providemore & more benefits to people, then we take charges for using this application which is affordable to our middle order people.

- **Environment Sustainability:**
    - The idea is much innovative and not implemented with such mixed technology so for which made the people to their comfort and so guide them to reach their desire.
    - Apart from this it is an initiative which has not been seen so far and it has Less development and maintenance cost and high profit values.
    - This Project is Environment Sustainable & much helpful to Remote areas public,rural area public & it does not cause any environmental damage.

4.4 Technology Used

- Web Development
    - **Frontend**: HTML, CSS, JavaScript.
    - **Backend**: NodeJs, ExpressJs MongoDB.
- Machine Learning & Deep Learning
    - Anaconda, Jupyter Notebook, EDA, Data Collection, Cleaning, Visualisation, NumPy, Pandas, Seaborn, SciPy, Matplotlib, Pandas, Sklearn Libraries, ML Algorithms
- Augmented Reality
    - Unity, Vuforia SDK, AR Core, A Frame, JDK, Augmented Reality integrated with Machine Learning
- Blockchain
    - MetaMask, Web3.Js,

CHAPTER 5. RESULTS AND DISCUSSIONS

5.1 User Interface Representation

Fig 5.1: User Interface of project

Online medical consultations can be convenient for patients who are unable to visit a healthcare provider in person due to geographical or mobility constraints. However, it's important to note that online consultations may not be suitable for all medical issues, and in some cases, an in-person examination may be necessary to properly diagnose and treat a condition.

5.2 Snapshots of Disease Prediction Module with brief detail of each

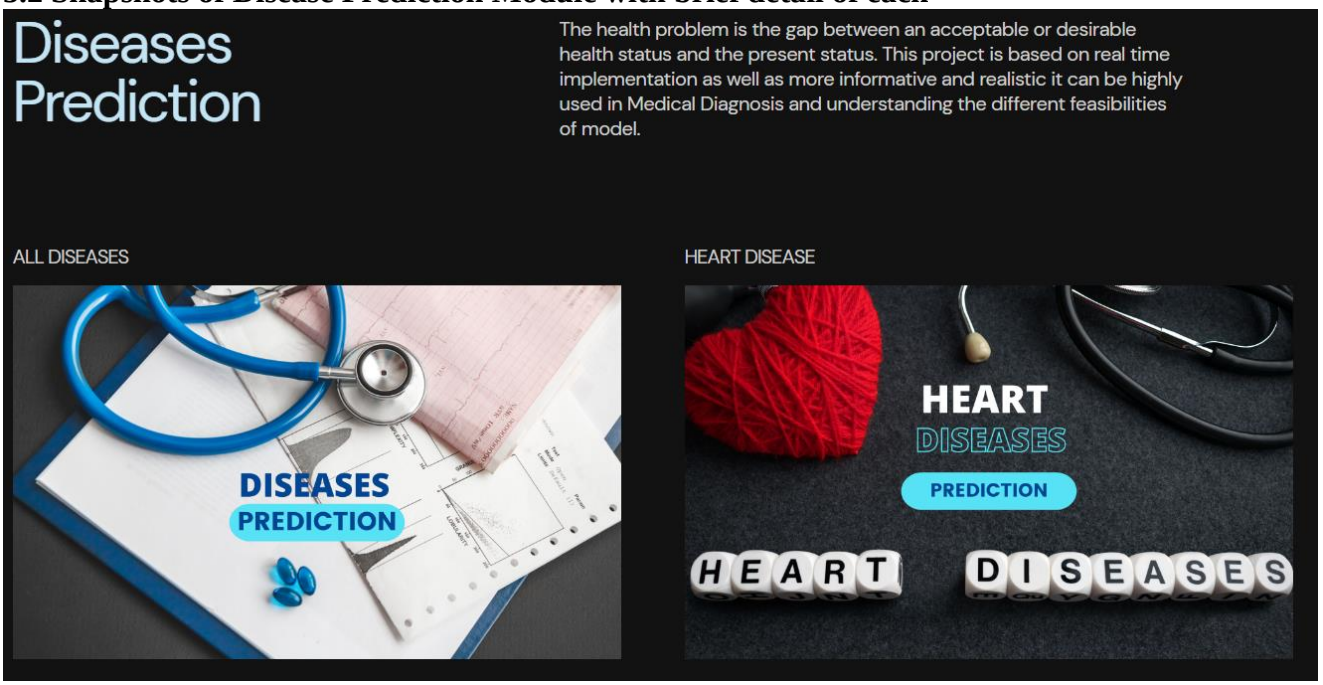

Fig 5.2: Disease Prediction Modules

Disease prediction in automated diagnosis involves using machine learning algorithms and other artificial intelligence (AI) techniques to analyze medical data and predict the likelihood of a patient having a particular disease or condition.

The process of disease prediction typically involves collecting and analyzing various types of patient data, including medical history, symptoms, test results, and other relevant information. This data is then fed into a machine learning algorithm or AI system, which uses statistical models and other techniques to analyze the data and make predictions about the patient's health.
Improved accuracy: Machine learning algorithms can analyze large amounts of medical data and identify patterns that may not be easily identifiable by humans, resulting in more accurate predictions.
Faster diagnosis: Automated diagnosis systems can analyze medical data quickly, allowing healthcare providers to diagnose and treat patients more quickly.
Reduced healthcare costs: Automated diagnosis can reduce the need for expensive diagnostic tests and procedures, resulting in lower healthcare costs.

5.3 Snapshots of 4 Main Modules of Automate Diagnosis & Consultant with a brief detail of each

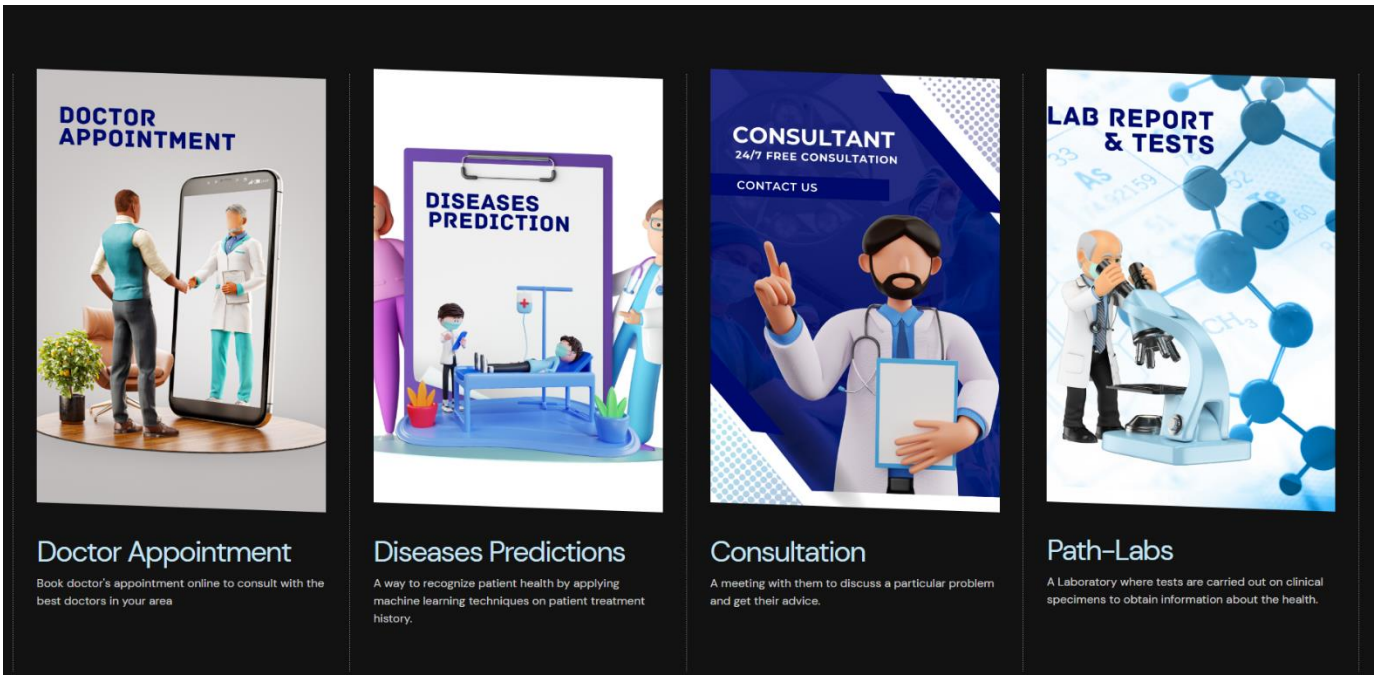

Fig 5.3: Various modules of project

Doctor Appointment: The Doctor Appointment module allows patients to schedule appointments with doctors online. Patients can view the availability of doctors, book appointments, and receive reminders about upcoming appointments. This module helps to streamline the appointment booking process, reducing wait times, and improving patient satisfaction.

