"Detection of Diseases using Machine Learning"

LY Project Report

**Submitted in partial fulfillment of the requirements of the Degree of
Bachelor of Technology in Electronics Engineering**

by

Raj Ghadi

Shrejay Patil

Rohit Desai

Project Guide
Prof. Pankaj Deshmukh

Department of Electronics Engineering

K.J Somaiya Institute of Engineering and Information Technology

Ayurvihar, Sion, Mumbai -400022

2021-22

"Detection of Diseases using Machine Learning"

LY Project Report

Submitted in partial fulfillment of the requirements of the Degree of Bachelor of Technology in Electronics Engineering

by

Raj Ghadi (Roll No: 23)

Shrejay Patil (Roll No: 59)

Rohit Desai (Roll No: 14)

Project Guide
Prof. Pankaj Deshmukh

Department of Electronics Engineering

K.J Somaiya Institute of Engineering and Information Technology

Ayurvihar, Sion, Mumbai -400022

2021-22

CERTIFICATE

*This is to certify that the project entitled **"Detection of Diseases using Machine Learning"** is a bonafide work of Student Raj Ghadi, Shrejay Patil and Rohit Desai submitted to University of Mumbai in partial fulfillment of the requirement for the Award of the degree of Bachelor of Technology in Electronics Engineering.*

_____

Prof. Pankaj Deshmukh

Project Guide

_____   _____

Dr. Milind Nemade   Dr. Suresh Ukrande

Head of Department   Principle

Date: -

Place: - Mumbai

PROJECT APPROVAL FOR L.Y.

This project report entitled **"Detection of Diseases using Machine Learning"** by

Raj Ghadi (Roll No 23)

Shrejay Patil (Roll No 59)

Rohit Desai (Roll No 14)

is approved for the degree of Bachelor of Technology in Electronics Engineering.

Examiners

1. _____

2. _____

Date:

Place: Mumbai

Declaration

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

Raj Ghadi _____

Shrejay Patil _____

Rohit Desai _____

(Signature)

Date:

Place: Mumbai

ACKNOWLEDGEMENT

We like to share our sincere gratitude to all those who help us in completion of this project. During the work we faced many challenges due to our lack of knowledge and experience but these people help us to get over from all the difficulties and in final compilation of our idea to a shaped sculpture.

We would like to thank Prof. Pankaj Deshmukh sir for his governance and guidance, because of which our whole team was able to learn the minute aspects of a project work.

We would also like to show our gratitude to our Project Coordinators Prof. Pankaj Deshmukh and Mr. Sharan Rahul Raghvendra (Alumnus) for their continuous help and monitoring during the project work.

In the last we would like to thank the management of K J Somaiya Institute of Engineering and Information Technology for providing us such an opportunity to learn from these experiences.

All of our team is thankful to all the Faculties and Staff of Department of Electronics Engineering, for their help and support towards this project and our team.

We are also thankful to our whole class and most of all to our parents who have inspired us to face all the challenges and win all the hurdles in life.

Thank you All.

<div align="right">

Raj Ghadi

Shrejay Patil

Rohit Desai

</div>

ABSTRACT

The world's growing population has put enormous pressure on the healthcare sector to offer high- quality treatment and services. Artificial Intelligence and Machine Learning are no exceptions in the healthcare industry, which has long been a strong supporter of cutting-edge technology. To relieve the burden on the healthcare system and to help a doctors and society, we have developed a project that will predict specific diseases, detect them sooner, and classify them more accurately and efficiently.

We use technologies such as machine learning, Python, Flask, and web development in this project, and we created a web application (Software) using all of them. This Machine Learning-based web application (Software) aids us in accurately predicting and classifying diseases such as Diabetes, Liver, Kidney, Heart, Breast Cancer, and Parkinson's disease. We simply need to enter the patient's information and medical data into the web application (software), and the software will generate results on the same page.

Our primary objective and outcome for this project will be to accurately and effectively predict diseases, to detect whether a person has this specific disease or not, to detect the disease at an earlier stage, and to classify the disease type. To help doctors in making vital decisions, to reduce human error in the healthcare system, and to assist them in saving more patients' lives or extending their lifespan.

K. J. Somaiya Institute of Engineering and Information Technology
Sion, Mumbai

(Accredited with 'A' Grade by NAAC, Accredited Three programs by NBA (Computer Engineering, Electronics Engineering and Electronics & Telecommunication Engineering, Permanently Affiliated to University of Mumbai)

Department of Electronics Engineering Course Outcomes of Major Project

Course Outcomes of Major Project

After taking this course, the student will be able to	PO Satisfied
1. **Design solutions, components, or processes for complex engineering problems based on research knowledge**	PO1, PO2, PO3
2. Select appropriate modern engineering tools and analyze and interpret data to meet the problem statement.	PO4 and PO5
3. Apply contextual knowledge to assess **public health, safety, societal, environmental issues** and sustainable development.	PO6, PO7
4. Comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	PO10
5. Apply ethical principles and commit to professional ethics, responsibilities norms of the engineering practice, and engage in independent and life-long learning.	PO8, PO12
6. Interact efficiently and effectively as an individual with the team members or leader for **timely and ethical management** of projects	PO9, PO11

CO-PO Mapping:

Map COs with POs, POs satisfied by respective COs are mentioned in the above table.

In the table given below, you have to map COs with POs in either of the three levels. Write levels against every COs. Write "- "if CO not mapping with PO.

"1": For less mapping, "2": Moderate mapping "3": Strong mapping

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	3	2	2	-	1	2	2	2	2	2	2	2
CO2	1	1	3	2	3	2	-	1	2	1	3	2	3	3	3
CO3	3	3	3	1	2	3	-	2	1	1	2	3	3	3	3
CO4	2	2	2	2	1	2	-	3	2	3	2	3	1	1	1
CO5	2	2	1	1	2	1	-	3	3	2	2	2	1	1	1
CO6	2	2	1	2	2	1	-	3	2	2	1	1	1	1	1

Course Outcomes

CO/PO mapping	Proof required
CO1	Literature review chapter, detailed problem statement based on research knowledge, discussion on results
CO2	Hardware and software used for project
CO3	Include a chapter on the Introduction and Context of the project (Context related to public **health, safety, societal, environmental issues,** and sustainable development)
CO4	Project Report writing and presentation
CO5	Plagiarism check certificate, proper citations in the report, acknowledgment of figures, data, observations in rubrics
CO6	Timely completion- marks awarded in rubrics, Budget documentation- Logbook maintenance

Program Outcomes

Program Outcomes	Explanation
PO_01 Engineering Knowledge (EK):	Applied the knowledge of Programming, Databases, Algorithms and data structures to develop this project.
PO_02 Problem Analysis (PA):	We Performed a literature review of various papers on diseases predication and the researches done to solve them. Also, analyzed the existing systems available.
PO_03 Design/Development of Solutions (DS):	Designed machine learning based software to predict the diseases and help our doctors to cross check their assumptions.
PO_04 Conduct Investigations of Complex Problems (IP):	Studied and investigates the various research papers to learn about the various machine learning algorithms and techniques to apply it on particular diseases effectively.
PO_05 Modern Tool Usage (MT):	Used modern tools such as Machine Learning to predict the diseases. Used PyCharm for the whole project development.
PO_06 The Engineer and Society (ES):	Applied reasoning is informed by the contextual knowledge to assess societal, health, safety, legal and agricultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO_07 Environment and Sustainability (Env):	The development of the project was done by ensuring that we are not harming the environment in any way.
PO_08 Ethics (Eth):	Applied ethical principles and the whole team was committed to professional ethics and responsibilities and norms of the engineering practice.
PO_09 Individual and Term Work (ITw):	We divided the project development into various sections and each section was handled by one member and the documentation was done with the participation of all the members respecting everyone's opinion.
PO_10 Communication (Cm):	Good communication between the team members was a key aspect of our project as each member was responsible for a particular section and at the end, we needed to put all the sections together to complete the entire project and present it effectively. We also communicated with various industry people about some topics which we had stuck upon.
PO_11 Project Management and Finance (PM):	As the mentioned project is a medical project, and hosted on cloud hence, it is publicly available to everyone and easy to access at free of cost. This open-source software project will help the doctors to a great extent.
PO_12 Life-long Learning (LL):	Recognized the need for using modern technologies in the healthcare sector to minimize the medical errors which will indirectly reduce deaths of the peoples of our country.

Program Specific Outcomes (PSO)

PSO_1:	Design and analyse software systems for Diseases Predication Applications.
PSO_2:	Identify and apply domain specific tools for design of Machine Learning based software.
PSO_3:	Develop cost effective software web application for doctor's use.

