SMART DISEASE PREDICTION

[1]Jayshree Pawar, [2]Habib Khan, [3]Nishant Walunj, [4]Sanskruti Sonwal, [5]Sachita Dodke

[1]Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Information Technology
[1] Vasantdada Patil Prathishtan College of Engineering, Mumbai, India

Abstract: We are living in an era where technology is advancing at unprecedented rate. Nowadays, Artificial Intelligence and Machine L earning are used in e very domain wherein machines or computers are being trained to work automatically with very less human efforts. Machine Learning Algorithms are very useful in prediction, analysis and training. We will use ML Algorithms in predicting and analysis of various diseases in human beings. Health is one of the precious asset for a human being but due to the ongoing pandemic people can't recognize and treat their diseases from their home so we are aiming to develop a disease prediction system using ML Algorithms for better prediction of diseases by providing their symptoms to recognize the diseases more precisely with its consequences and treatment with the ease of using it at their own comfort zone. We are developing this system also because people can consult with their respective doctors via live consultation without physical contact.

Our interface would help people to some extent in-order to reduce the risk associate with predicted diseases to reduce the impact on other body parts. With the help of extensive powerful ML Algorithms accuracy is highest. People can also maintain safety protocols in this pandemic by predicting the diseases from their home.

Index Terms - **Scikit-learn, NumPy, Data Preprocessing, Dataset, Algorithms, Training Data, Training Set, Machine Learning, Naive Bayes, KNN, Decision Tree, Kernel SVM, Logistic Regression, Random Forest, Django, Training Model, Web Module, Data Collection Module, APIs, Artificial Intelligence, Disease Detection Module, Authentication, User Interface, SDP**

1.INTRODUCTION

Smart Disease Prediction is a Web Application used for predicting Human diseases by providing respective symptoms. Our system uses powerful Machine Learning Algorithms to predict the disease based on the symptoms provided by the users.

1.1 Purpose

As we know that there is quite confusion amongst patients in recognizing and predicting their diseases based on their symptoms. This problem of patients should be minimized by developing a Disease Prediction System with providing their own symptoms with accurate prediction and analysis of that disease using powerful Machine Learning Algorithms. So this interface would surely facilitate them.

1.2 Project Scope

SDP will provide powerful analysis features using various libraries, APIs, and Machine Learning Algorithms with appropriate and precise disease prediction by providing symptoms. This will not only reduce anxiety of people but will also give full satisfactory results. SDP will have specialist recommendation for the respective diseases.

### 1.3 Project Goals and Objectives
1.3.1 Goals
Early detection of Diseases

Prediction and detecting the diseases with the help of some symptoms at early stage would be much more beneficial to users in improving their health.

Effective and Accurate Prediction

Use of Extensive Machine learning algorithms to develop an interface with much better accuracy to avoid Emergency situations. To provide quick results in critical situations.
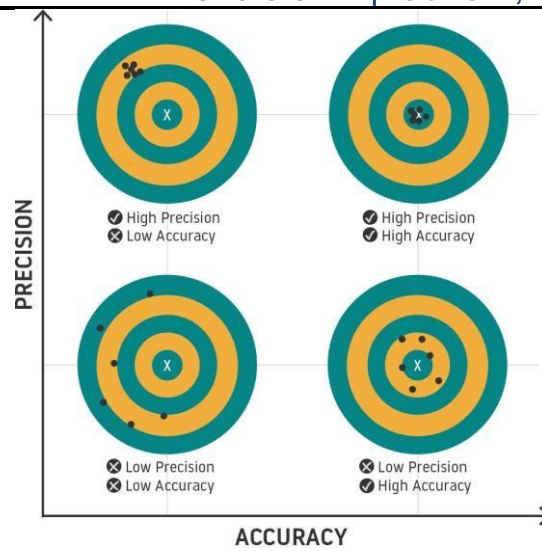
To Reduce Anxiety

Patients can have a Virtual meet with the Doctor so that they can convey their symptoms and it will diagnose and assist them in more effective manner rather than creating stress due to traditional disease predictions.

1.3.2 Objectives
Powerful Analysis Features

This system will diagnose and analyze patient's symptoms more effectively using powerful Machine learning algorithms and classifiers and can recommend best available doctors to them.

Take care of Patient's Health

Patients can set reminders for their medicinal routine checkup and also for their regular appointments with their consulting Doctors.

2. Literature Survey
2.1 Designing Disease Prediction Model Using Machine Learning Approach

Now-a-days, people face various diseases due to the environmental condition and their living habits. So, the prediction of disease at earlier stage becomes important task. But the accurate prediction on the basis of symptoms becomes too difficult for doctor.
The correct prediction of disease is the most challenging task. To overcome this problem data mining plays an important role to predict the disease. Medical science has large amount of data growth per year. Due to increase amount of data growth in medical and healthcare field the accurate analysis on medical data which has been benefits from early patient care. With the help of disease data, data mining finds hidden pattern information in the huge amount of medical data.