Disease Prediction: The Disease Prediction module uses machine learning algorithms to analyze medical data and predict the likelihood of a patient having a particular disease or condition. The module can analyze various types of data, including medical history, symptoms, and test results, to provide accurate and fast

predictions. Disease prediction can help doctors to diagnose diseases early and accurately, resulting in better treatment outcomes.

Consultation: The Consultation module allows patients to consult with doctors remotely using video conferencing, phone calls, or chat platforms. This module enables patients to receive medical advice and treatment from the comfort of their homes, making healthcare more accessible and convenient. The consultation module is particularly useful for patients who live in remote or rural areas, as well as those who have mobility issues.

Path Labs: The Path Labs module allows patients to access laboratory services online. Patients can schedule appointments for lab tests, receive test results online, and consult with doctors about their test results. This module can help to reduce the time and cost associated with laboratory testing, and enable patients to receive faster and more accurate diagnoses.

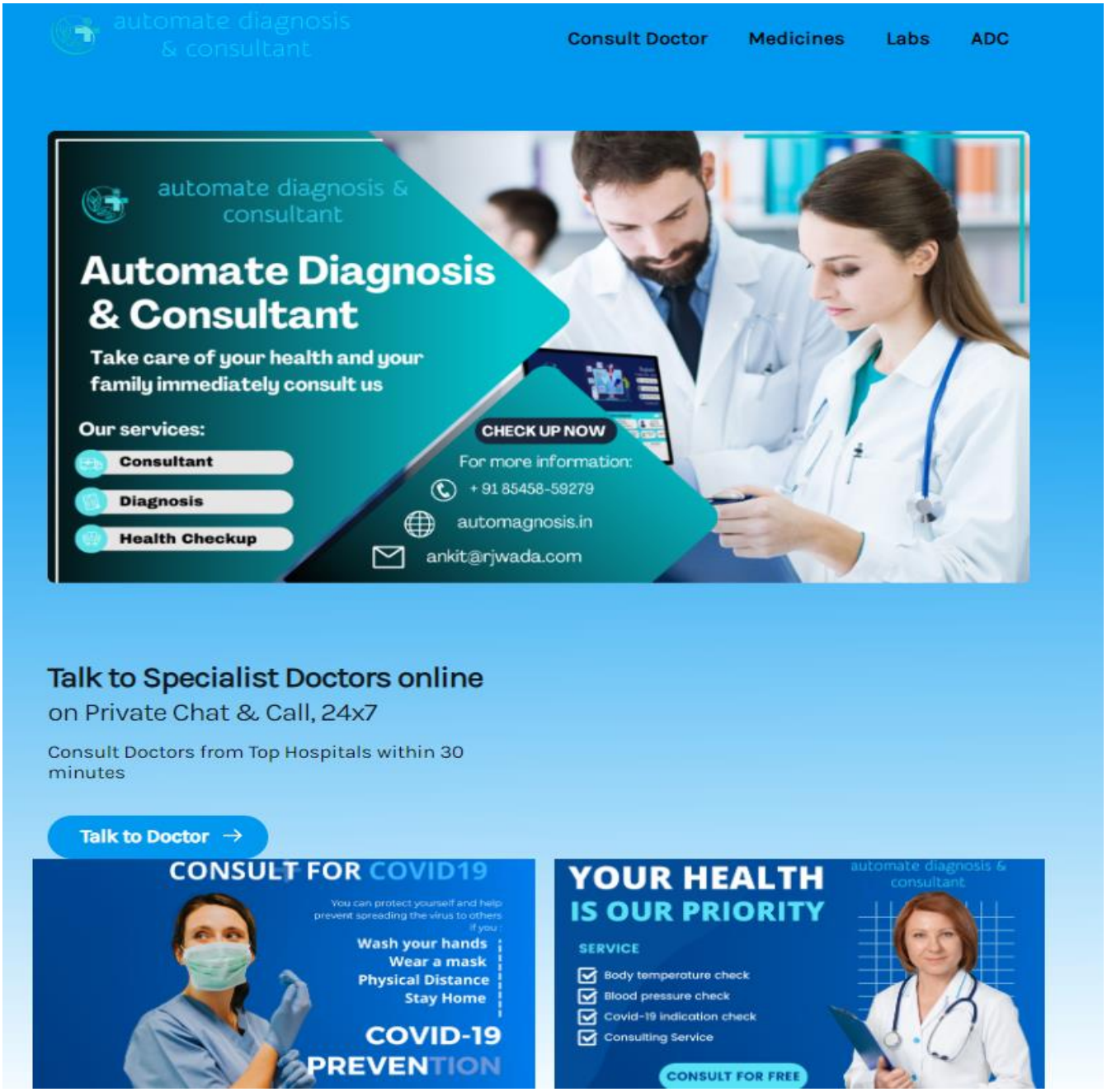

Fig 5.4: Snapshots of Consultant web application of Automate Diagnosis & Consultant.

The main purpose of a consultant web application is to facilitate the delivery of consulting services remotely, making it more convenient for both consultants and clients. This type of application is particularly useful for consultants who work independently or have a distributed team, as it enables them to connect with clients from anywhere in the world.

Consultant web applications typically have a user-friendly interface that allows clients to easily book appointments, communicate with their consultant, and share documents. The application may also include features such as project management tools, invoicing, and analytics to help consultants manage their business operations more efficiently.

Security is a critical consideration for consultant web applications, as they handle sensitive client

information. These applications typically include measures such as encryption, two-factor authentication, and regular security audits to ensure that client data is protected from unauthorized access.