CONTENTS

List of Abbreviations

Abbreviations	Meaning	Page No
Hz	The number of periods or cycles per second is called frequency. The SI unit for frequency is the hertz (Hz)	21
mm	A millimeter (abbreviated as mm and sometimes spelled as millimeter) is a small unit of displacement (length/distance) in the metric system.	21
4dB	The decibel is a relative unit of measurement equal to one tenth of a bel. It expresses the ratio of two values of a power or root-power quantity on a logarithmic scale. Two signals whose levels differ by one decibel have a power ratio of 101/10 or root-power ratio of $10^{1/20}$.	21
mg/ dl	According to the American Diabetes Association, mg/dL (MG/DL) stands for milligrams per deciliter, which is a unit of measure that shows the concentration of a substance in a specific amount of fluid. In the United States, blood glucose test results are reported as mg/dL	21
IU/L	An international unit is an arbitrary amount of a substance agreed upon by scientists and doctors; a liter is a measure of volume that is slightly greater than a quart. Some medical tests report results using international units (IU) per liter (L).	21
dl	Some medical tests report results in grams (g) per deciliter (dL). A gram is equal to the weight of one milliliter or 16 drops of water. It is about 1/30 of an ounce.	21
mm Hg	A millimeter of mercury is a manometric unit of pressure, formerly defined as the extra pressure generated by a column of mercury one millimeter high, and currently defined as exactly 133.322387415 pascals. It is denoted mmHg or mm Hg.	21
kg/ m^2	The BMI is defined as the body mass divided by the square of the body height, and is expressed in units of kg/m2, resulting from mass in kilograms and height in meters.	21

List of Figures

List of Tables

1. Introduction

The ever-incrementing population of the world has put tremendous pressure on the healthcare sector to provide quality treatment and healthcare accommodations. Now, more than ever, people are authoritatively mandating keenly intellective healthcare accommodations, applications, and wearables that will avail them to lead better lives and perpetuate their lifespan.

The healthcare sector has always been one of the greatest proponents of innovative technology, and Artificial Perspicacity and Machine Learning are no exceptions. In fact, Machine Learning has come to play a pivotal role in the realm of healthcare – from ameliorating the distribution system of healthcare accommodations, cutting down costs, and handling patient data to the development of incipient treatment procedures and drugs, remote monitoring and so much more.

Withal, the fact that the healthcare sector's data burden is incrementing by the minute (owing to the ever-growing population and higher incidence of diseases) is making it all the more essential to incorporate Machine Learning into its canvas. With Machine Learning, there are illimitable possibilities. Through its cutting-edge applications, ML is availing transform the healthcare industry for the preponderant.

To minimize the pressure of the healthcare system and to avail the medicos and society we have engendered the project which will prognosticate the particular disease, detect the disease in earlier stage and relegate the diseases very accurately and efficiently. This will avail medicos to substantiate or cross- check their posit and analysis. It will avail them in critical situations and decisions.

2. Literature Survey

2.1 Literature Survey

1. University of California – Irvine, December 17, 2020 "Researchers engender model to calculate COVID-19 health outcomes"

University of California, Irvine health sciences researchers have engendered a machine-learning model to prognosticate the probability that a COVID-19 patient will require a ventilator or ICU care. The implement is free and available online for any healthcare organization to utilize.[1]

"The goal is to give an earlier alert to clinicians to identify patients who may be vulnerably susceptible at the onset," verbally expressed Daniel S. Chow, an assistant pedagogia in residence in radiological sciences and first author of the study, published in PLOS ONE. The implement presages whether a patient's condition will worsen within 72 hours.[1]

Coupled with decision-making concrete to the healthcare setting in which the implement is utilized, the model utilizes a patient's medical history to determine who can be sent home and who will require critical care. The study found that at UCI Health, the implement's predictions were precise about 95 percent of the time.[1]

Fig 2.1 Researchers create model to calculate COVID-19 health outcomes

2. Marouane Ferjani, December 2020 "Disease Presage Utilizing Machine Learning"

The wide adaptation of computer-predicated technology in the health care industry result in the accumulation of electronic data. Due to the substantial amplitudes of data, medical medicos are facing challenges to analyze symptoms accurately and identify diseases at an early stage. However, supervised machine learning (ML) algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and availing medical experts in the early detection of high- risk diseases. In this literature, the aim is to apperceive trends across sundry types of supervised ML models in disease detection through the examination of performance metrics.

The most prominently discussed supervised ML algorithms were Ingenuous Bayes (NB), Decision Trees (DT), K-Most proximate Neighbour (KNN). As per findings, Support Vector Machine (SVM) is the most adequate at detecting kidney diseases and Parkinson's disease. The Logistic Regression (LR) performed highly at the presage of heart diseases. Conclusively, Desultory Forest (RF), and Convolutional Neural Networks (CNN)prognosticated in precision breast diseases and mundane diseases.

ML algorithms are highly susceptible to errors because of two factors. Firstly, it depends on the quality and the selection of the datasets, which is crucial to achieve accurate and unbiased decisions. Secondly, ML algorithms relies heavily on the right selection of features extracted from the dataset, which proved to be difficult, time consuming, and required high computational power. These factors hinder the performance of the learning model and generate fatal errors that can endanger the lives of patients. [2]

2.2 Comparative Survey of Existing system

Pr. No.	Author	Topic Name	Use Case	Algorithm	Technology	Learning
1	University of California – Irvine [1]	Researchers created a model to predict whether a patient needs ICU, general ward or can be sent home.	COVID-19	Ensemble and tree-based models	Software Development, Machine Learning, Data Analysis, Data Mining and Pre-Processing Techniques	The idea to develop one open-source software that will help our doctors to predict the diseases and to cross their assumptions.
2	Marouane Fergana [2]	Disease Prediction Using Machine Learning	Heart Disease Common Diseases	Logistic Regression Random Forest	Machine Learning, Data Pre-Processing techniques	Learn about the different types of algorithms that should be used to predict specific diseases.

Table 2.1 Comparative Survey of Existing system outcomes

2.3 Limitation of the Existing system or research gap

1) Can only be used by doctors or healthcare professionals.

The data which is asked while using this web application is the data which is only available and understandable by the medical professionals as it consists of complex parameters such as enzymes contain, blood report and diabetes etc.

2) Results of the prediction are totally dependent on the user's input.

Proper data input while using the application is also important, it can be considered as human error, hence leading to false report.

3) The prediction accuracy is approximately 80% only.

Accuracy of system depends on many factors such as amount of data, quality of data, repeated data and computational or mathematical boundaries etc.

4) One cannot take decision or treatment only through these predictions, it just provides some insight to the user about possibilities of health issues.

5) After getting an insight about the health, one must consult the doctor for the same and continue with the treatment.

3. Proposed System

3.1 Proposed System

World is suffering because of many diseases and to surmount it we can utilize artificial Perspicacity and machine learning.

The purport of artificial Intelligence and machine learning is to make the machine more prosperous, efficient, precise and reliable than afore. However, in a healthcare system, it will definitely avail medico a lot in critical situations and decisions.

To minimize the pressure of the healthcare system and to avail the medicos and society we have engendered the project which will presage the particular disease, detect the disease in earlier stage and relegate the diseases very accurately and efficiently. This will avail medicos to attest or cross- check their postulation and analysis. It will avail them in critical situations and decisions.

Proposed Diseases prediction System to Improve quality and efficiency of current healthcare system. "It is introduced with the aid of web creation. In order to prevent pitfalls in the current system, the proposed system is built as a web-based system where it can be accessed at anytime and anywhere on their mobile/PC. The user can use their credential to access the server system. The website faculties will now keep track of user's health by providing insights of health issues if any. Additional features include generation of report.

A login page is a web page that needs user identification and authentication by entering a username and password combination on a regular basis. Logins can provide access to the entire website. Logging in not only gives the user access to the site, but also enables the website to monitor user activities and behavior. Logging off a website or site may be a manual for the user or may occur automatically when such circumstances (such as page closure, device shutdown, long delay, etc.) occur.

3.2 Problem Statement

• Around 15% deaths occur because of the medical errors. In the medical errors includes the erroneous diagnosis, erroneous prescription and erroneous treatment.

• World Health Organization verbally expressed that "Globally, as many as 4 in 10 patients are harmed in primary and outpatient health care. Up to 80% of harm is preventable. The most detrimental errors are cognate to diagnosis, prescription and the utilization of medicines."