In this paper, they have proposed a general disease prediction based on symptoms of the patient. For the disease prediction, they used K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) machine learning algorithm for accurate prediction of disease. For disease prediction required disease symptoms dataset. Ensemble classification technique is used in this model before prediction. In this general disease prediction, the living habits of person and check-up information considerfor the accurate prediction.

Machine Learning Algorithms Used: KNN (K-Nearest Neighbor), CNN (Convolutional neural network)
Accuracy: CNN – 84.50%, KNN – 81.12%

2.1.1 Advantages of Paper
a. Precise and accurate.
b. Algorithms are working efficiently with larger datasets.

2.1.2 Disadvantages of Paper
a. Risks associated with disease are not displayed.
b. Higher Memory Requirement.
c. Less Efficiency and accuracy.
d. Time required for classification is higher about 12900ms.

2.1.3 How to overcome the problems mentioned in Paper
a. Prediction will be done at faster rate.
b. Memory Requirement will be reduced.
c. Risks associated with the diseases will be displayed.

2.2 Symptom Based Health Prediction using Data Mining and ML

The general day to day health of a person is vital for the efficient functioning of the human body. In this Paper they are taking certain prominent symptoms and their diseases to build a Machine learning model to predict common diseases based on real symptoms is the objective of their research with the dataset of the most commonly exhibited diseases, they built a relation for predicting the possible disease based on the input of symptoms.

The proposed model utilizes the capability of different Machine learning algorithms combined with text processing to achieve accurate prediction. Text processing has been implemented using Tokenization and, is combined with various algorithms to test the similarities and the outputs. In health industry, it provides several benefits such as pre-emptive detection of diseases, faster diagnosis, medical history for review of patients etc.

Training Models, ML Algorithms Used: Scikit Learn Library and Pandas data frames of Python to process data. Decision Trees, Random Forest and Naïve Bayes Algorithms.

Accuracy: Decision tree Algorithm - 98.18%, Random Forest Algorithm - 98.05%, Naïve Bayes Algorithm – 98.55%

2.2.1 Advantages of Paper

a. Less Preprocessing.

b. Strong independent Assumptions.

c. Faster due to usage of conditional probability.

2.2.2 Disadvantages of Paper

a. Less data set and number of parameters for classification.

b. Decision trees are built on the entire dataset.

c. Time required for training dataset is higher.

d. It does not show all components of the Prediction system, but it focuses on the feature of classification.

2.2.3 How to overcome the problems mentioned in Paper

a. More number of parameters will take for the accurate prediction.

b. Time required for training dataset will be reduced by doing preprocessing of the dataset for faster results.

c. Selection of best ML Algorithm will take so as to improve the efficiency and accuracy of our project in terms of all requirements.

d. Features of the predictions system will be showcased in a detailed better analysis.manner for

2.3 A Novel Approach to Predict Diabetes by Using Naive Bayes Classifier

Diabetes can be mentioned as one of the most fatal and constant sicknesses that may cause a rise in the glucose levels. In this IEEE Paper, the main target of this model is to analyze the database of diabetic patients and to predict the diabetic disease in the early stage.

In this proposed system, Naïve Bayes Classification is used for predicting the diabetes. Information mining is a procedure extricating data from a dataset and change it into justifiable structure for additional utilization. The information arrangement is diabetic patients informational collection is created by gathering information from clinic storehouse comprises of 1865 occurrences with various qualities. The outcomes show that the proposed novel strategy can foresee the diabetes with higher exactness levels than the customary/existing techniques.

In this proposed system using Naïve Bayes Classifier, Output will be the Web Interface showing the Outcome of having diabetes or not by taking the input values like Insulin level, age and so on. This increases the accuracy of the system.

Machine Learning Algorithms Used: Naïve Bayes Accuracy: 96%

2.3.1 Advantages of Paper

a. Simple and easy to predict.

b. Performs well due to multiple class expectation.

c. Accuracy is highest.

2.3.2 Disadvantages of Paper

a. Implicitly assumes that all the attributes are mutually independent.

b. The assumption of independent predictor features.

2.3.3 How to overcome the problems mentioned in Paper

a. Prediction will be done faster and with less memory requirement and classification will be done more precisely with appropriate symptoms.

b. Risk associated with the disease will be showcased after prediction.

c. Selection of best ML Algorithms will be taken for development.

d. Implementation of Live Consultation with Doctors for better treatment and diagnosing of diseases.

e. Reduce the limitations of less accurate Algorithms by implementing them in a Sequenced manner.

f. Specialized Doctor Recommendation and nearest health centers from your location available for quick treatment.

2.4 Technical Review

We have used Flask Framework to build the functioning of our webapp and Machine Learning algorithms like Naive Bayes, K-Nearest Neighbor, Kernel SVM, Decision Trees and Random Forest Algorithms for accurate pre-diction and analysis of diseases.