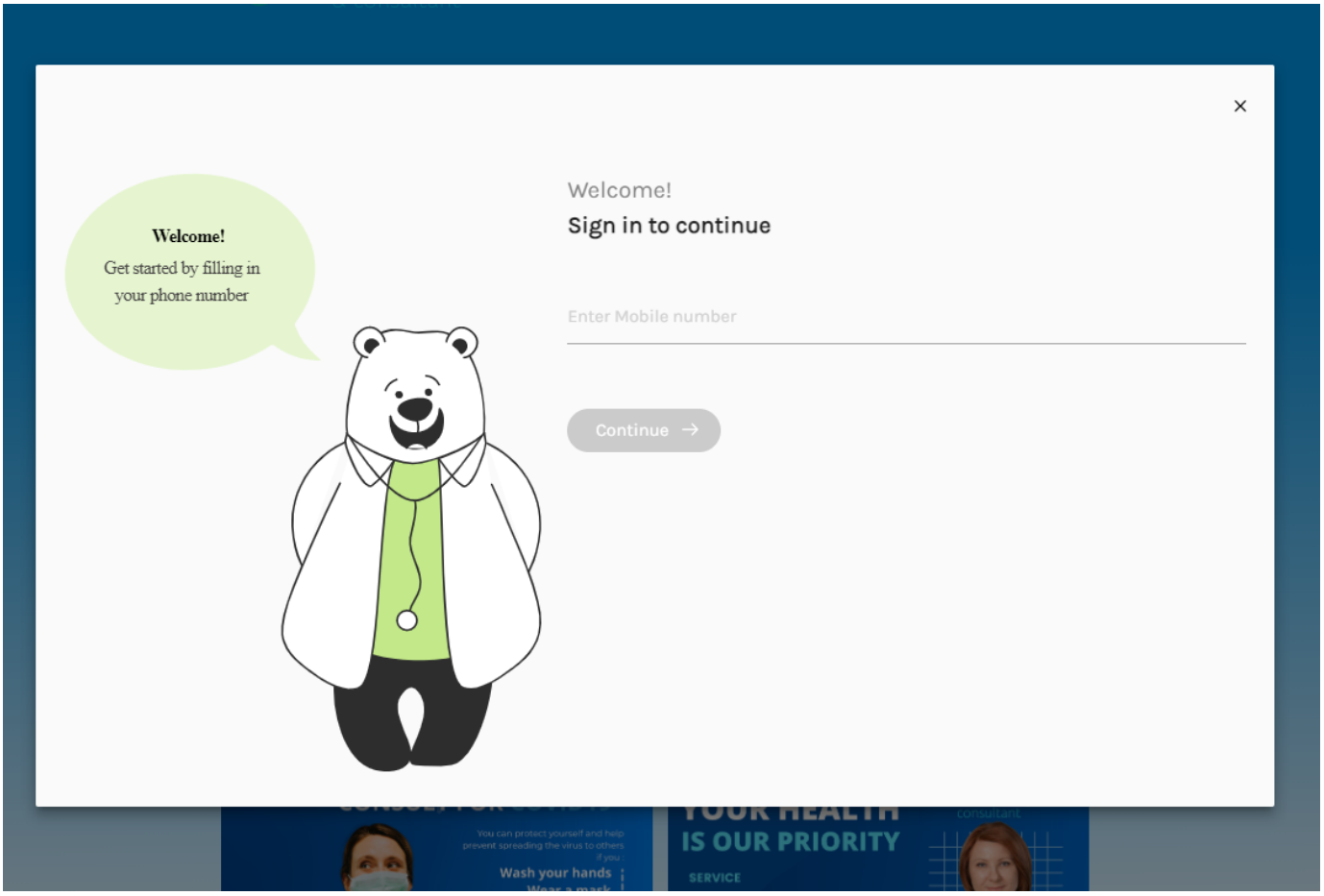

Fig 5.5: Snapshots of Login Module of Automate Diagnosis & Consultant with Mobile Number.

Sign-In Module: The sign-in module of an automated diagnosis system typically requires users to provide their mobile phone numbers for registration and verification purposes. This provides a secure way for users to access the system and protects their personal information. The sign-in module of an automated diagnosis system typically requires users to provide their mobile phone number as part of the registration process.

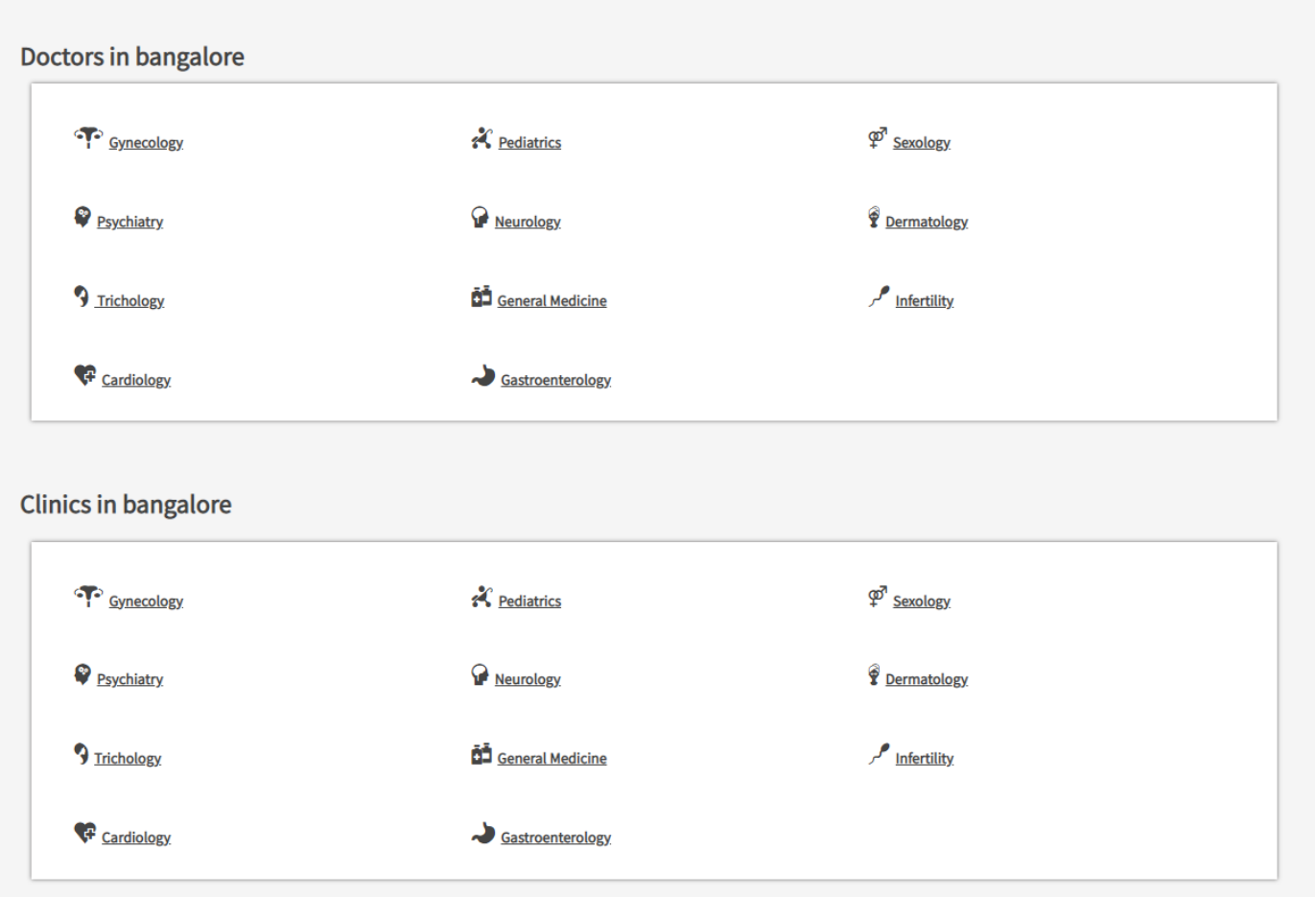

Fig 5.6: Snapshots of Doctor Search Module of Automate Diagnosis & Consultant

Doctor Search by Hospital: This feature enables users to search for doctors based on the hospital or medical facility they work in. Users can enter the name of the hospital or medical center they are interested in, and the system will display a list of doctors who work there. This feature is particularly useful for users who prefer to see a doctor associated with a specific hospital or medical center, or who may need to visit a doctor at a particular location due to travel or other reasons.

Doctor Search by City: This feature enables users to search for doctors based on their location. Users can enter the name of the city or zip code they are interested in, and the system will display a list of doctors who are in that area. This feature is particularly useful for users who are new to an area or who may be traveling and need to find a doctor nearby.