• Use of Artificial Perspicacity and Machine Learning can avail the healthcare sector to truncate all these human errors and to preserve more lives.
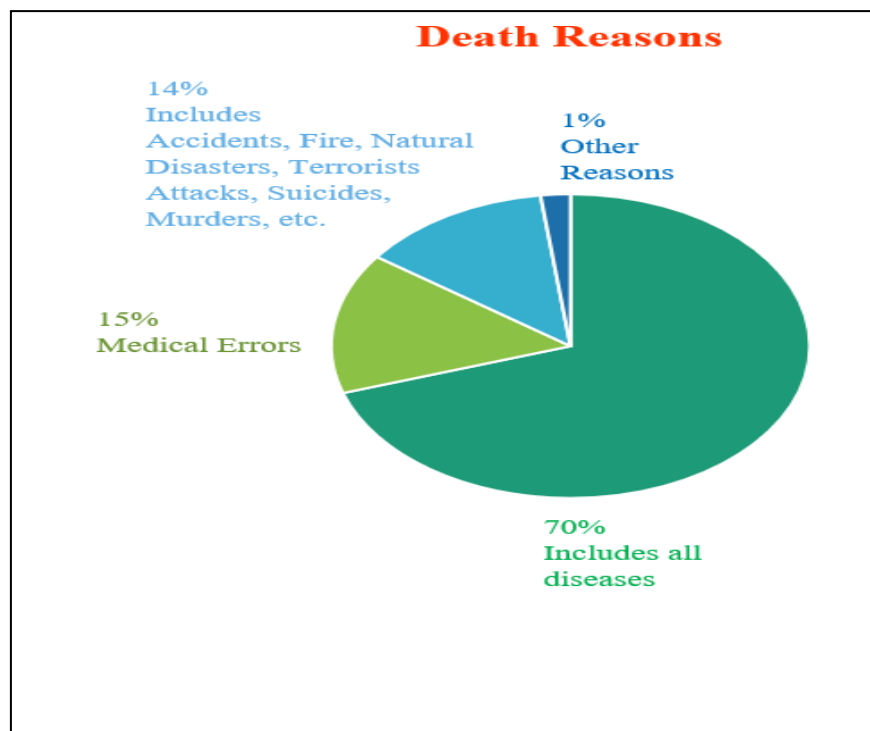

Fig 3.1 Death Reasons

3.3 Objectives

- Prediction of specific diseases (Heart, Liver, Diabetes, Cancer, Parkinson's, Kidney)

- Accuracy of prediction must be more than 70%.

- Detection of diseases in an earlier stage.

- Increase the speed of diagnosis and efficiency in the healthcare system.

- Provide the best user interface.

3.4 Methodology

The life cycle of our machine learning project we have represented as a multi-component flow, where each consecutive step affects the rest of the flow.

1. **Gathering Data:**

Data Amassing is the first step of the Project. In this step, we have identified the different data sources, as data can be amassed from sundry sources such as files, database, internet. The quantity and quality of the accumulated data will determine the efficiency of the output. The more will be the data, the more precise will be the prediction.

2. **Data Pre-processing:**

Feature Cull is the process where you automatically or manually cull those features which contribute most to your presage variable or output in which you are intrigued with.

3. **Feature Extraction:**

Feature Extraction aims to minimize the number of features in a dataset by engendering incipient features from the subsisting ones.
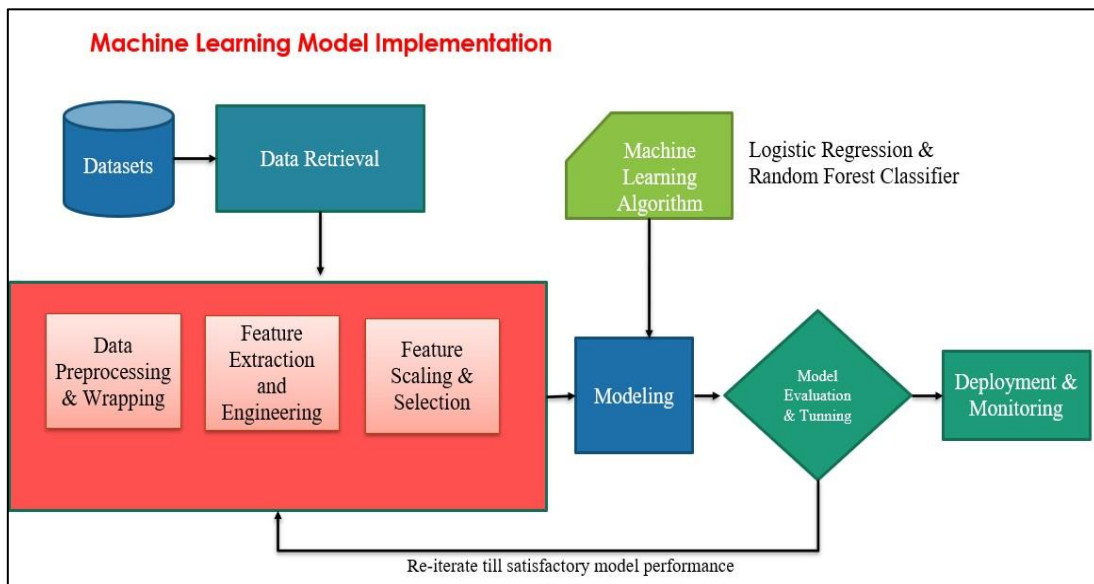

Fig 3.2 Machine Learning Model Implementation

4. **Feature Cull:**

Feature Cull is the process where you automatically or manually cull those features which contribute most to your prediction variable or output in which you are fascinated with.

5. **Train Model:**

We utilize datasets to train the model utilizing sundry machine learning algorithms. In this project we have use Arbitrary Forest Classifier and Logistic Regression Algorithm.

6. **Evaluation:**

Once our machine learning model has been trained on a given dataset, then we test the model. Testing the model determines the percentage precision of the model as per the requisite of project or quandary.

7. **Deployment:**

The last step of machine learning life cycle is deployment, where we deploy the model in the genuine- world system.

3.5 Analysis / Framework/ Algorithm

➢ **Framework**

The purport of framework is to sanction designers and developers to fixate on building a unique feature for their web-predicated projects rather than re-inventing by coding. Framework is specially engendered to avail you boost the performance and efficiency of your web app development task. In this project we have utilized some popular web frameworks like Bootstrap and Flask.

Bootstrap:

Bootstrap is the most popular CSS Framework for developing responsive and mobile-first websites. It is a free accumulation of implements for engendering a websites and web applications. It contains HTML and CSS-predicated design templates for typography, forms, buttons, navigation and other interface components, as well as optional JavaScript extensions.

Bootstrap is a massive amassment of reusable and multifarious pieces of code which are indited in CSS, HTML and JavaScript. The primary purport of integrating it to a web project is to apply Bootstrap's culls of color, size, font and layout to that project.

JavaScript:

JavaScript is a text-based programming language used both on the client-side and server-side that allows you to make web pages interactive. Where HTML and CSS are languages that give structure and style to web pages, JavaScript gives web pages interactive elements that engage a user.

JavaScript is a light-weight object-oriented programming language that is used by several websites for scripting the webpages. It is an interpreted, full-fledged programming language. JavaScript enables dynamic interactivity on websites when it is applied to an HTML document.

Flask:

Flask is a web application framework inscribed in Python. It is a Python module that lets you develop web applications facilely.

Flask is predicated on the Werkzeg WSGI toolkit and the Jinja2 template engine. It does have many cool features like URL routing, template engine.

Flask is considered more Pythonic than the Django web framework because in common situations the equivalent Flask web application is more explicit. Flask is also easy to get started with as a beginner because there is little boilerplate code for getting a simple app up and running.

Advantages:

- Higher compatibility with latest technologies

- Easier to utilize for simple cases

- Codebase size is relatively more diminutive

- High scalability for simple applications,

- Easy to build an expeditious archetype

- Routing URL is facile

- Easy to develop and maintain applications

- Minimal yet puissant platform

➢ **Algorithm**

Random Forest Classifier:

Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, customarily trained with the "bagging" method. The general conception of the bagging method is that a coalescence of learning models increases the overall result. Random Forest builds multiple decision trees and merges them together to get a more precise and stable prediction.
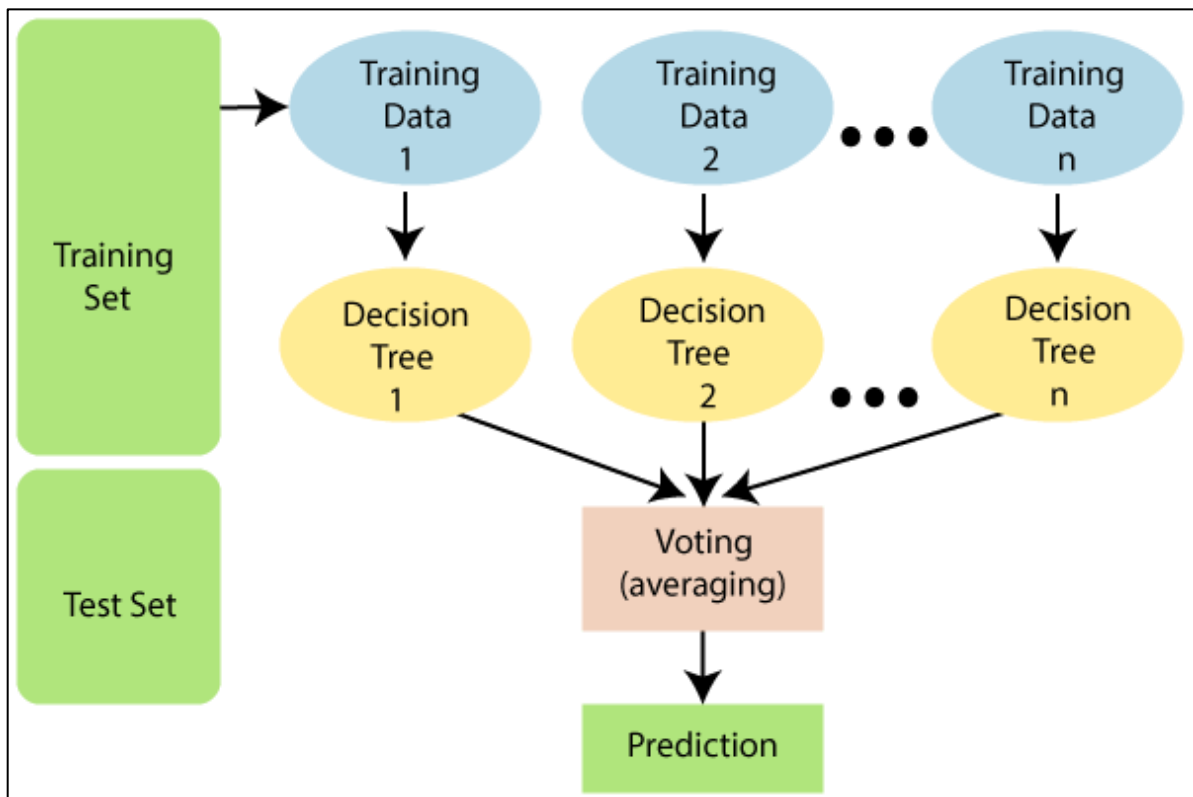

Fig 3.3 Random Forest Classifier Working

Step 1 − First, start with the selection of random samples from a given dataset.

Step 2 − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

Step 3 − In this step, voting will be performed for every predicted result.

Step 4 − At last, select the most voted prediction result as the final prediction result.

Advantages of Random Forest

- ➢ Random Forest is capable of performing both Classification and Regression tasks.

- ➢ It is capable of handling large datasets with high dimensionality.

- ➢ It enhances the accuracy of the model and prevents the overfitting issue.

Logistic Regression:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is utilized for soothsaying the categorical dependent variable utilizing a given set of independent variables.
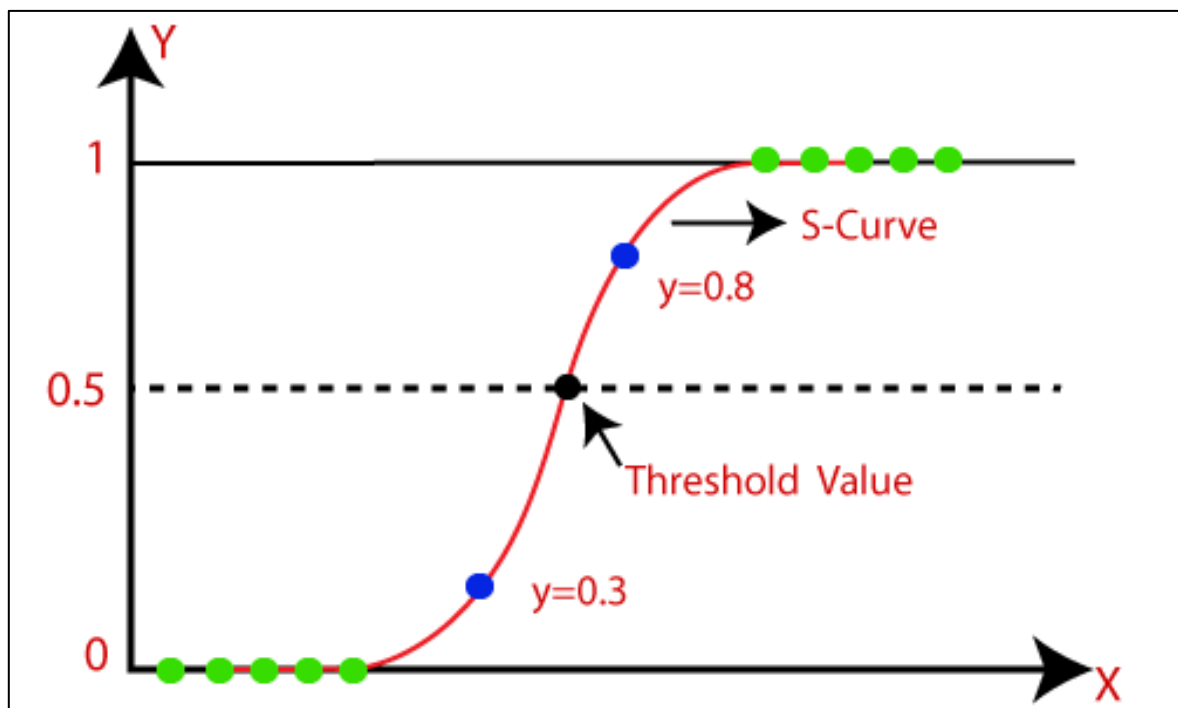

Fig 3.4 Logistic Regression Working

The logistic function that is a sigmoid function is an 'S' shaped curve that takes any real values and converts them between 0 to 1. If the output given by a sigmoid function is more than 0.5, the output is classified as 1 & if is less than 0.5, the output is classified as 0. If the graph goes to a negative end, then y predicted will be 0 and vice versa.

3.6 Details of Hardware & Software

Software Technology Used:

- ➢ **Core Technologies / Skills Used:** Machine Learning, Python Flask, Web Development
- ➢ **Front-End (Client-Side Scripting):** HTML, CSS, Bootstrap
- ➢ **Backend (Server-Side Scripting):** Python Flask, Javascript
- ➢ **Machine Learning Algorithm Used:** Random Forest Classifier, Logistic Regression
- ➢ **Cloud Used:** Python Anywhere
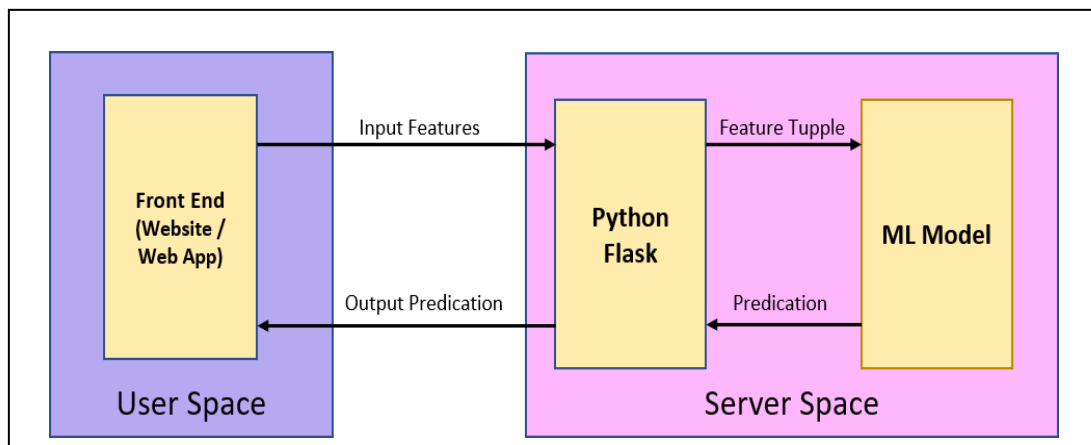
Software Working:

Fig 3.5 Working of Web Application

- ➢ Login page / Sign-Up (User Authorization)

   Website has a login/Sign Up page as a start Up page. New users need to sign up to access the main content of the web application. It consists of basic information such as name, email, contact no., date of birth, blood grp and user-defined password, these credentials will be stored in the database for future reference. If you are an existing user, you just need to Login using email and password provided during the Sign-Up process.

   Our software will perform user authentication here. After successful authentication, we will be redirected to the home page.

- ➢ Display all the different diseases on the home page.

  We have provided an attractive and easy-to-use user interface to our users so that anyone can use it easily. Every disease has separate web page.

  With regard to our top navigation bar, we have included a service button. After clicking on that button, services section will be displayed, which provides all our main services (disease diagnosis services).

- ➢ Each disease has separate page.

  In accordance with our requirements, we can select our services. If we click on one of the buttons, we'll be redirected to the page for that particular disease.

