FACULTY OF COMPUTER SCIENCE & ENGINEERING

DEPT. OF COMPUTER SCIENCE & ENGINEERING

BACHELOR OF TECHNOLOGY (Computer Engineering)

(ACADEMIC SESSION: 2022-2023)

MAJOR PROJECT REPORT

ON

"MeDiagnose- Identification of Disease and Suggesting References Based on Symptoms Developed using Machine Learning and Integrating it with Web App"

Niharika Jain

2019PUSCEBCEX07365 IV Year

Prakhar Saxena

2019PUSCEBCCX07276 IV Year

Prikshit Saini

2019PUSCEBCCX07513 IV Year

Dr. A. Anushaya
(Project Guide)

Mr. Shikha Sharma
(Head of Department)

Poornima University

Plot No. 2027-2031, Ramchandrapura P.O. VidhaniVatika Road, Sitapura, Jaipur, Rajasthan- 303905 (India)

www.poornima.edu.in

A

MAJOR PROJECT REPORT

On

"MeDiagnose- Identification of Disease and Suggesting References Based on Symptoms Developed using Machine Learning and Integrating it with Web App"

Submitted in partial fulfilment of the requirementsfor the award of the Degree of

BACHELOR OF TECHNOLOGY (Computer Engineering)

Poornima University, Jaipur(Academic Session: 2022-23)

Submitted By:

Niharika Jain 2019PUSCEBCEX07365

Prakhar Saxena 2019PUSCEBCCX07276

Prikshit Saini 2019PUSCEBCCX07513

IV Year, Computer Science & Engineering

Submitted To:

Department of Computer Science & Engineering

Faculty of Computer Science & Engineering

Poornima University

Ramchandrapura, Sitapura Ext., Jaipur, Rajasthan-(303905)

CANDIDATE'S DECLARATION

I hereby declare that the work presented in the Project report entitled **"MeDiagnose-Identification of Disease and Suggesting References Based on Symptoms Developed using Machine Learning and Integrating it with Web App"** is submitted by **Niharika Jain 2019PUSCEBCEX07365, Prakhar Saxena 2019PUSCEBCCX07276** and **Prikshit Saini 2019PUSCEBCCX07513** is in the fulfillment of the requirements for the award of Bachelor of Technology in Computer Engineering from Poornima University, Jaipur during the academic year 2022-23. The work has been found satisfactory, authentic of my own work carried out during me degree and approved for submission.

The work reported in this has not been submitted by me for award of any other degree or diploma.

Date: 28/04/2023

Niharika Jain
2019PUSCEBCEX07365
Prakhar Saxena
2019PUSCEBCCX07276
Prikshit Saini
2019PUSCEBCCX07513

CERTIFICATE

This is to certify that project work entitled **"MeDiagnose- Identification of Disease and Suggesting References Based on Symptoms Developed using Machine Learning and Integrating it with Web App"** is a bonafide work carried out in the eight semester by **Niharika Jain 2019PUSCEBCEX07365, Prakhar Saxena 2019PUSCEBCCX07276 and Prikshit Saini 2019PUSCEBCCX07513** in partial fulfillment for the award of Bachelor of Technology in Computer Science & Engineering from Poornima University Jaipur, during the academic year 2022 -2023.

Dr. A. Anushya
Associat Professor,
Faculty of Computer Science and Engineering
Poornima University, Jaipur

Mr. Shikha Sharma
HoD
Department of Computer Science & Engineering,
Faculty of Computer Science and Engineering,
Poornima University, Jaipur

Dr Ajay Khunteta
Dean,
Faculty of Computer Science and Engineering,
Poornima University, Jaipur

ACKNOWLEDGEMENT

I have undergone a Major project which was meticulously planned and guided at every stage so that it became a life time experience for me. This could not be realized without the help from numerous sources and people in the Poornima University and Programming Express.

I am thankful to **Dr. Manoj Gupta, Pro President, and Poornima University for** providing usa platform to carry out this activity successfully.

I am also very grateful to **Dr. Ajay Khunteta,** Dean, Faculty of Computer Science & Engineering, and **Ms. Shikha Sharma**, HoD, Department of Computer Science & Engineering, for their kind support and guidance. I would like to take this opportunity to show our gratitude towards **Dr. A. Anushya** who helped me in successful completion of my project. She has been a guide, motivator & source of inspiration for us to carry out the necessary proceedings for completing this training and related activities successfully and grateful for her guidance and support.