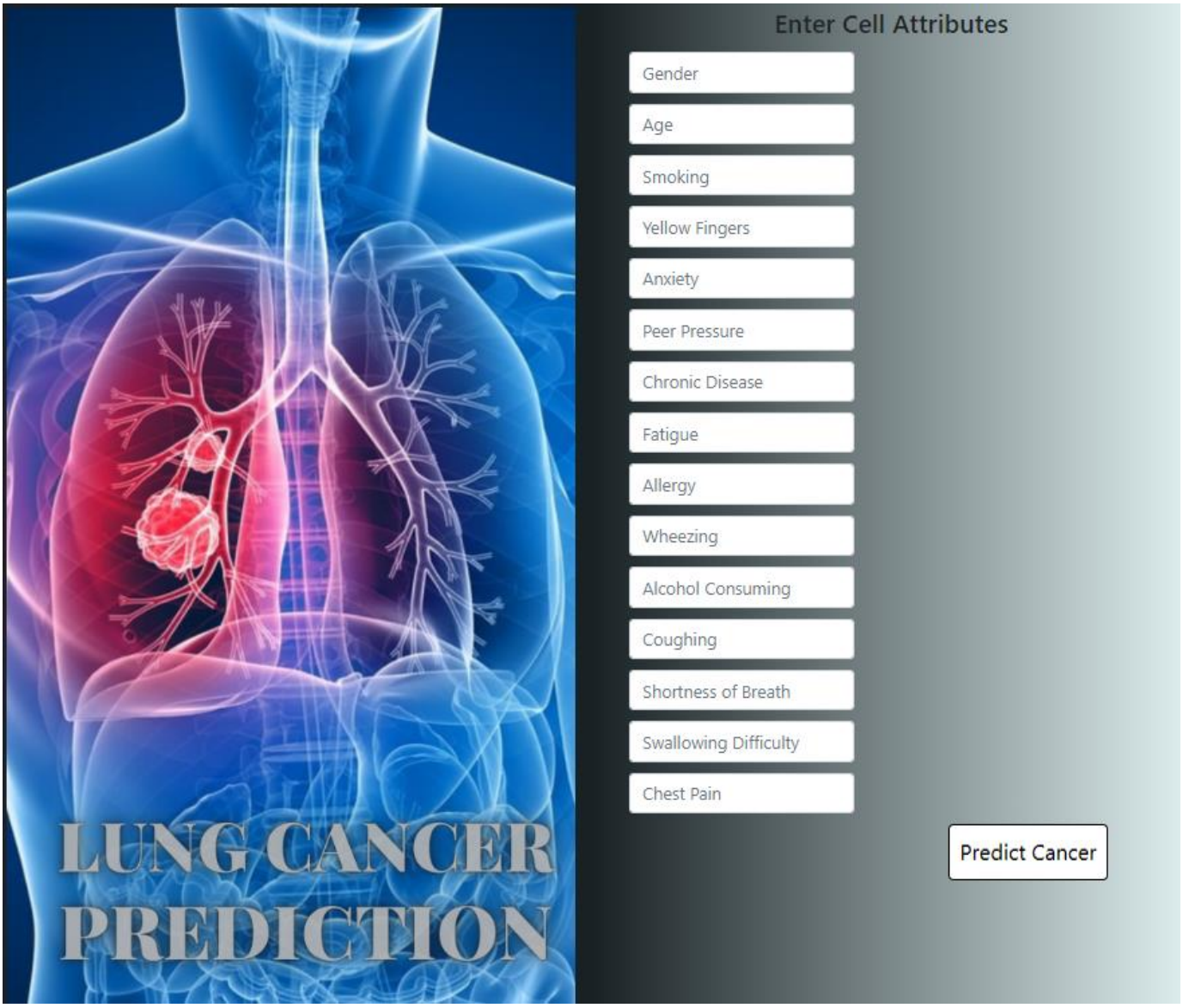

Fig 5.6: Snapshots of Lung Cancer Prediction Module of Automate Diagnosis

Lung cancer prediction is one feature in many automated diagnosis systems. These systems use machine learning algorithms and predictive models to analyze patient data and identify individuals who may be at high risk for developing lung cancer.
Prediction: Based on the data analysis, the system generates a prediction of the patient's likelihood of developing lung cancer. The prediction may be expressed as a probability, such as a percentage or risk score.
Recommendations: The system may provide recommendations for further testing or follow-up based on the predicted risk of lung cancer. For example, patients who are identified as high-risk may be advised to undergo additional screening tests, such as a chest X-ray or CT scan.

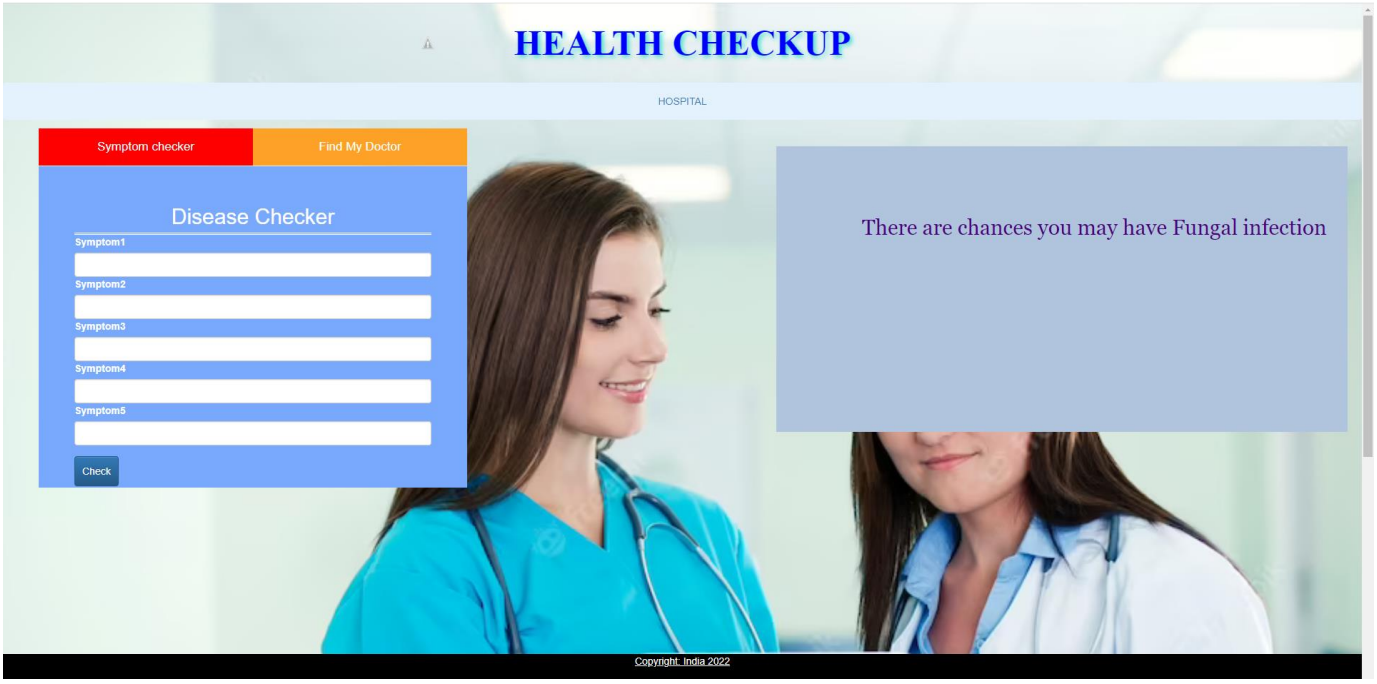

Fig 5.7: Snapshots of Multiple Disease Prediction Module of Automate Diagnosis

Multiple Disease Prediction: The website would use machine learning algorithms to analyze the user's data and generate predictions of the user's risk for multiple diseases, such as diabetes, heart disease, and cancer. These predictions would be displayed to the user in an easy-to-understand format, such as a risk score or percentage.

Recommendations: Based on the predicted risk for each disease, the website might provide recommendations to the user on how to reduce their risk, such as lifestyle changes or additional medical testing. The website might also provide information on the symptoms of each disease and how to detect them early.

Follow-Up: The website might offer follow-up services, such as regular check-ins or reminders for medical testing, to help users stay on top of their health and reduce their risk of developing serious disease

CHAPTER 6 CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

As healthcare services are an important part of our society, automating these services lessenthe burden on humans and diseases through the measuring process. Also, the transparency of this system helps patients to trust it. When the threshold value is reached, the alarm system thatconsists of a buzzer and LED alerts the doctor and they can act more quickly.

The IoT technology helps the server to update the patient data on the website. Many further improvements can be made to our system to make it better and easily adaptable such as adding more advanced sensors. The biometric information of the patient which is stored and published online & transmit to cellular can be given to scientists and researchers in medical fields to analyze the value and find patterns or for other research work.

Future work will focus on monitoring additional health-related parameters using a broader combination of transducers, sensors, and correlation techniques, and on improving system reliability and robustness to patient movement and connectivity losses.

.