2.4.1 Advantages of Technology
a. Flask is implemented in Python
b. Better CDN connectivity and Content Management
c. Batteries Included Framework
d. Fast Processing
e. Naive Bayes Algorithm is simple and easy to implement. It doesn't require as much training data. It handles both continuous and discrete data

2.4.2 Reasons to use this Technology
a. Flask Offers Rapid-development and provides Security
b. Flask has a Model-View-Controller (MVC) architecture. The MVC (Model View Controller) is a software design pattern.
c. It is a collection of three essential components Model, View, and Template. These three layers are responsible for different things and can be used independently.
d. Naive Bayes Algorithm is highly scalable with the number of predictors and data points.
e. It is fast and can be used to make real-time predictions
f. It is not sensitive to irrelevant features
g. Naive Bayes Algorithm performs better than other models with less training data if the assumption of independence of features holds.
h. If you have categorical input variables, the Naive Bayes algorithm performs exceptionally well in comparison to numerical variables.

3. Project Planning
3.1 Roles and Responsibilities

SR. No	Name of Member	Role	Responsibilities
1	Habib Khan	Team Leader	Flask Integration, ML Model Design
2	Sachita Dodke	Member	Frontend, UI Design, Backend
3	Sanskruti Sonwal	Member	UI / UX Design
4	Nishant Walunj	Member	ML Model Training and Design

3.2Assumptions and Constraints
• Users of this Webapp can use this system.
•Users of this Webapp should have active internet.
• Users should provide appropriate symptoms for prediction.
• There will be a admin present 24/7 at the server.
• Highest accuracy is fetched with more than 2 or 3 symptoms.
• We have to work with the available resources.
• We need to manage the entire project within the team of developers.

3.3 Project Management Approach
    We will be working in an agile project management approach. As Agile is a project management methodology that uses short development cycles called "sprints" to focus on continuous improvement in the development of a product or service. We will be using the following principles while working:
• Changing environments are embraced at any stage of the process to provide the customer with a competitive advantage.
• We will deliver a product or service with higher frequency.
• Our Stakeholders and developers collaborate closely on a daily basis.
• All stakeholders and team members remain motivated for optimal project out- comes, while teams are provided with all the   necessary tools and support, and are trusted to accomplish project goals.
• Face-to-face meetings are deemed the most efficient and effective format for project success.
• A final working product is the ultimate measure of success.
• Sustainable development is accomplished through agile processes whereby development teams and stakeholders are able to maintain a constant and ongoing pace.
• Agility is enhanced through a continuous focus on technical excellence and proper design.
• Simplicity is an essential element.
• Self-organizing teams are most likely to develop the best architectures and designs and to meet requirements. Regular intervals are used by teams to improve efficiency through fine-tuning behaviors.

3.4 Ground Rules for the Project
• Be on time for all team meetings.
• Team leader must create and disseminate agendas for each team meeting.
• Team leader must create and disseminate minutes after each team meeting.
• Attend full duration of all team meetings unless a case of emergency.
• Avoid informal/social talk during team meetings.
• Avoid apathetic/passive decision making (e.g., "whatever you all think is right").
• Inform team leader if unable to complete work on time.
• Set deadlines for each deliverable in advance of due date to allow for collaborative revisions.
• Rotate responsibilities so each person gets experience with several aspects regardless of quality or qualifications.
• Make criticisms constructive with suggestions for improvement and non-judgmental language.

• Confront issues directly and promptly.
• Promptly relay all interpersonal concerns/conflicts to team leader.
• Keep a positive attitude toward the team, individual members, projects and course.
• Take initiative by offering ideas and volunteering for tasks.
• Play an equal role in the team by contributing equally to every task.
• Be honest with any team member who is not pulling her/his weight.
• Help one another with difficult or time-consuming deliverables.
• Ask for help from the team or other resources if "stuck" or falling behind.
• Treat each other with respect.
• Accept responsibility and accountability along with the authority given.

3.5 Project Budget
• It is cost efficient Project.
• Easily deployable across compatible devices.

4. Software Requirements Specification

4.1 Overall Description
4.1.1 Product Perspective
   Our disease prediction system web application is based that tries to ease the development process by simplifying repetitive tasks used in most of today's web applications, including routing, authentication, caching and sessions. The vertically integrated web development environment is meant to offer an improved and smooth workflow of data.

4.1.2 Product Features
The following are the main features that are included in our Smart Health Prediction System:
1. Cross platform support: Offers operating support for most of the known and commercial operating systems.
2. Number of users being supported by the system: Though the number is precisely not mentioned, the system is able to support   a large number of online users at a time.
 3. Analysis summary: Provides a dashboard to review the results of disease prediction based on their entered input symptoms.

4.1.3 Characteristics
Doctors can view the details of patient about the disease and also the symptoms that mentioned while predicting through our portals so that doctors can study more while contacting with the patient.

4.1.4 Operating Environment
The application is developed in the Flask framework to enable the creation of a web-based application, which can be accessed from any web browser.

4.1.5 Design and Implementation Constraints
a. The application should run in a latest JavaScript enabled web browser.
b. The application might take a few seconds to load the analysis data generated by the API.
c. This system is provisioned to be built on the Flask framework which is highly flexible.

4.2 System Features
4.2.1 Accurate Prediction with faster results
Smart Disease Prediction will predict accurate diseases based on the user's symptoms with full-fledged efficiency and swiftness. SDP will also show-case the consequences of that disease and the impact of that disease on other organs of the body. SDP will work on the fastest Machine Learning Algorithm with the data set and it will predict the diseases with its powerful prediction capabilities.