- ➢ Enter medical data carefully.

  The webpage after clicking on our preferred disease to predict consists of a form. This form has parameter of the medical data of the specific disease selected. Sliders are available through which user just needs to slide it back and forth accordingly the value will be forwarded to the machine learning model deployed. The output for the same will be displayed on the same webpage itself.

- ➢ Machine Learning Model will predict the output.

  Following that, the data we have entered in the form will be sent to the machine learning model that is running in the back end using Python Flask. Consequently, the machine learning model will predict its output based on the medical data we have entered.

- ➢ The results will be displayed on the same page itself.

  We are going to use Python Flask to display the final output that is determined by our machine learning model on the same page of our website.

  The system withal generates reports that can be subsidiary in-order to keep records. The report generated can be downloaded locally on user's device as well as provide on user's personal emails.

3.7 Design details

Sign-Up Page

Fig 3.6 Sign-Up Page

A signup page (also known as a registration page) enables users and organizations to independently register and gain access to the system.

Use:

In order to access the system, we need to create an account. After creating an account, we need to fill all our personal details: Name, Email, Date of Birth, Phone Number, Blood Group, and Password. Then our data is stored in the database.

Features:

- Indicates Password Strength. (Weak or Strong)

- Calendear to select date of birth.

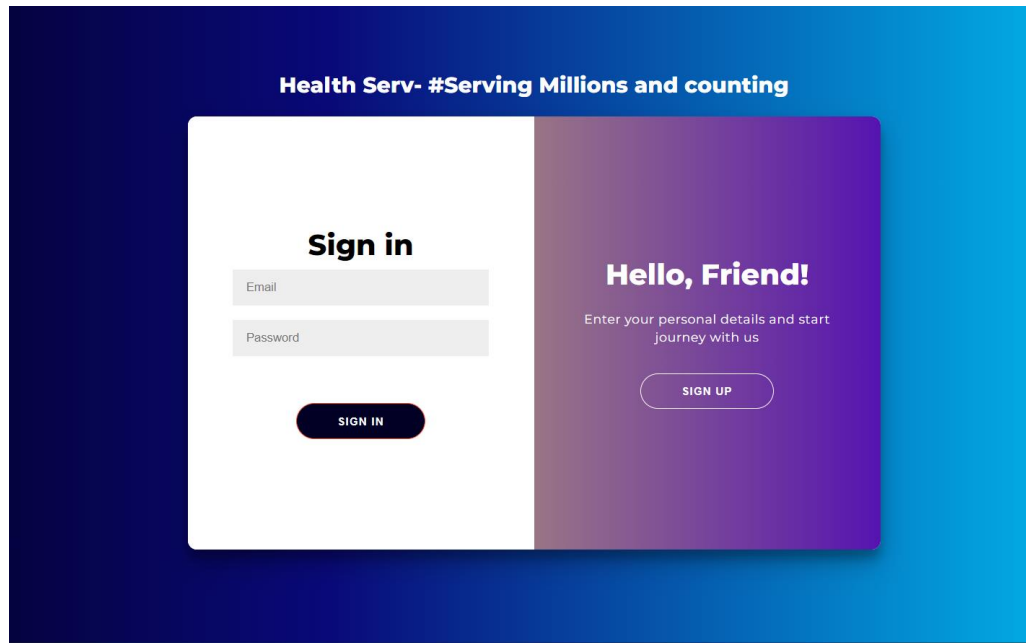- The Interface is easy to understand.

Login Page

Fig 3.7 Login Page

A login page is a web page that needs user identification and authentication by entering a username and password combination on a regular basis. Logins can provide access to the entire website. Logging in not only gives the user access to the site, but also enables the website to monitor user activities and behavior. Logging off a website or site may be a manual for the user or may occur automatically when such circumstances (such as page closure, device shutdown, long delay, etc.) occur.
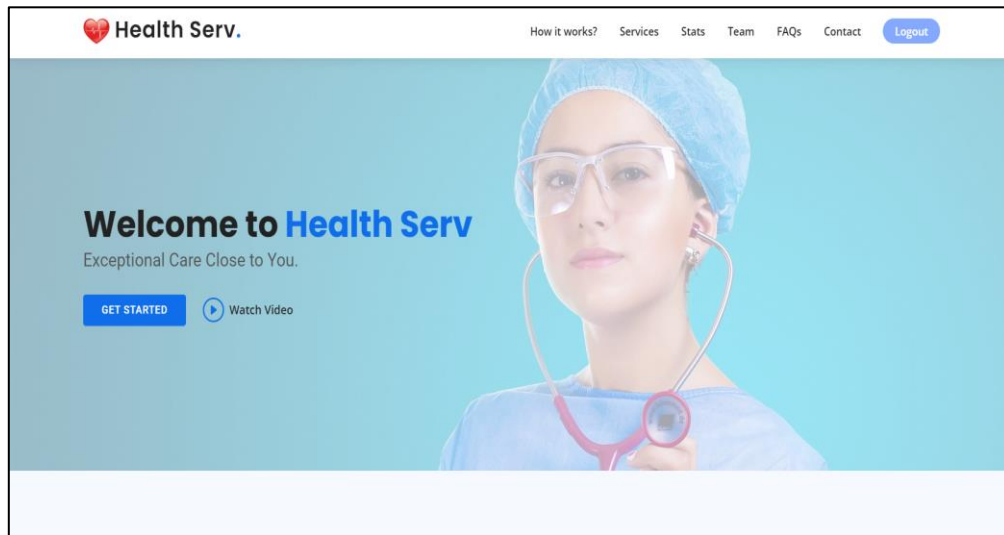
Use:

We can access the main website content by logging-in. Existing users just need to login using their email and password provided at the time of registration.

Features:

- Accurate user authentication process.

- The Interface is easy to understand.

Homepage

Fig 3.8 Home Page

Use: -

- Primary web page which a visitor navigating to a website from a search engine will visually perceive, and it may additionally accommodate as a landing page to magnetize visitors.

Navbar (for Navigation purport)

Navigation bar has links to each section present on the webpage providing better user experience. It also consists of a professional welcome to the user logged in. There is an introductory video based on comparison of everyday activities of a healthy person to that of an unhealthy person.

Use: -

- link to congruous sections/pages in a website and or avails readers in traversing the online document.

- a graphical utilizer interface intended to avail visitors in accessing information.

Buttons for selection of Diseases
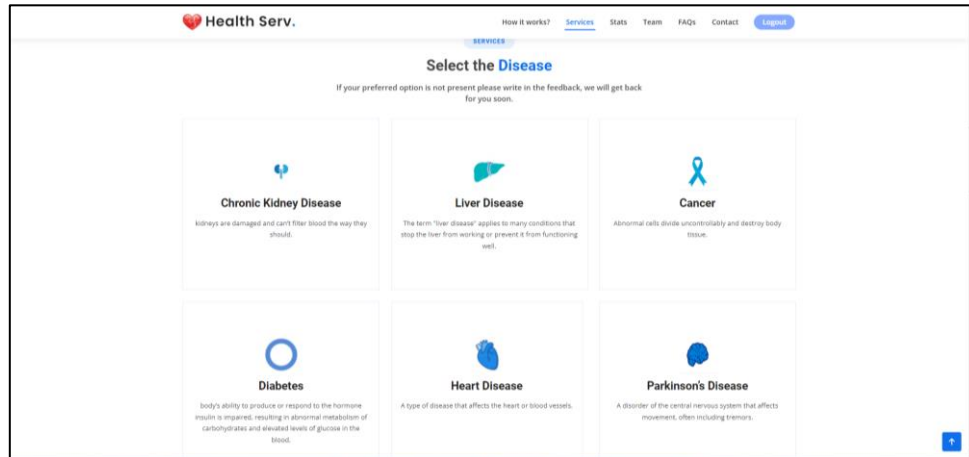
Fig 3.9 Buttons for selection of Diseases

- Button links are habituated to direct users to other pages (Particular Diseases Page where we can fill the required data of patient).

Disease Page

Fig 3.10 Disease Page

Disease's parameter:

Sr No.	Disease	Parameters	Algorithm Used	Accuracy in (%)
1	Diabetes	BP (mm Hg), Glucose, Insulin (mu U/ml), BMI (kg/m^2), Diabetes Pedigree Function, Age	Random Forest Classifier	82
2	Heart	Cholesterol, Fasting Blood Sugar, Chest Pain type	Logistic Regression	80
3	Liver	Proteins, Albumin, Bilirubin, Albumin and Globulin Ratio	Random Forest Classifier	79
4	Kidney	Sugar, Red Blood Cells, Blood Urea, Hyper Tension	Random Forest Classifier	99
5	Parkinson's	range of biomedical voice measurements	Random Forest Classifier	94
6	Breast Cancer	Radius, Perimeter, Area, Concavity, Concave Points	Logistic Regression	97

Table 3.1 Disease's parameter and deployed algorithms with accuracy

Backend / Database Part

We have use MYSQL database to store data of users.