6.2 FUTURE SCOPE

The use of machine learning-based automated medical diagnosis has tremendous potential to enhance the healthcare infrastructure in the future. Some of the possible future scopes of this technology are: Personalized Medicine: With the help of machine learning algorithms, it is possible to analyze a patient's health data to identify patterns and provide personalized treatment plans. This can lead to more accurate diagnoses and better outcomes for patients. Early Disease Detection: Machine learning can be used to identify early signs of diseases and predict their progression, allowing for earlier intervention and treatment. Medical Imaging: Machine learning algorithms can be used to analyze medical images, such as CT scans, MRIs, and X-rays, to provide accurate diagnoses and reduce the need for invasive procedures. Drug Discovery: Machine learning algorithms can be used to analyze large amounts of data from clinical trials and other sources to identify potential new drugs or repurpose existing drugs for new uses. Remote Patient Monitoring: Machine learning can be used to monitor patients remotely, allowing healthcare professionals to track patients' vital signs and other data in real time and provide timely interventions when needed. Telemedicine: With the help of machine learning, telemedicine can be made more effective by providing patients with accurate diagnoses and treatment plans remotely. Patient Care Management: Machine learning algorithms can be used to identify patients at high risk of readmission and provide proactive care management to prevent readmissions. Resource Optimization: Machine learning algorithms can be used to optimize the use of resources, such as hospital beds, medical equipment, and staff, to improve efficiency and reduce costs. In conclusion, machine learning-based automated medical diagnosis has the potential to revolutionize the healthcare industry by improving patient outcomes, reducing costs, and enhancing the overall quality of care.

LIST OF PUBLICATIONS

The outcome of the project contains 2 publications:

1- Patent

(12) PATENT APPLICATION PUBLICATION (21) Application No.202211064553 A

(19) INDIA

(22) Date of filing of Application :11/11/2022 (43) Publication Date : 25/11/2022

(54) Title of the invention : REAL-TIME AUTOMATED DIAGNOSIS & CONSULTATION SYSTEM

(51) International classification	:G16H0050200000, G16H0010600000, G16H0050300000, G06N0020000000, G16H0050700000
(86) International Application No Filing Date	:NA :NA
(87) International Publication No	: NA
(61) Patent of Addition to Application Number Filing Date	:NA :NA
(62) Divisional to Application Number Filing Date	:NA :NA

(71)**Name of Applicant :**
1)Ankit Yadav
Address of Applicant :Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India 201206 ---------- ----------
2)Ayush Srivastava
3)Harsh Khatter
Name of Applicant : NA
Address of Applicant : NA
(72)**Name of Inventor :**
1)Ankit Yadav
Address of Applicant :Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India 201206 Ghaziabad ---------- ----------
2)Ayush Srivastava
Address of Applicant :Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India 201206 Ghaziabad ---------- ----------
3)Harsh khatter
Address of Applicant :Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India 201206 Ghaziabad ---------- ----------

(57) Abstract :

The present invention is a system and method which A sizable segment of the world's population lacks access to quality healthcare. The success of healthcare ultimately depends on the doctor's skill. In this study, we investigate if this knowledge may be represented as an information corpus, or as data that has been retrieved using data mining methods, particularly the Machine Learning & Deep Learning Model, to make a diagnosis. When the medical diagnosis is made widely available, coverage increases and life quality improves. In order to determine whether inferences about the causes of various diseases can be made from the data, this paper provides an overview of machine learning approaches used in the classification of various diseases. We outline a few of our findings from the trials we ran before offering some suggestions for the future. The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures, monitoring systems are designed to lower health care expenditures. Using the data mining modeling technique, the integration of clinical decision support with computer-based patient records could decrease medical errors, increase patient safety, stop unwelcome practice variance, and improve practice outcomes. Connecting patients and doctors through a user-friendly interface will make it easier for patients to use in emergency situations The figures of the present invention showed the detail description of the work.

No. of Pages : 19 No. of Claims : 5

2-Research Paper

Conference ID : 57294

GLA UNIVERSITY MATHURA

Accredited with A+ Grade by NAAC

12-B Status from UGC

6th IEEE International Conference on
Information Systems and Computer Networks (ISCON 2023)
March 03rd - 04th,2023

Financial Sponsors

Technical Co-Sponsor

IEEE UP SECTION (INDIA)

Certificate of Participation

This is to certify that ***Mr./Ms./Dr./Prof. Dr. Harsh Khatter*** *from* ***KIET Group of Institutions, Delhi NCR, Ghaziabad*** *has presented a paper entitled* ***"Machine learning based automated medical diagnosis for healthcare"*** *in* ***6th International Conference on Information Systems and Computer Networks (ISCON-2023)*** *organized by Department of Computer Engineering & Applications at GLA University Mathura, Uttar Pradesh, INDIA during March 03rd-04th, 2023.*

Prof. Dilip Kumar Sharma *General Chair*

Prof. Ashok Bhansali *General Chair*

Dr. Rohit Agrawal *General Chair*

Dr. Ashish Sharma *Convener*

Knowledge Partners

REFERENCES

[1] D. Fagella, Applications of Machine Learning in Pharma and Medicine, 2017, [online] accessed on 14 Nov 2022.

[2] V. S. Pendyala and S. Figueira, "Automated Medical Diagnosis from Clinical Data," 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService), 2017, pp. 185-190, doi: 10.1109/BigDataService.2017.14.

[3] F. Santos, F. Silva and P. Georgieva, "Automated Diagnosis of Skin Lesions," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 545-550, doi: 10.1109/IS48319.2020.9200090.

[4] V. Datla et al., "Automated clinical diagnosis: The role of content in various sections of a clinical document," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017, pp. 1004-1011, doi: 10.1109/BIBM.2017.8217794.

[5] J. Siddiqi, B. Akhgar, A. Gruzdz, G. Zaefarian and A. Ihnatowicz, "Automated Diagnosis System to Support Colon Cancer Treatment: MATCH," Fifth International Conference on Information Technology: New Generations (itng 2008), 2008, pp. 201-205, doi: 10.1109/ITNG.2008.62.

[6] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 2017, pp. 1-6, doi: 10.1109/ICTKE.2017.8259629.

[7] R. Chellappa, S. Theodoridis and A. van Schaik, "Advances in Machine Learning and Deep Neural Networks," in Proceedings of the IEEE, vol. 109, no. 5, pp. 607-611, May 2021, doi: 10.1109/JPROC.2021.3072172.

[8] F. Ertam and G. Aydın, "Data classification with deep learning using Tensorflow," 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 755-758, doi: 10.1109/UBMK.2017.8093521.

[9] H. Božiković and M. Štula, "Web design — Past, present and future," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 1476-1481, doi: 10.23919/MIPRO.2018.8400266.

[10] D. S. Blyth, "Web page design for HTML (and friends)," IPCC 96: Communication on the Fast Track. IPCC 96 Proceedings, 1996, pp. 89-103, doi: 10.1109/IPCC.1996.552585.

[11] H. Ran, W. Zhuo and X. Jianfeng, "Web Quality of Agile Web Development," 2009 IITA International Conference on Services Science, Management and Engineering, 2009, pp. 426-429, doi: 10.1109/SSME.2009.112.

[12] S. Drobi, "Play2: A New Era of Web Application Development," in IEEE Internet Computing, vol. 16, no. 4, pp. 89-94, July-Aug. 2012, doi: 10.1109/MIC.2012.84.

[13] H. Sugimura et al., "Development of immersive display system of web service in living space," 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), 2014, pp. 49-50, doi: 10.1109/GCCE.2014.7031106.

[14] N. Sandhya and K.R. Charanjeet, "A review on Machine Learning Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, pp. 395-399, 2016, ISSN 2321-8169.

[15] Harsh Khatter, Amit Kumar Gupta, Ruchi Rani Garg, and Mangal Sain. 2022. "Analysis of the S-ANFIS Algorithm for the Detection of Blood Infections Using Hybrid Computing" Electronics 11, no. 22: 3733.