4.2.2 Feedback
User can send feedback regarding any issue related to the web application by typing message in our contact us section.

4.3 External Interface Requirements
4.3.1 User Interfaces
a. Front-end software: HTML, CSS, JavaScript, Bootstrap, jQuery
b. Back-end software: Flask Framework, Python

4.3.2 Hardware Interfaces
a. Windows or Mac operating systems
b. Android or IOS mobile phones.
c. Devices should be enabled with the Internet.

4.3.3 Software Interfaces
a. The user's browser should be HTML5 and JavaScript compatible for all the functionalities to work.

4.4 Nonfunctional Requirements

4.4.1 Performance Requirements

a. The dashboard page is displayed to the user immediately. It takes few seconds to display the predicted data based on different symptoms and training of date to predict disease.

b. The disease prediction part of the application is flexible and smooth so that it does not consume any time of the user.

5. System Design

5.1 System Requirements Definition

System requirement definitions specify what the system should do, its functionality and its essential and desirable system properties. The techniques applied to elicit and collect information in order to create system specifications and requirement definitions involve consultations, interviews, requirements workshop with customers and end users. The objective of the requirements definition phase is to derive the two types of requirements.

Data-flow Diagram

Level 0

DFD Level 0 for our systems focuses on the main output where patient gets the result of his/her disease analysis.
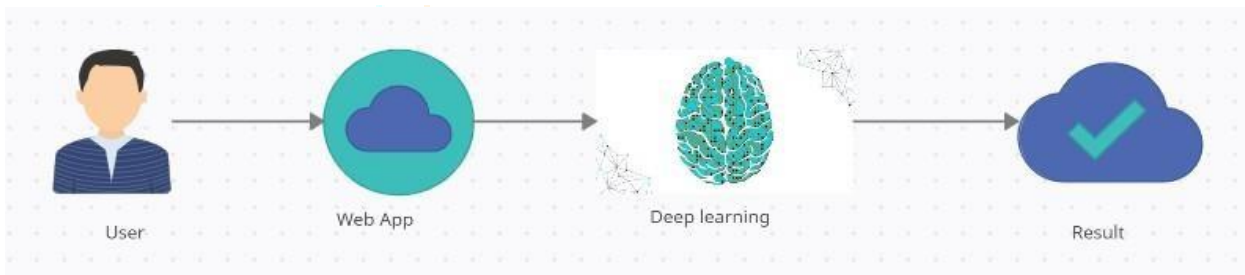
DFD Level 0 for SDP

Level 1

DFD level 1 depicts that Patients symptom will be saved in database as well as sent for classification and prediction accompanied by trained Machine Learning Model.
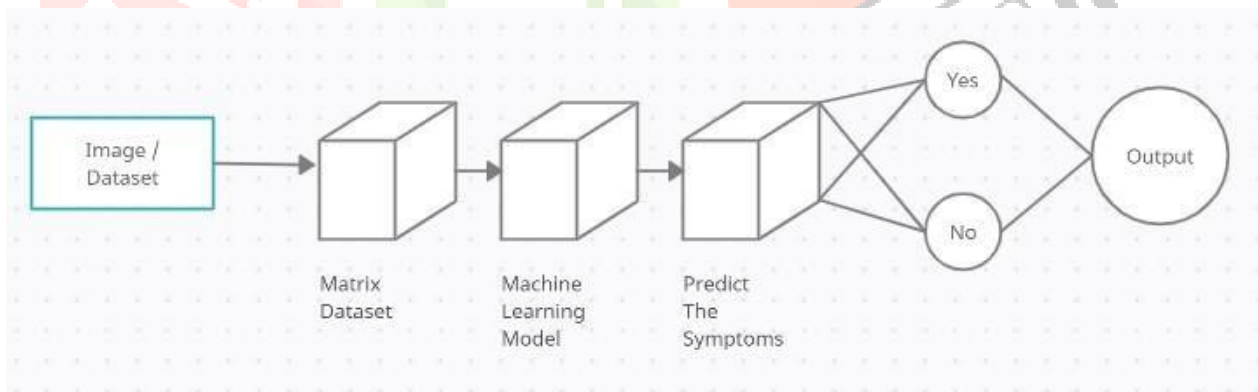
DFD Level 1 for SDP