I am thankful for their kind support and providing us expertise of the domain to develop the project. I would also like to express our heartfelt appreciation to all of our friends whose direct or indirect suggestions help us to develop this project and to entire team members for their valuable suggestions.

Lastly, thanks to all faculty members of Department of Computer Science & Engineering for their moral support and guidance.

Niharika Jain
2019PUSCEBCEX07365
Prakhar Saxena
2019PUSCEBCCX07276
Prikshit Saini
2019PUSCEBCCX07513

ABSTRACT

People often concern the relationships between symptoms and diseases when seeking medical advices. In this project, medical data are divided into three copies, records related to main disease categories, records related to subclass disease types, and records of specific diseases firstly; then two disease recognition methods only based on symptoms for the main disease category identification, subclass disease type identification, and specific disease identification are given. In the methods, a neural network and a support vector machine (SVM) algorithms are adopted, respectively. In the method validation part, accuracy of the two diagnosis methods is tested and compared. Results show that automatic disease prediction only based on symptoms is possible for intelligent medical triage and common disease diagnosis. Accurate and on-time analysis of any health-related problem is important for the prevention and treatment of the illness. The traditional way of diagnosis may not be sufficient in the case of a serious ailment. Developing a medical diagnosis system based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis than the conventional method. We have designed a disease prediction system using multiple ML algorithms. The data set used had more than 230 diseases for processing. Based on the symptoms, age, and gender of an individual, the diagnosis system gives the output as the disease that the individual might be suffering from. After the disease Identification Health tips

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF GRAPHS

LIST OF TABLES

.

Chapter 1

Introduction to MeDiagnose

At present, there is shortage of per capita medical resources, and high-quality medical resources are concentrated in large cities and large hospitals. In China, many patients have strong health awareness, even if their symptoms are not serious; they also flock to large hospitals to seek quality medical services. Constraints and conflicts between medical resource supply and demand are a long-standing phenomenon.

In medical consultations, what people intuitively care about are the relationships between symptoms and diseases. Nowadays, many people provide symptoms online to obtain prediagnosis results, and their objective is to screen critical illnesses and seek an advice for further accurate medical treatment.

An intelligent information system, which can automatically perform prediagnosis based on the symptoms provided by patients, can alleviate the problem of medical resource shortage. In this project, such diagnosis methods are proposed. Through these methods, preliminary diagnoses can be provided for specialized diseases, and it can help medical workers in areas having underdeveloped medical resources implement medical triage and provide consultation services for people who will seek precise treatment in big hospitals. Additionally, a common disease diagnosis service can be realized for people who can seek medical treatment by themselves.

1.1 Introduction

The intelligent diagnosis problem to be solved in this project includes two aspects. The first one is seeking diagnosis experience according to the relationships between symptoms and diseases. Here, supervised machine learning methods are adopted. The second one is disease prediction based on the symptoms provided by visitors. The first one is the main problem. People often concern the relationships between symptoms and diseases when seeking medical advices. In this project, medical data are divided into three copies, records related to main disease categories, records related to subclass disease types, and records of specific diseases firstly; then two disease recognition methods only based on symptoms for the main disease category identification, subclass disease type identification, and specific disease identification are given. In the method validation part, accuracy of the two diagnosis methods is tested and compared.

Results show that automatic disease prediction only based on symptoms is possible for intelligent medical triage and common disease diagnosis. The Features of Me diagnose includes Health tips , Providing Awareness about disease and latest News updates from Medical Field.