[16] Kanika Gupta, Nandita Goyal, Harsh Khatter "Optimal reduction of noise in image processing using collaborative inpainting filtering with Pillar K-Mean clustering", The Imaging Science Journal (2019), Vol. 67(2), pp:100-114, ISSN 1743-131X
[17] Sharma, Sarang; Gupta, Sheifali; Gupta, Deepali; Juneja, Abhinav; Khatter, Harsh; Malik, Sapna; Bitsue, Zelalem Kiros "Deep Learning Model for Automatic Classification and Prediction of Brain Tumor", Journal of Sensors (2022), Hindawi, Vol 2022(3065656), 8 April 2022.
[18] Gunjan Raghav, Harsh Khatter, "Intelligent Curation Fuzzy Inference System for Blood Infections in Android", IEEE 4th International Conference on Computational Intelligence & Communication Technology (CICT), 2018, Ghaziabad, India.
[19] Vaishali Aggarwal, Harsh Khatter, Anil Kumar Ahlawat, "Performance analysis of the competitive learning algorithms on Gaussian data in automatic cluster selection", 2nd IEEE International Conference CICT 2016, Feb 12-13 2016, ABES EC Ghaziabad.
[20] Anjali Jain, Harsh Khatter, Rajesh Kumar Tewari, (2020) "Insight To Pandemic Covid-19: Analysis, Statistics and Prevention", Aegaeum Journal, Vol. 8 (7), pp: 551-561, ISSN 0776-3808 DOI:16.10089.AJ.2020.V8I7.285311.3758
[21] Harsh Khatter, Amit Kumar Gupta, Ruchi Rani Garg, and Mangal Sain. 2022. "Analysis of the S-ANFIS Algorithm for the Detection of Blood Infections Using Hybrid Computing" Electronics 11, no. 22: 3733, 14 Nov 2022, https://doi.org/10.3390/electronics11223733
[22] Harsh Khatter, Anil K Ahlawat, "An intelligent personalized web blog searching technique using fuzzy-based feedback recurrent neural network". Soft Computing, 24, 9321–9333, 29 April 2020, (2020). https://doi.org/10.1007/s00500-020-04891-y
[23] Harsh Khatter, Anjali Jain, Poonam Pandey, "Classification and Categorization of Blood Infection using Fuzzy Inference System", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-5 Issue-2, May 2015, pp. 95-97

Machine learning-based automated medical diagnosis for healthcare

Harsh Khatter[1], Ankit Yadav[2], Ayush Srivastava[3]

[1,2,3]*Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India*

[1]harsh.khatter@kiet.edu

[2]ankit.1923co1041@kiet.edu

[3]ayush.1923co1040@kiet.edu

Abstract— **A sizable segment of the world's population lacks access to quality healthcare. The success of healthcare ultimately depends on the doctor's skill. In this study, we investigate if this knowledge may be represented as an information corpus, or as data that has been retrieved using data mining methods, particularly the Machine Learning & Deep Learning Model, to make a diagnosis. When the medical diagnosis is made widely available, coverage increases, and life quality improves. In order to determine whether inferences about the causes of various diseases can be made from the data, this paper provides an overview of machine learning approaches used in the classification of various diseases. We outline a few of our findings from the trials we ran before offering some suggestions for the future.**

The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures, monitoring systems are designed to lower health care expenditures. Using the data mining modeling technique, the integration of clinical decision support with computer-based patient records could decrease medical errors, increase patient safety, stop unwelcome practice variance, and improve practice outcomes. Connecting patients and doctors through a user-friendly interface will make it easier for patients to use in emergency situations.

Keywords— Artificial Intelligence, Machine Learning, Health Analysis, Preventive systems

I. INTRODUCTION

Machine learning (ML) is a branch of artificial intelligence that uses "learning ability provided to computers without additional programming" to address problems in the real world. Research into whether computers could learn to emulate the human brain led to the development of machine learning. When Arthur Samuel created the first checkers game-playing software in 1952, ML made its initial attempts to generate the necessary skills to defeat a world champion. Later in 1957, Frank Rosenblatt developed an electronic system that can mimic the way the human brain works to learn how to handle complex issues [5].

The growth of ML has facilitated the increased use of computers in medicine.There are three primary goals of data mining that tend to be prediction, description, and presentation. Prediction involves some variables or fields in the dataset to predict unknown or future values of other variables of interest. Description focuses on finding patterns describing the data that can be interpreted by humans, on the other hand Presentation plays an important role to be easily understandable to humans.

Health issues are more prevalent today than they were 25 to 30 years ago due to urbanization and industrialization. Unexpected changes in the environment have a direct impact on health. As a result, more health issues emerge daily, necessitating daily health condition checks. This project, which is based on real-time implementation and is more informative and realistic, can be very useful in helping to diagnose medical conditions and comprehend the many model feasibility options.

The availability of medical facilities for people is limited under these unheard-of economic conditions, and it is even more so for those with specific requirements. People with other illnesses didn't even get an opportunity to access medical aid because the healthcare industry became so preoccupied with Covid-19 instances [20]. We are here to provide a model that will help alleviate the lack of facilities and poor medical management.

The goal of creating automated systems is to decrease the amount of time and money spent on health care by lowering the number of doctor visits, hospital stays, and diagnostic testing procedures. The second main method is supervised learning, which uses machine learning techniques to learn a function from a collection of training data. Prediction and generalization are two key requirements for supervised learning algorithm performance. The trained function ought to be

capable of accurately predicting the results for data that are not included in the training set. It should also serve as a model that generalizes to new data points and captures the underlying traits of the training data.

Machine learning is frequently used to categorize diseases, and scientists are becoming increasingly more interested in creating such systems for many disease diagnosis and tracking as well. Diabetes, cardiovascular disease & many more are the top 10 killers worldwide, according to the World Health Organization (WHO). According to a study from January 2017, CDs are the leading cause of death worldwide. In the list of the top 10 causes of death over the past 15 years, the world's worst disease, which claimed 15 million lives in 2015, holds the top spot.

In this paper, the authors focuses on the various methods studied on preventive health and how to analysis the patient's health.

II. Literature Survey

Biosignals from patients have been used in the past to construct AI systems for clinical decision assistance. Such organized clinical data includes unprocessed signals. a lack of sufficient background for appropriate interpretation, whereas clinical publications with unstructured free text contain comprehensive explanations of broader clinical situations.

Tamilselvan. P [1]. Monitors based on blood pressure and ECG readings are available. Reactions are kept the signals that these sensors transmit to the Using a signal conditioner and amplifier, Raspberry Pi because the signals' levels are low (gain), an amplifier is necessary (scu). The signals that these sensors transmit to the Raspberry Pi are processed by a signal conditioner and amplifier, which is required because the signal levels are low (gain). Using a circuit, the signal is amplified and sent to a Raspberry Pi. Linux is used to run the Raspberry Pi computer. The system operates like a minicomputer processing system. Here, patients' ECG and blood pressure are measured with the appropriate sensors. Moreover, it is monitorable on a computer's monitor. Utilizing a Raspberry Pi and monitoring from anywhere, Internet sources are used everywhere.

Vivek Datla, Sadid A. Hasan [2]. outlines our Knowledge Graph (KG)-system for based clinical diagnostic inference. We performed substantial testing on the MIMIC-III benchmark dataset, analyzing different parts of a clinical note. Results proved that the details of the current illness's history were relevant. The parts on prior health histories often offer the greatest insight. inference of a clinical diagnosis in comparison to all portions. Furthermore, we demonstrated that the KG-based system can perform admirably with a loose accuracy metric in comparison to the cutting-edge CMemNN model.

Hiroshi Sugimura, Kazuki Utsumi [3]. They suggest a method that superimposes online service information onto the daily environment. Natural behaviors like speech and gestures are used to operate the system. The system is put together using three input/output devices: a microphone for sound recognition, a camera for gesture detection, and a projector for information presentation. We provided a detailed account of the conception, design, execution, and assessment of a prototype system. The proposal system's value was then confirmed.

Fabio Santos, Filipe Silva and Petia Georgieva [4]. focused on the skin lesion diagnosis techniques integrated inside eHealth apps that help individuals and medical professionals and are clearly needed as the prevalence of skin cancer grows. Meanwhile, recent developments in deep learning techniques enable performance that is close to that of a dermatologist and has a large room for growth, outperforming previous approaches. Before putting such tools into use in the real world, issues like the need for large datasets or the high computing needs must be resolved because they negatively affect how well models function. However, effective methods like these reduce their impacts through transfer learning and data augmentation, according to research. Finally, it is anticipated that when more information on skin lesions is made publicly available, these difficulties will lose some of their significance.

Berina Ali [5] provides an overview of machine learning methods for categorizing CVD and diabetes using artificial neural networks (ANNs) and Bayesian networks (BNs). A comparative study was carried out on a few publications released throughout the time period, between 2008 and 2017. In a few

articles, multilayer feedforward neural networks using the Levenberg-Marquardt learning method are the most commonly utilized ANN type. Additionally, utilizing ANN improved the computation of the mean accuracy of observed networks, indicating a greater likelihood of obtaining more accurate findings for the categorization of CVD and/or diabetes.