Level 2

DFD Level 2 the data will be analyzed and a complete prediction will be made. The predicted disease will be accompanied by some of its information as well as a doctor will be referred according to the classified disease
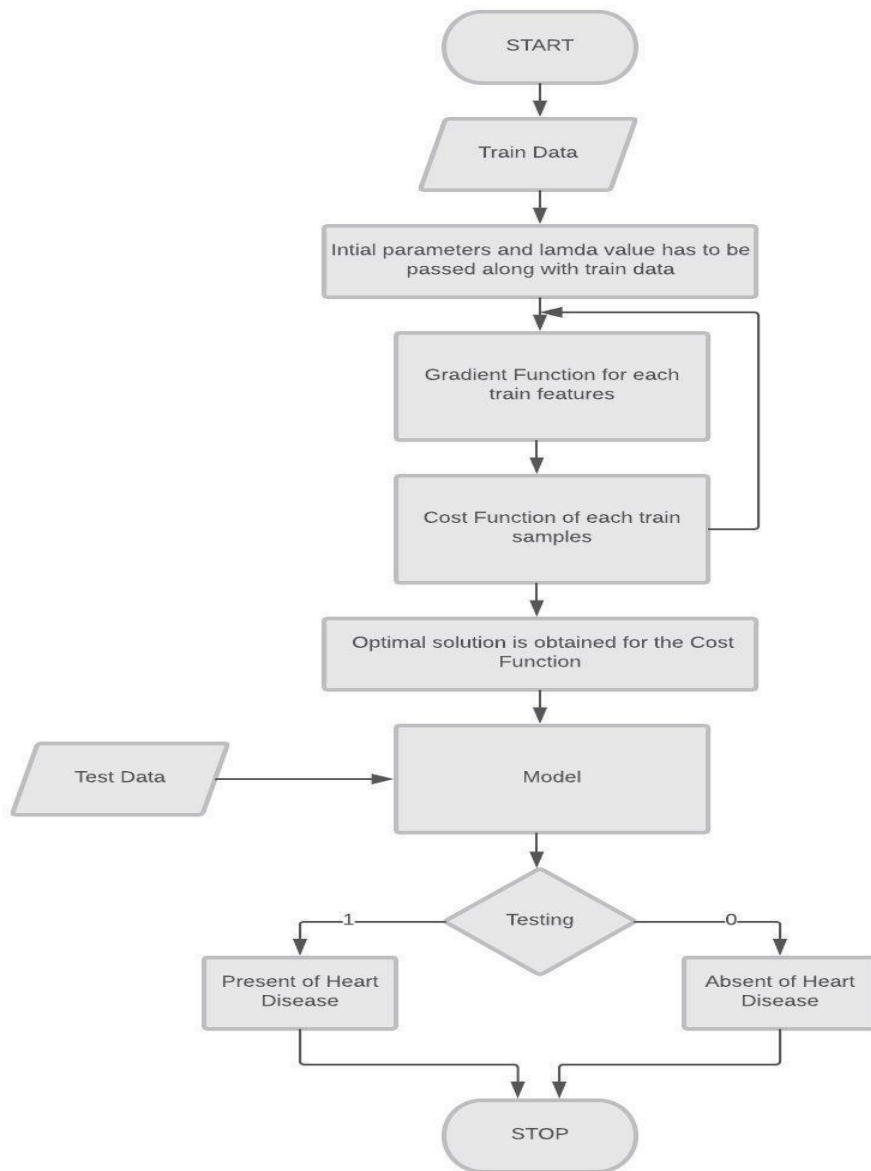
DFD Level 2 for SDP

5.2 System Architecture Design

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.
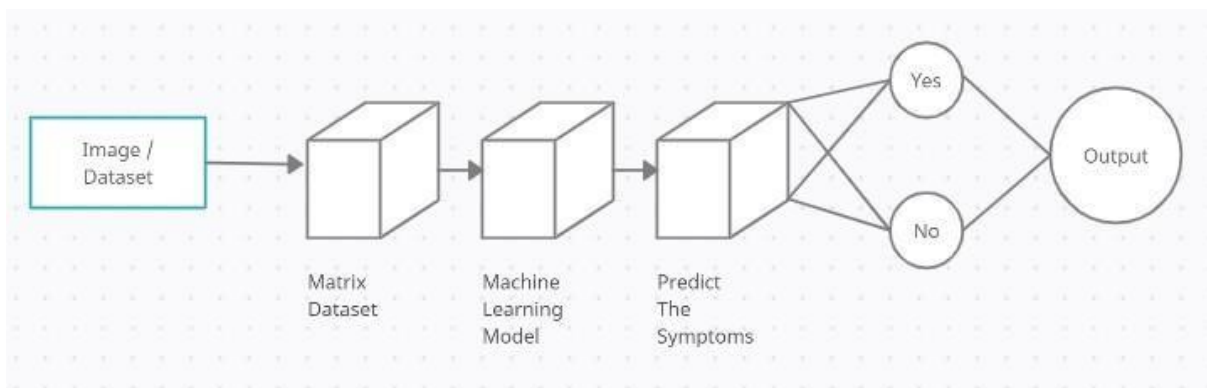
System Architecture Design Diagram for SDP

5.3 Sub-system Development

This system consist of four modules Data Collection Module, Disease Prediction Module and Web Module. In Data collection Module user symptoms input will be taken and based on that input symptoms Disease Prediction Module will analyses the disease probability and that will display through Web Module.