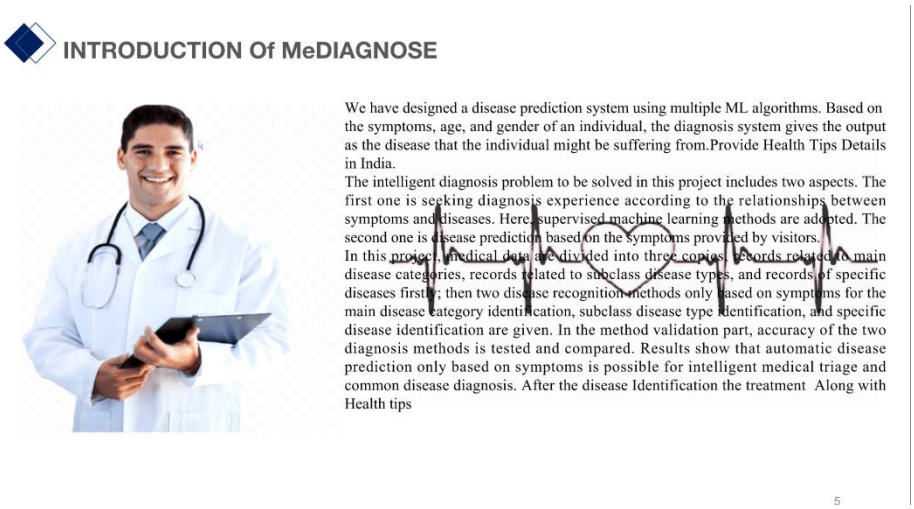


Figure 1 *Introduction to Me Diagnose*

Computer aided diagnosis research has begun since the last century. Most intelligent disease diagnosis researches focus on a certain type disease or only a specific disease. The contents mostly are intelligent diagnosis using machine learning algorithms based on pathology data, influencing factors, examination data, physiological performance, or images when disease types are known previously [1–7]. Some exploratory works have discussed the disease diagnosis only based on the symptoms provided by patients. A simple method is to compare the symptoms provided by a patient to record symptoms in each data item, and the disease in the most similar entry is an output result. In [8], the user gives out feathers related to the diseases such as gender, age, affected part, and related symptoms firstly. Jackcard similarities are calculated based on symptom matrixes, and the similarities are arranged in descending order. Diseases in the first 3 items are selected as alternative recommended answers. In [9], the similarities, which are evaluated by differences between a symptom vector provided by the user and characteristic symptom sets of different diseases, are calculated. The similarities are also arranged in descending order, and the diseases in the first 3 selected items are alternative recommended answers. Disease diagnosis only based on symptoms and without disease type limitation is a general practice (GP) problem. If the above methods are used to solve this kind of diagnosis problem, the efficiency is extremely low, and repetition calculations are involved in each diagnosis case. In related works [10, 11], automatic disease diagnoses based on machine learning algorithms are proposed; in these works, symptoms are extracted firstly, and then, the

diagnosis is implemented using deep learning algorithms. There are many diseases, while all proposed methods are limited to discussions on few diseases in the above projects.

Without detailed medical examination data and pathology support, accuracy of diagnosis methods based on symptoms cannot be guaranteed, while, in current online applications, reports, and documents, diagnosis only based on symptoms can be a disease screening method and used to help fast disease type recognition and disease triage in hospital. The key problem is the adaptability of this kind of diagnosis methods. At present, there is no discussion about which disease type levels or which diseases this kind of diagnosis methods is suitable for. To fill this gap, in this project, this issue is considered.

Disease prediagnosis based on symptoms, which are contained in consultation words, is indeed a text classification problem. In these works, the first step would mostly be lexical feather extraction, and then classification based on different feather properties is implemented [12–14]. Considering the particularity in clinic and immature Chinese word segmentations, in this project, we only discuss the core prediagnosis problem, and the symptoms, which are also disease feathers, have been extracted according to clinical experience previously. A hierarchical frame is provided in this project. Firstly, the diseases are divided into major categories and then are divided into several subtypes. Furthermore, specific diseases are filled into subclass disease types. In this project, two automatic diagnosis methods using a neural network technology and a support vector machine (SVM) technology, respectively, are given to solve this general practice (GP) problem. In the methods, the first is the major disease category identification, and then it is based on the results to identify disease subtypes. Further process is the training for specific disease identification. To observe the effectiveness, the two diagnosis methods are tested and compared.

1.2 Objective Of Me Diagnose

Disease Prediction using Machine Learning is a system that predicts the disease based on the information provided by the user. It also predicts the disease of the patient or the user based on the information or the symptoms he/she enters into the system and provides accurate results based on that information. If the patient is not very serious and the user just wants to know the type of disease, he/she has been through. It is a system that provides the user the tips and tricks to maintain the health system of the user and it provides a way to find out the disease using this prediction. Now a day’s health industry plays a serious role in curing the diseases of the patients

so this is often also some quite help for the health industry to tell the user and also it's useful for the user just in case he/she doesn't want to travel to the hospital or the other clinics, so just by entering the symptoms and every one other useful information the user can get to understand the disease he/she is affected by and therefore the health industry also can get enjoy this technique by just asking the symptoms from the user and entering in the system and in just a few seconds they can tell the exact and up to some extent the accurate diseases.