- MySQL is a database system used on the web

- MySQL is a database system that runs on a server

- MySQL is ideal for both small and large applications

- MySQL is very fast, reliable, and easy to use

- MySQL uses standard SQL

- MySQL compiles on a number of platforms

- MySQL is free to download and use

- MySQL is developed, distributed, and supported by Oracle Corporation

- MYSQL Databases are useful for storing information categorically.

We have created two tables in the database.

Users:

The user table contains information about every user, such as their username, name, email, password, date of birth, mobile number, and blood group.

Feedback:

Feedback table contains information such as user id, name, email id, feedback subject, and feedback message about user feedback.
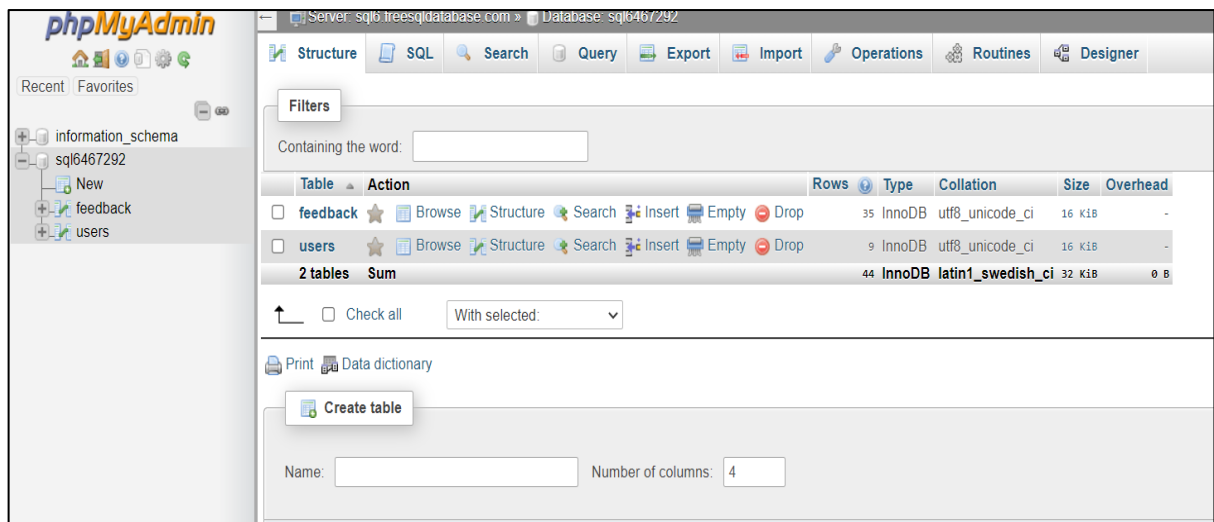
Fig 3.11 Backend Database

Fig 3.12 Users Table

Fig 3.13 Feedback Table

3.8 Budget Details

Sr. No.	Equipment Name	Quantity	Amount/ unit	Total Amount
1	Laptop / Desktop	1	>40,000	>40,000
2.	PyCharm IDE / Anaconda IDE and Jypyter Notebook	1	Free / Open Source	Free / Open Source
3	Cloud	1	Free / Open Source	Free / Open Source

Table 3.2 Budget Details

3.9 IMPLEMENTATION PLAN (TIME PLAN)

Activity	July	Aug.	Sept.	Oct.	Jan.	Feb.	Mar	Apr
Project Planning and Fulfillment of requirements	Start	End						
Learning Skills	Start	End						
Frontend Work		Start	Partially Completed	Partially Completed	Partially Completed	End		
Backend Work		Start	Partially Completed	Partially Completed	Partially Completed	End		
Testing						Start	End	
Modification						Start	End	
Deployment	`					Start	End	
Documentation				Start				End

Table 3.3 Implementation Plan

4. Results

4.1 Results

- This project will predict the diseases of the patient based on medical data utilizing datasets.

- Precise results after the prediction (above 70%)

- Less loading time of website.

- Felicitous working of report distribution system.

- Provide best user-friendly experience.

4.2 Analysis
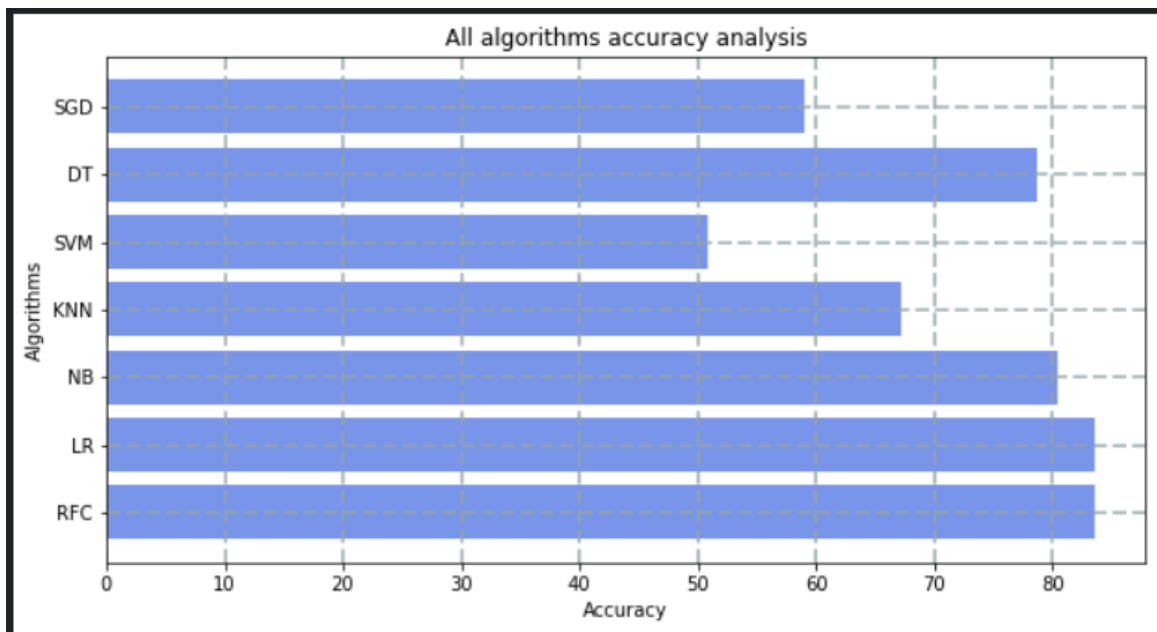
Fig 4.1 Algorithm Analysis

We used various machine learning algorithms such as random forest classifier, logistic regression, support vector machine, decision tree, K-nearest neighbor method, and naive Bayes classifier.

However, all analysis shows that random forest classifier and logistic regression provide better accuracy than any other machine learning algorithm. Therefore, we deployed only these two algorithms in our machine learning model.

5. Conclusion

5.1 Conclusion

The overall aim is to define various data mining techniques utilizable in efficacious disease prediction. Efficient and precise prediction with a lesser number of attributes and tests is our goal. We found the precision after implementing algorithms to be above 70%.

Another crucial goal we are looking forward to is to soothsay the disease which affects the patients salubrity a lot.

We all might have wondered utilizing online algorithm for these kinds of stuffs is not reliable, it provides you with even a slightest insight of your health that might get worse in future if ignored than its likely propitious for us.

5.2 Future Scope of work

- Addition of more diseases.

- Mobile-based / Android application in future work directions to expand the framework.

- Fixing Bugs to improve speed of website.

- Improvement in security of website.

- Facility for modifying user details.

- Provide more detailed report to the user.

6. REFERENCES

[1]   University of California, Irvine.: "Researchers create model to calculate COVID-19 health outcomes". California, 12-10, 2020.

[2]   Ferjani, Marouane. (2020). Disease Prediction Using Machine Learning. 10.13140/RG.2.2.18279.47521.

[3]   Sriram, T.V.S.; Rao, M.V.; Narayana, G.V.S.; Kaladhar, D.S.V.G.K. "A Comparison and Prediction Analysis for the Diagnosis of Parkinson Disease Using Data Mining Techniques on Voice Datasets". Int. J. Appl.Eng. Res. 2016,11, 6355–6360.

[4]   N. Lavesson, Evaluation and Analysis of Supervised Learning Algorithms and Classifiers, 2006

[5]   Kleinbaum, D.G.; Klein, "Analysis of matched data using logistic regression". In Logistic regression; Springer: Berlin, Germany, 2010.

[6]   Huang, M.J.; Chen, M.Y.; Lee, S.C. "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis" Expert Syst. Appl.