5.3.1 Data Collection Module

In Data Collection Module Doctor will enter the symptoms of patient and it will be further sent to Disease Prediction Module to predict the result.
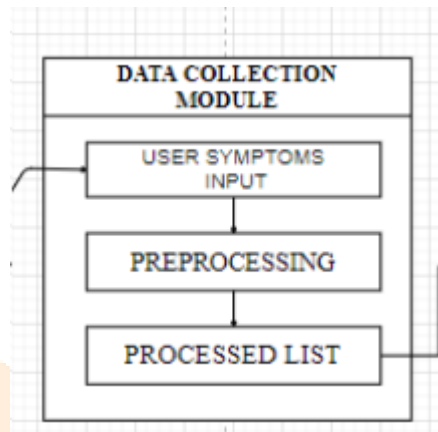
Data Collection Module

5.3.2 Disease Prediction Module

In Data prediction Module there a prediction of disease will be done with the help of algorithms provided in back-end. The better probability of disease prediction result will be selected for further procedure.
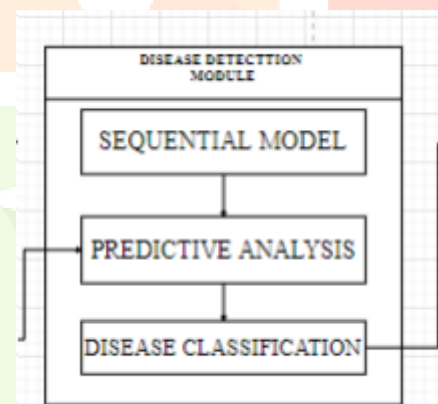
Disease Prediction Module

5.3.3 Web Module

In web module the result will be displayed to the doctor so they can further work on patient treatment and predicted through previous module and they will contact with each other due this module in which we design a chat body system for them.
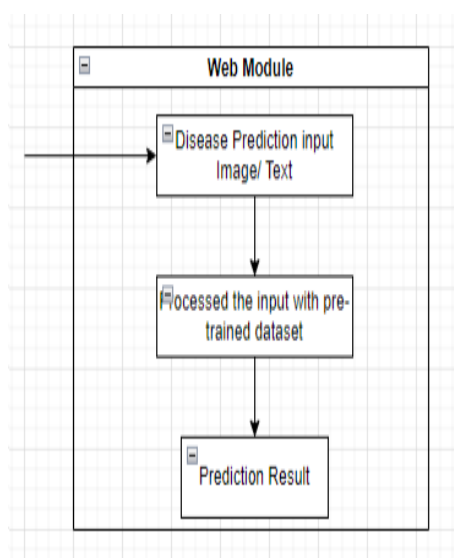
Web Module

6. Implementation

6.1 Disease Prediction

Disease Prediction the core and sole feature of our project named Smart Disease Prediction. The main goal is to provide powerful analysis features with utmost accuracy in predicting diseases because health is an essential factor with which no one is willing to take any risk. As we know that Health is the most important asset for a human being. So to maintain a healthy lifestyle by anticipating or predicting diseases and impact of them on other organs at early stage would be more beneficial and constructive in analysis for the treatment of that dis-ease. To overcome this Real-life problem, we developed a Smart Health Disease Predictor with the help of Machine Learning.

Smart Disease Prediction will predict accurate diseases based on the user's symptoms with full-fledged efficiency and swiftness. SDP will also showcase the consequences of that disease and the impact of that disease on other organs of the body. SDP will work on the fastest Machine Learning Algorithm with the data set and it will predict the diseases with its powerful prediction capabilities.

The precise goal of our project is accuracy. Since our system deals with problems related to health and it could be sometimes serious if it lacks accuracy hence, we will try to achieve as much accuracy as we can in comparison with the systems previously proposed. Our system will not only bound till prediction we will try to make it more User specific so that both patient's as well medical professionals can seek profit through our system.