Authors focus to check if seeding strategies have strong influences on the success of viral marketing campaigns or not. We can determine the era of marketing how one can have successful approach in getting viral videos [11].

III. METHODOLOGY

The methodology's goal is to forecast an individual's risk of developing kidney, lung, and breast cancer, heart disease, and diabetes using a few questions and machine learning models in an end-to-end procedure. the system I used for my research has the following software and system configurations: On an Intel(R) Core(TM) i5-2310M GPU @1650Ti with 8 GB RAM, Jupyter Notebook 5.5.0 and VS Code 1.73 are used to implement Python 3 and the Flask framework.

Figure 1 displays a block schematic of the fundamental procedures used for each machine-learning model. To transform the raw data into a form that can be used, data cleaning is done first. Data analysis is carried out after data cleansing to ascertain the significance of characteristics. Once a goal has been established, it is time to start gathering the data required for analysis. Your data team will be responsible with cleaning and sifting through the data once it has been gathered from all the required sources. Because not all data is good data, data cleansing is crucial during the data analysis process. Data mining, which is referred to as "knowledge discovery inside databases," is one method. In order to predict what will probably happen next in the future, predictive analysis looks ahead to the future. These methods are a component of inferential statistics, which is the act of examining statistical data in order to make inferences about the connections between various sets of data. We have to utilize the Train/Test methodology to evaluate the model's performance. The train/test approach is a way to gauge how accurate your model is. Because you divide the data set into two sets, a training set and a testing set, the method is known as Train/Test. 20% for testing, 80% for training. Using the training set, you train the model. Utilizing the testing set, you test the model. Create the model by training it. To test a model is to determine its correctness. After confirming that our model is sound, we can now begin making fresh value predictions.

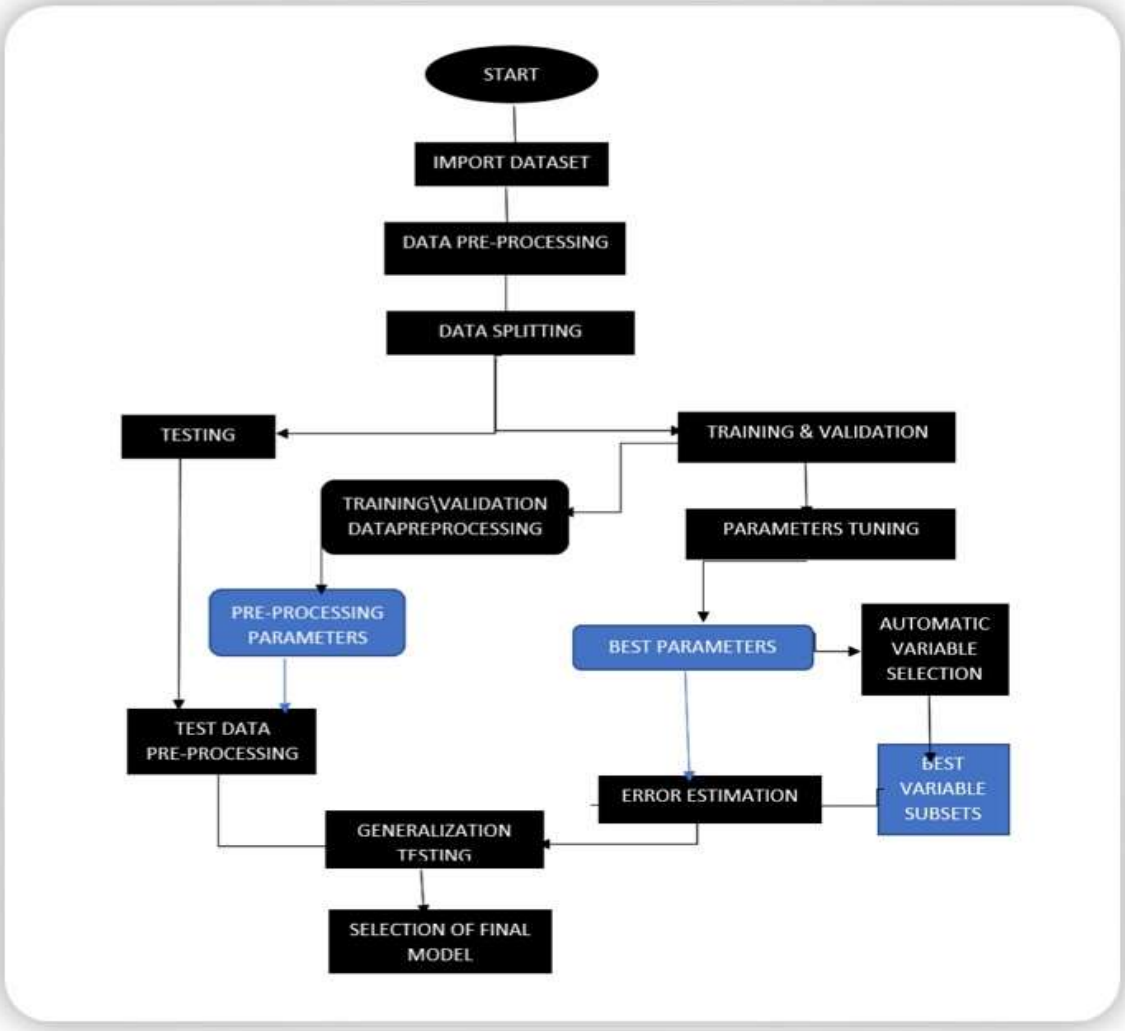

Figure 1: Flow chart of the proposed model

IV. ALGORITHMIC APPROACHES FOR PROPOSED WORK

A. *Random Forest Classifier Algorithm:*

A group of several decision trees is called a random forest. Decision trees are employed as parallel estimators in the bagging technique, which is used to construct random forests. When used in a classification issue, the outcome is determined by the majority vote of the findings from each decision tree. In a regression, the mean value of something like the target values in a leaf node serves as the prediction. The mean value of the decision tree outcomes is taken into account by random forest regression. [6,7]

B. *Logistic Regression:*

A set of independent variables is used to estimate discrete values (often binary values like 0/1) using logistic regression. By fitting data to a logit function, it aids in the prediction of an event's likelihood. Logistic regression is another name for it. The following techniques are frequently used to enhance logistic regression models: incorporate interaction terms, remove features, regularize approaches, and employ a non-linear mode. [12,13]

C. *Decision Tree:*

The Decision Tree method, a supervised learning technique used for issue classification, is one of the most widely used algorithms in machine learning today. Both continuous and categorical dependent variables may be classified using it. Based on the most important characteristics/independent variables, this method splits the population into two or more homogenous groupings.

Step 1: Gathering and identifying the data that will be provided to the network as input is the first stage in the categorization of diabetes or heart disease using Logistic Regression, Decision Tree & Random Forest. [14]

Step 2: The network receives a defined training dataset and the selected training method. After the training phase, the Logistic Regression, Decision Tree & Random Forest are also put to the test to get feedback on how well they categorize the condition.[15]

Step 3: The "training data set" is used to train the classifier, the "validation set" is used to fine-tune the parameters, and the "test data set" is used to evaluate the performance of the classifier. It's vital to keep in mind that only the training and/or validation set is available when the classifier is being trained. The test data set must not be used for classifier training. Only when the classifier is being tested will the test set be accessible.

Step 4: We may then create a confusion matrix, which indicates how effectively our model has been trained, after this is complete. True Positives, True Negatives, False Positives, and False Negatives are the four parameters that make up a confusion matrix. [16] In order to create a model that is more accurate, we would want to obtain more data for the true negatives and true positives. The number of classes has a direct impact on the size of the confusion matrix.

Step 5: The model creation process includes a step called model evaluation. Finding the model that best depicts our data and predicts how well the model will perform in the future is helpful. In order to enhance the model, we may adjust its hyper-parameters to attempt to raise its accuracy while also looking at the confusion matrix to try to increase the proportion of true positives and true negatives.