[7]   V. S and D. S, "Data Mining Classification Algorithms for Kidney Dis-ease Prediction," International Journal on Cybernetics & Informatics, vol. 4, no. 4, pp. 13–25, 2015

[8]   Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006.

[9]   Mehmed, K.: "Data mining: Concepts, Models, Methods and Algorithms", New Jersey: John Wiley, 2003.

[10]   Mohd, H., Mohamed, S. H. S.: "Acceptance Model of Electronic Medical Record", Journal of Advancing Information and Management Studies. 2(1), 75-92, 2005.

[11]   Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[12]   Sellappan, P., Chua, S.L.: "Model-based Healthcare Decision Support System", Proc. Of Int. Conf. on Information Technology in Asia CITA'05, 45-50, Kuching, Sarawak, Malaysia, 2005

7. Achievements

Sr. No.	Competition Name	Certificate
1	AVISHKAR Research Convention 2021-2022	Participation
2	INTECH 2K22	Participation
3	Oscillation 2K22 Technical Paper Presentation	Participation

Table 7.1 Achievements

- **Avishkar**

University of Mumbai
16th Inter-Collegiate/
Institute/Department

AVISHKAR
Research Convention: 2021-22
SELECTION ROUND
(District/Zonal Level Research Proposal Competition)

Certificate of Participation

This is to certify that **Prof. Pankaj Deshmukh** of **K. J. Somaiya Institute of Engineering and Information Technology, Sion** has guided a research proposal titled **Detection of Diseases using Machine Learning** which was submitted by his student **Mr. Ghadi Raj** in **Medicine and Pharmacy** category and **UG** level for the **Selection Round** of **16th Inter-Collegiate / Institute / Department Avishkar Research Convention: 2021-22** for zone.

Dr. (Mrs.) Minakshi Gurav
OSD,
Avishkar Research Convention,
University of Mumbai

Dr. Sunil Patil
Director,
Department of Students' Development,
University of Mumbai

Mumbai
April 25, 2022

University of Mumbai

University of Mumbai
16ᵗʰ Inter-Collegiate/
Institute/Department

AVISHKAR
Research Convention: 2021-22
SELECTION ROUND
(District/Zonal Level Research Proposal Competition)

Certificate of Participation

This is to certify that **Mr. Ghadi Raj** of **K. J. Somaiya Institute of Engineering and Information Technology, Sion** has participated and submitted a research proposal titled **Detection of Diseases using Machine Learning** in **Medicine and Pharmacy** category and **UG** level for the **Selection Round** of **16th Inter-Collegiate / Institute / Department Avishkar Research Convention: 2021-22** for zone.

Dr. (Mrs.) Minakshi Gurav
OSD,
Avishkar Research Convention,
University of Mumbai

Dr. Sunil Patil
Director,
Department of Students' Development,
University of Mumbai

Mumbai
April 25, 2022

University of Mumbai
16th Inter-Collegiate/
Institute/Department

AVISHKAR
Research Convention: 2021-22
SELECTION ROUND
(District/Zonal Level Research Proposal Competition)

Certificate of Participation

This is to certify that **Mr. Patil Shrejay** of **K. J. Somaiya Institute of Engineering and Information Technology, Sion** has participated and submitted a research proposal titled **Detection of Diseases using Machine Learning** in **Medicine and Pharmacy** category and **UG** level for the **Selection Round** of **16th Inter-Collegiate / Institute / Department Avishkar Research Convention: 2021-22** for zone.

Dr. (Mrs.) Minakshi Gurav
OSD,
Avishkar Research Convention,
University of Mumbai

Dr. Sunil Patil
Director,
Department of Students' Development,
University of Mumbai

Mumbai
April 25, 2022

University of Mumbai

University of Mumbai
16th Inter-Collegiate/
Institute/Department

AVISHKAR
Research Convention: 2021-22
SELECTION ROUND
(District/Zonal Level Research Proposal Competition)

Certificate of Participation

This is to certify that **Mr. Desai Rohit** of **K. J. Somaiya Institute of Engineering and Information Technology, Sion** has participated and submitted a research proposal titled **Detection of Diseases using Machine Learning** in **Medicine and Pharmacy** category and **UG** level for the **Selection Round** of **16th Inter-Collegiate / Institute / Department Avishkar Research Convention: 2021-22** for zone.

Dr. (Mrs.) Minakshi Gurav
OSD,
Avishkar Research Convention,
University of Mumbai

Dr. Sunil Patil
Director,
Department of Students' Development,
University of Mumbai

Mumbai
April 25, 2022

- **INTECH**

SOMAIYA VIDYAVIHAR

K. J. Somaiya Institute of Engineering and Information Technology
An Autonomous Institute Affiliated to the University of Mumbai
Accredited by NAAC and NBA, Approved by AICTE, New Delhi

The Institution of Engineering and Technology

Certificate of Participation

Shrejay Rajesh Patil

has actively participated in National Level Project Competition **"KJSIEIT - INTECH 2K22"**

organized by

K J Somaiya Institute of Engineering & Information Technology, Sion, Mumbai

in association with

IET Mumbai Local Network on April 16, 2022

Vikrant Sankhe

Mr. Vikrant Sankhe
Chairman, IET Mumbai LN

Dr. Sunita Patil
Vice Principal, KJSIEIT

Dr. Suresh Ukarande
Principal, KJSIEIT

The Institution of
Engineering and Technology

Certificate of Participation

Rohit Ramesh Desai

has actively participated in National Level Project Competition "KJSIEIT - INTECH 2K22"

organized by

K J Somaiya Institute of Engineering & Information Technology, Sion, Mumbai

in association with

IET Mumbai Local Network on April 16, 2022

Vikrant Sankhe

Mr. Vikrant Sankhe
Chairman, IET Mumbai LN

Dr. Sunita Patil
Vice Principal, KJSIEIT

Dr. Suresh Ukarande
Principal, KJSIEIT

8. Conference Paper

"Detection of Diseases using Machine Learning"

Raj Ghadi
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information
Technology
Mumbai, India
r.ghadi@somaiya.edu

Shrejay Patil
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information
Technology
Mumbai, India
shrejay.p@somaiya.edu

Rohit Desai
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information
Technology
Mumbai, India
rohit.desai@somaiya.edu

Prof. Pankaj Deshmukh
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information
Technology
Mumbai, India
pankaj@somaiya.edu

Abstract—*The world's growing population has put enormous pressure on the healthcare sector to offer high- quality treatment and accommodations. Artificial intelligence and Machine Learning are no exceptions in the healthcare industry, which has long been a vigorous adherent of cutting-edge technology. We have developed a web application using flask framework. It consists of web pages designed for different functionality. It is a disease prediction system which can be deployed on any network for communication among ecumenical users. Report is generated that can be subsidiary in-order to keep records. The report generated can be downloaded locally on user's device as well as provided on user's personal emails.*

Keywords—prediction, MySQL, python, session, admin, machine learning,

I. INTRODUCTION

World is suffering because of Many Diseases and to surmount it we can utilize artificial intelligence and machine learning. The purport of artificial intelligence and machine learning is to make the machine more prosperous, efficient, precise and reliable than afore. However, in a healthcare system, it will definitely avail medicos a lot in critical situations and decisions. To minimize the pressure of the healthcare system and to avail the medicos and society we have engendered the project which will predict the particular disease, detect the disease in earlier stage and predict the diseases very accurately and efficiently. This will avail medicos to attest or cross- check their postulation and analysis. It will avail them in critical situations and decisions. The interface has a navigation-bar-driven programme that enables facile utilizer interaction with some GUI applications. Login and Signup forms are a component of user authentication. All details acquired during signup process are stored in the database which can only be accessed by Admin. Admin here is the one who manages the website and works in the backend maintaining all the data extracted during the user's session. Sessions are engendered that avails maintain users state and data all over the application. Different sections such as contact, FAQ, feedback and analysis are present on the webpage. Especially the analysis section provides transparency to the users about how our model works in the backend as healthcare is a consequential issue and just can't be ignored for.

II. LITERATURE SURVEY

A. UCI researchers create model to calculate COVID-19 health outcomes

University of California, Irvine health sciences researchers have engendered a machine-learning model to predict the probability that a COVID-19 patient will require a ventilator or ICU care. The implement is free and available online for any healthcare organization to utilize.

"The goal is to give an earlier alert to clinicians to identify patients who may be vulnerably susceptible at the onset," verbally expressed Daniel S. Chow, an assistant pedagogia in residence in radiological sciences and first author of the study, published in PLOS ONE. The implement predictions whether a patient's condition will worsen within 72 hours.

Coupled with decision-making concrete to the healthcare setting in which the implement is utilized, the model utilizes a patient's medical history to determine who can be sent home and who will require critical care. The study found that at UCI Health, the implement's predictions were precise about 95 percent of the time.