6.2 Machine Learning Module

In our system we have used immense power of Machine learning algorithms top redict diseases for the end user on the basis of symptoms which they provide with the confidence score with which it is predicting. We have implemented six Ma- chine
Learning algorithms and trained them in so that we can use them in our future endeavors. The Algorithms implemented are as follows:-
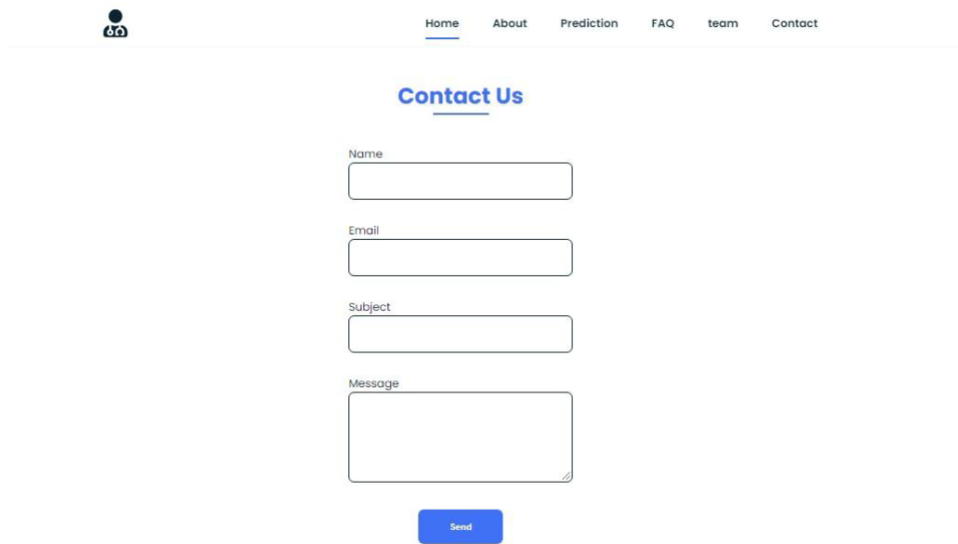1. Logistic Regression
2. K - Nearest Neighbors
3. Naive Bayes
4. Decision Tree
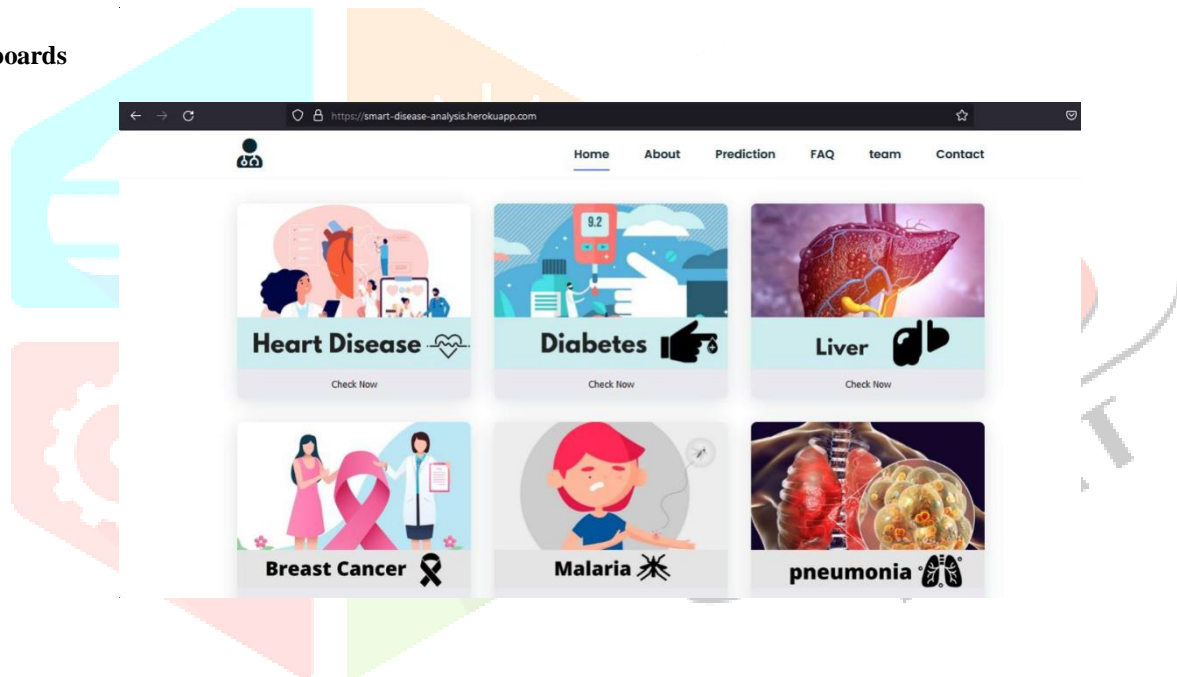5. Random Forest