The dashboard will run the function that connects patients and doctors. The dashboard has been designed to get a better understanding of a patient's condition. The medical profession is being transformed by technological advancements, which is causing a rapid change in the medical business. Artificial intelligence (AI) is being used to diagnose illnesses, predict diseases, and develop treatments. In this article, it will be discussed how AI can support the management of level-three 911 calls and assist physicians in making better patient decisions. According to the doctors' available time slots, those who have appointments can show up for their appointments. To avoid any unneeded waiting times, this is done. Modern medical professionals are employing cutting-edge technology to provide more precise diagnoses and individualized treatment strategies [17,18]. Doctors can now give patients extensive medical data, an in-depth report on their symptoms, and personalized care plans that can be seen on a mobile device thanks to artificial intelligence (AI). With the use of an app, the doctor may view the patient's information, write him a prescription, and suggest testing. This software assists in quickly diagnosing the patient since it has all the information required to understand the ailment, its symptoms, and recommended treatments. The outdated paper lab slip is being replaced with a modern dashboard. Patients will be able to schedule an appointment at a time that works for them because these slips are now digital. As soon as the reports are finished being processed by the laboratory, they will also be automatically uploaded into the patient's database. Doctors use information from a variety of sources to provide the most precise diagnosis possible. With the more recent use of electronic medical records, a more complete image of patients is now possible, improving care. [19]

V. Discussion

Data mining techniques play a significant role in medical systems, which will significantly contribute to the advancement of the medical industry. This paper provides a disease categorization based on several data mining and artificial intelligence technologies. Furthermore, we discovered in the literature that there are two kinds of factors utilized in disease categorization.

We noticed in the literature that the classification of illnesses still has room for improvement. Despite the fact that proper characteristics may be derived from the ECG, it is a noninvasive strategy used to diagnose patients, and the ECG signal does not provide the necessary information. Because biosignals have an irregular structure, creating an effective technique for hidden factor extraction from

ECG signals is particularly challenging. Several studies have found that the feature extraction method is incapable of determining the exact values of unmasked ECG signal parameters. Furthermore, using a restricted dataset for classification may result in misclassification; hence, in order to overcome the error rate, it is critical to avoid using a short dataset for classification.

VI. CONCLUSION

In this project, the focus is on healthcare services, which are such an important aspect of our society, automating them relieves human stress while also making measurement simpler. Furthermore, the system's transparency promotes patient trust. When the threshold value is achieved, an alarm system consisting of a buzzer and an LED alerts the physicians, allowing them to respond more quickly.

This work focused on a variety of health concerns that affect not only India but the entire world. Visualize the dataset and identify diseases using all ML and DL algorithms. Integrate physicians and patients with an interactive and user-friendly interface so that patients may use it in an emergency. This model's three key purposes are prediction, description, and presentation. The patient's biometric data, which is captured, published online, and transferred to cellular devices, may be made available to scientists and researchers in medical disciplines in order to evaluate its value and reveal patterns, as well as for other research objectives.

Future research will focus on monitoring additional health-related indicators with a bigger collection of transducers, sensors, and correlation algorithms, as well as improving system dependability and resilience in the face of patient movement and connectivity losses.

REFERENCES

[1] D. Fagella, Applications of Machine Learning in Pharma and Medicine, 2017, [online] accessed on 14 Nov 2022.

[2] V. S. Pendyala and S. Figueira, "Automated Medical Diagnosis from Clinical Data," 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService), 2017, pp. 185-190, doi: 10.1109/BigDataService.2017.14.

[3] F. Santos, F. Silva and P. Georgieva, "Automated Diagnosis of Skin Lesions," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 545-550, doi: 10.1109/IS48319.2020.9200090.

[4] V. Datla et al., "Automated clinical diagnosis: The role of content in various sections of a clinical document," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017, pp. 1004-1011, doi: 10.1109/BIBM.2017.8217794.

[5] J. Siddiqi, B. Akhgar, A. Gruzdz, G. Zaefarian and A. Ihnatowicz, "Automated Diagnosis System to Support Colon Cancer Treatment: MATCH," Fifth International Conference on Information Technology: New Generations (itng 2008), 2008, pp. 201-205, doi: 10.1109/ITNG.2008.62.

[6] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 2017, pp. 1-6, doi: 10.1109/ICTKE.2017.8259629.

[7] R. Chellappa, S. Theodoridis and A. van Schaik, "Advances in Machine Learning and Deep Neural Networks," in Proceedings of the IEEE, vol. 109, no. 5, pp. 607-611, May 2021, doi: 10.1109/JPROC.2021.3072172.

[8] F. Ertam and G. Aydın, "Data classification with deep learning using Tensorflow," 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 755-758, doi: 10.1109/UBMK.2017.8093521.

[9] H. Božiković and M. Štula, "Web design — Past, present and future," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 1476-1481, doi: 10.23919/MIPRO.2018.8400266.

[10] D. S. Blyth, "Web page design for HTML (and friends)," IPCC 96: Communication on the Fast Track. IPCC 96 Proceedings, 1996, pp. 89-103, doi: 10.1109/IPCC.1996.552585.

[11] H. Ran, W. Zhuo and X. Jianfeng, "Web Quality of Agile Web Development," 2009 IITA International Conference on Services Science, Management and Engineering, 2009, pp. 426-429, doi: 10.1109/SSME.2009.112.

[12] S. Drobi, "Play2: A New Era of Web Application Development," in IEEE Internet Computing, vol. 16, no. 4, pp. 89-94, July-Aug. 2012, doi: 10.1109/MIC.2012.84.

[13] H. Sugimura et al., "Development of immersive display system of web service in living space," 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), 2014, pp. 49-50, doi: 10.1109/GCCE.2014.7031106.

[14] N. Sandhya and K.R. Charanjeet, "A review on Machine Learning Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, pp. 395-399, 2016, ISSN 2321-8169.

[15] Harsh Khatter, Amit Kumar Gupta, Ruchi Rani Garg, and Mangal Sain. 2022. "Analysis of the S-ANFIS Algorithm for the Detection of Blood Infections Using Hybrid Computing" Electronics 11, no. 22: 3733.

[16] Kanika Gupta, Nandita Goyal, Harsh Khatter "Optimal reduction of noise in image processing using collaborative inpainting filtering with Pillar K-Mean clustering", The Imaging Science Journal (2019), Vol. 67(2), pp:100-114, ISSN 1743-131X

[17] Sharma, Sarang; Gupta, Sheifali; Gupta, Deepali; Juneja, Abhinav; Khatter, Harsh; Malik, Sapna; Bitsue, Zelalem Kiros "Deep Learning Model for Automatic Classification and Prediction of Brain Tumor", Journal of Sensors (2022), Hindawi, Vol 2022(3065656), 8 April 2022.

[18] Gunjan Raghav, Harsh Khatter, "Intelligent Curation Fuzzy Inference System for Blood Infections in Android", IEEE 4th International Conference on Computational Intelligence & Communication Technology (CICT), 2018, Ghaziabad, India.

[19] Vaishali Aggarwal, Harsh Khatter, Anil Kumar Ahlawat, "Performance analysis of the competitive learning algorithms on Gaussian data in automatic cluster selection", 2nd IEEE International Conference CICT 2016, Feb 12-13 2016, ABES EC Ghaziabad.

[20] Anjali Jain, Harsh Khatter, Rajesh Kumar Tewari, (2020) "Insight To Pandemic Covid-19: Analysis, Statistics and Prevention", Aegaeum Journal, Vol. 8 (7), pp: 551-561, ISSN 0776-3808 DOI:16.10089.AJ.2020.V8I7.285311.3758

[21] Harsh Khatter, Amit Kumar Gupta, Ruchi Rani Garg, and Mangal Sain. 2022. "Analysis of the S-ANFIS Algorithm for the Detection of Blood Infections Using Hybrid Computing" Electronics 11, no. 22: 3733, 14 Nov 2022, https://doi.org/10.3390/electronics11223733

[22] Harsh Khatter, Anil K Ahlawat, "An intelligent personalized web blog searching technique using fuzzy-based feedback recurrent neural network". Soft Computing, 24, 9321–9333, 29 April 2020, (2020). https://doi.org/10.1007/s00500-020-04891-y

[23] Harsh Khatter, Anjali Jain, Poonam Pandey, "Classification and Categorization of Blood Infection using Fuzzy Inference System", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-5 Issue-2, May 2015, pp. 95-97