B. Disease Prediction Using Machine Learning

Computerized systems are currently considered to be much more efficient than the traditional ones, similarly adapting these systems in the healthcare sector would yield better results comparatively. The concept of supervised machine learning algorithms holds enormous potential for disease diagnosis. Huge amount of data is required in such systems in order to gain high precision output. There are many types of algorithms available, selection of these algorithms is very crucial at the time of designing the machine learning model. In this literature, the aim is to apperceive trends across various types of supervised ML models in disease detection through the examination of performance metrics.

There are some algorithms such as Naves Bayes (NB), Decision Trees (DT), And K-Nearest Neighbor (KNN) etc. is considered to be most prominent among others. According to the research Support Vector Machine (SVM) was found to be most eligible at detecting Kidney and Parkinson's

diseases. Similarly Logistic Regression (LR) for heart disease, Random Forest Classifier (RFC) and Convolutional Neural Networks (CNN) for breast and common diseases were selected respectively.

III. PROPOSED SYSTEM

Proposed Diseases prediction System to Improve quality and efficiency of current healthcare system. "It is introduced with the aid of web creation. In order to prevent pitfalls in the current system, the proposed system is built as a web-based system where it can be accessed at anytime and anywhere on their mobile/PC. The user can use their credential to access the server system. The website faculties will now keep track of user's health by providing insights of health issues if any. Additional features include generation of report.

Figure 1 shows the flowchart of proposed web-based system. A login page is a web page that needs user identification and authentication by entering a username and password combination on a regular basis. Logins can provide access to the entire website. Logging in not only gives the user access to the site, but also enables the website to monitor user activities and behaviour. Logging off a website or site may be a manual for the user or may occur automatically when such circumstances (such as page closure, device shutdown, long delay, etc.) occur.

A. Diseases Parameters

Every disease has different parameters on the basis of which it is detected. Below are few most significant parameters tabulated in table 1 after data pre-processing.

Table I. Most prominent parameters for disease detection

Disease	Parameters	Algorithm Used	Accuracy in (%)
Diabetes	BP (mm Hg), Glucose, Insulin (mu U/ml), BMI (kg/m2), Diabetes Pedigree Function, Age	Random Forest Classifier	82
Heart	Cholesterol, Fasting Blood Sugar, Chest Pain type	Logistic Regression	80
Liver	Proteins, Albumin, Bilirubin, Albumin and Globulin Ratio	Random Forest Classifier	79
Kidney	Sugar, Red Blood Cells, Blood Urea, Hyper Tension	Random Forest Classifier	99
Parkinson's	range of biomedical voice measurements (Hz)	Random Forest Classifier	94
Breast Cancer	Radius, Perimeter, Area, Concavity, Concave Points	Logistic Regression	97

B. Units

- The number of cycles per second is called frequency. The SI unit for frequency is the hertz (Hz).

- A millimetre of mercury is a manometric unit of pressure, described as the extra pressure generated by a column of mercury one millimetres high, defined as exactly 133.322387415 pascals. It is denoted mmHg.

- Bbody mass index defined by the square of the body height, and is expressed in units of kg/m2, formulated from mass in kilograms and height in meters.

- A litre is a measure of volume that is a bit greater than a quart.

- A millimeter abbreviated as mm is a small unit of displacement in the metric system.

- All the parameters utilized in Parkinson's diseases are the measure of voice in Hertz and db.(decibel) of healthy person and some of unhealthy person.

C. Developing Tools

The developing tools are utilized for web page plan and database building. To begin with, the internet pages of Diseases Prediction Framework were outlined by the HTML5, CSS3, J Query and JavaScript, because the software is simple to induce and simple to use. We have utilized PYTHON for the programming, since the internet page designed by it is more proficient in preparing the complex working environment providing vast range of libraries. Flask is a web application framework written in Python.

The application can be deployed on desktops with any operating systems. About the database apparatuses, the MySQL server is used.

Several algorithms were implemented, Figure 10 shows comparison between them on basis of accuracy. Hence considering accuracy two most prominent algorithms were shortlisted and deployed mentioned below:

Random Forest Classifier:
It's a type of supervised learning algorithm that focuses on ensemble technique. Multiple decision trees make a forest and are trained with the "bagging" method. Combining various learning models increases the accuracy relatively. Decisions are provided individually by the trees and the most voted one becomes the prediction of our models.

Logistic Regression:
Another most popular supervised learning algorithm is the Logistic Regression. It utilizes the famous sigmoid function providing a S-shaped graph. It basically works similar to that of mathematical probability resulting in classification of categorical data.

IV. IMPLEMENTATION OF PROPOSED PROJECT

Fig. 2. Authentication Login form

Figure 2 shows the diseases prediction system login page and it is password-protected. A unique email and password have to be entered. If the correct email and password are entered, it will lead to a Home Page of the website.

Fig. 3. Sign Up Form

Figure 3 shows the diseases prediction system Sign Up page in which the user needs to fill its credentials which is stored in a database for further processes. It contains fields such as name, email, contact, blood group, date of birth and password.

Fig. 4. Home page

Top navigation bar consists of links to each section present on the webpage providing better user experience as shown in figure 4. It also consists of a professional welcome to the user logged in.

Sessions are created for every user in order to maintain data of a specific user throughout the web application.

Fig. 6. Diseases section

In Figure 6 these are the various diseases that are available for the users for check-up using their medical data. It consists of kidney disease, liver disease, breast cancer, diabetes, heart disease and Parkinson's disease.
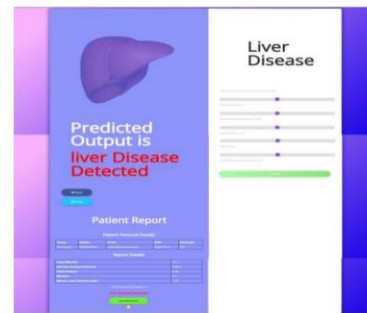

Fig. 9. Disease prediction page

Figure 9 shows a disease prediction page. Each disease has their separate prediction page. As soon as the user clicks on the disease that he/she prefer to check will land on this page where the user need to enter the medical data and click on predict button. The output generated will be displayed on the same page itself.
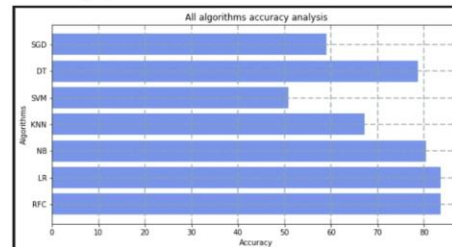

Fig. 10. Algorithm comparison

Algorithms and their analysis are provided to the user as shown in figure 10.

V. Conclusion

The overall aim is to define various data mining techniques utilizable in efficacious disease prediction. Efficient and precise prediction with a lesser number of attributes and tests is our primary goal. In this study. The data were pre-processed and then utilized in the model. We found the precision after implementing algorithms to be above 70 percent.

Another crucial goal we are looking forward to is to soothsay the disease in its early stage which affects the patients salubrity a lot. As we all might be aware of the fact that early stages of any diseases are too tough to detect and if detected will open gates for lots of treatment processes.

We all might have wondered utilizing online algorithm for these sensitive issues is not reliable, we can verbalize that it's partly veritable as it is a machine and we cannot plenarily rely on its prediction, but still if it provides you with even a slightest insight of your health that might get worse in future if ignored than its likely propitious for us.

Reference

[1] Daniel S. Chow, Justin Glavis-Bloom, Jennifer E. Soun, Brent Weinberg, Alpesh N. Amin, Peter D. Chang. Development and external validation of a prognostic implement for COVID-19 critical disease. PLOS ONE, 2020 ; 15 (12): e0242953 DOI: 10.1371/journal.pone.0242953.

[2] Ferjani, Marouane. (2020). Disease Prediction Using Machine Learning. 10.13140/RG.2.2.18279.47521.

[3] M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning," 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391453.

[4] A. C. Lyngdoh, N. A. Choudhury and S. Moulik, "Diabetes Disease Prediction Using Machine Learning Algorithms," 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2021, pp. 517-521, doi: 10.1109/IECBES48179.2021.9398759.

[5] A. Grover, A. Kalani and S. K. Dubey, "Analytical Approach towards Prediction of Diseases Using Machine Learning Algorithms," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 793-797, doi: 10.1109/Confluence47617.2020.9058120.

[6] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic Diseases Diagnosis using Machine Learning," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018, pp. 1-6, doi: 10.1109/ICCSDET.2018.8821235.

[7] Goel, Rati, Heart Disease Prediction Using Various Algorithms of Machine Learning (July 12, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, Available at SSRN: https://ssrn.com/abstract=3884968 or http://dx.doi.org/10.2139/ssrn.3884968.