7. Screenshots of Project
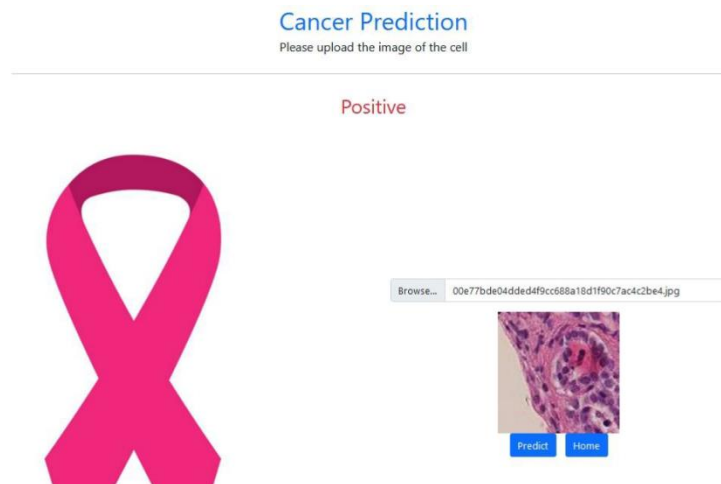
7.1 Home Page
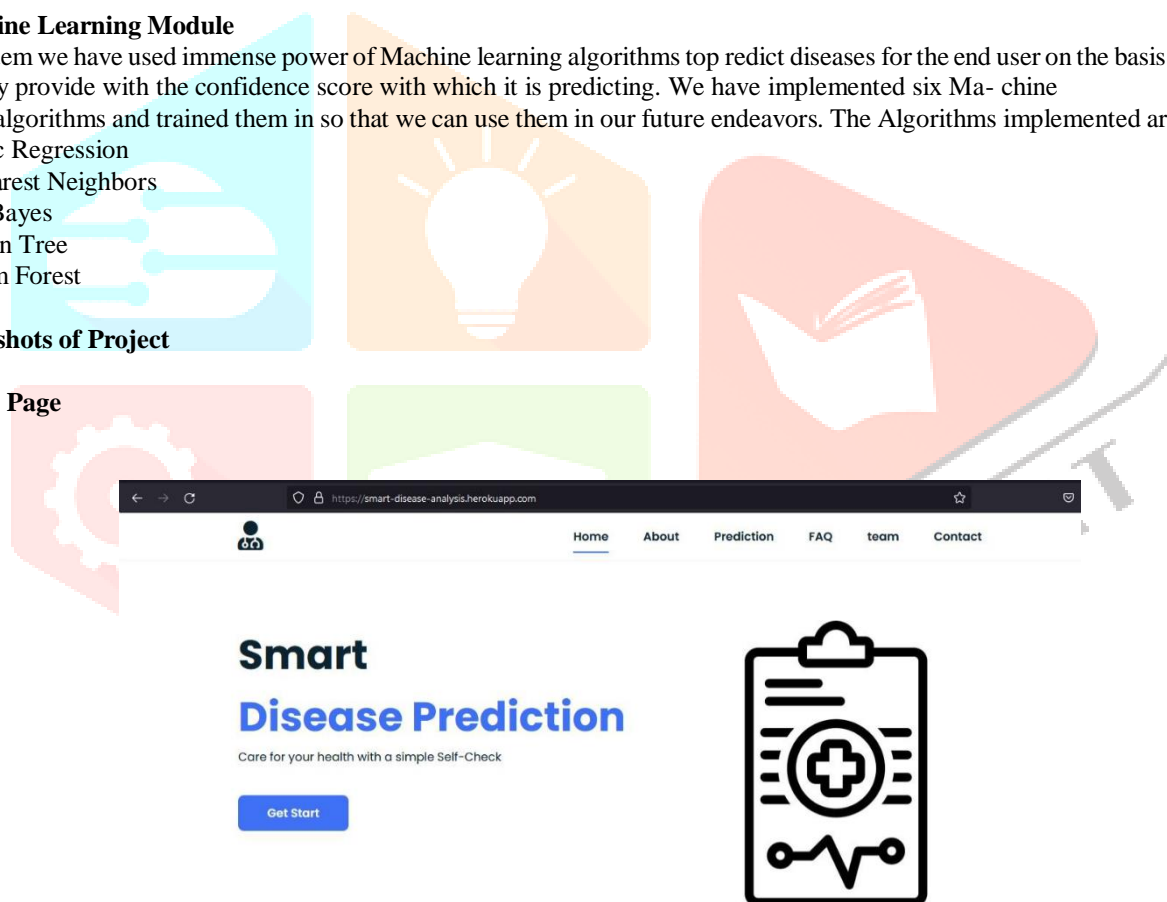
7.2 Contact Page

7.3 Dashboards

7.4 Disease Prediction

Malaria Prediction

Please upload the image of the cell

The sample is positive

Choose File C100P61ThinF_IMG_20150918_144104_cell_162.png

Predict Home

Pneumonia Prediction

Please upload the X-ray of lungs

The sample is positive

Pneumonia

Choose File person100_bacteria_477.jpeg

Predict Home

Liver Disease Prediction

Age

Enter the Age

Total Bilirubin

Enter the Total Bilirubin

Direct Bilirubin

Enter the Direct_Bilirubin

Alkaline Phosphotase

Enter the Alkaline Phosphotase

Alamine Aminotransferase

Enter the Alamine Aminotransferase

Aspartate Aminotransferase

Enter the Aspartate Aminotransferase

Total Protiens

Enter the Total Protiens

8. Conclusion and Future Scope

8.1 Conclusion

As we have seen the real importance and value of Health and lives in the current ongoing pandemic. There is no wealth more than a Healthy life. As a part of serving the common people and solving their health-related problems was our motto throughout the Development of this project.

We developed this Webapp for predicting and analyzing the diseases using various Machine Learning Algorithms for early cure and ailment of diseases based on the symptoms provided by the user which indeed became very fruitful and successful by providing highest accuracy and accurate prediction. We also facilitated our users with live consultation with doctors in ongoing pandemic so that they can get proper medical care from their home.

This project is a well-planned Combo for Generic Health care with best optimal solutions be it in predicting a disease and analyzing it. Also getting the information of the risks associated to other organs of body from predicted disease. This project also recommended Specialist Doctor for communication and getting cure as early as possible.

This project involved our immense and rigorous efforts for development and resolving the hurdles in day- t o - d a y health problems of people. We would like to make more Projects for the help of needy people and also for the upliftment of society in future.

8.2 Future Scope

We wish to keep working on our project to make it a marketable and leading product. Some of the additions that we feel are needed for it to happen are:
• Making it deployable on Cloud for global accessibility.
• Increasing the dataset by using our Disease Addition Program for predicting more diseases.
• Making User Interface more User Friendly.
• Adding more functionalities for Real Time Health Care using various parameters like BP, Sugar Level, Body Temperature and RBC WBC Readings, etc.
• Customizing this project with live Health Blogs using data Scrapping.