The Features of Me diagnose includes Health tips , Providing Awareness about disease and latest News updates from Medical Field.

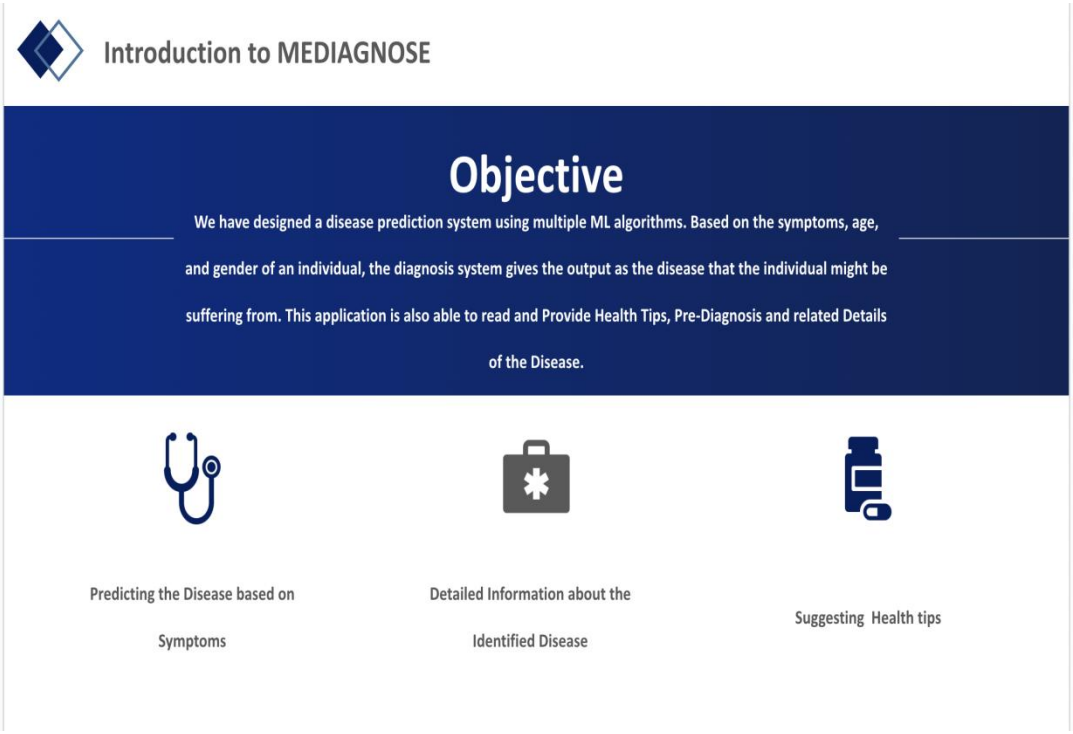


Figure 2 *Objective of Me Diagnose*

The background of this study was that the Internet medical service has a large amount of structured information such as diseases and hospitals, but the service to users is very limited. Most service sites only offer features such as online searches and consultation with doctors. Some websites offer simple self-diagnosis services. However, most only allow users to enter symptom keywords. The list of diseases given by the system is very long. Therefore, this project proposes a method for extracting disease symptom keywords by a related machine learning algorithm. The application of this method can effectively retrieve information for intelligent recommendation of doctors and hospitals and greatly improve the efficiency of patients' medical treatment. It is of great significance and can effectively improve the medical environment of the

hospital, but the disadvantage is that not every patient can use the system proficiently, and it needs a lot of operation training for the system later to be familiar with the use of the system.

The method adopted in this project is that the system integrates machine learning algorithms and knowledge graph technology to help patients conduct online consultations. The power of this method is to effectively improve the level of urban medical care and services. The system analyses the misclassified data of different departments through high-frequency word analysis.

Chapter 2

Technologies Used

Machine learning is the process of teaching computers to learn from data, without being explicitly programmed. It enables computers to make decisions and predictions by recognizing patterns in data. The applications for machine learning are endless – it can be used for anything from detecting credit card fraud, to automatically tagging friends in photos on Facebook, to predicting sales trends.

The field of machine learning is constantly evolving, with new concepts and technologies being developed all the time. In order to stay ahead of the curve, it's important for data scientists to Stay up to date with the latest advancements by following some of these blogs. This will help you to understand how machine learning can be applied in practice and give you ideas for potential applications in your own business or field of work.

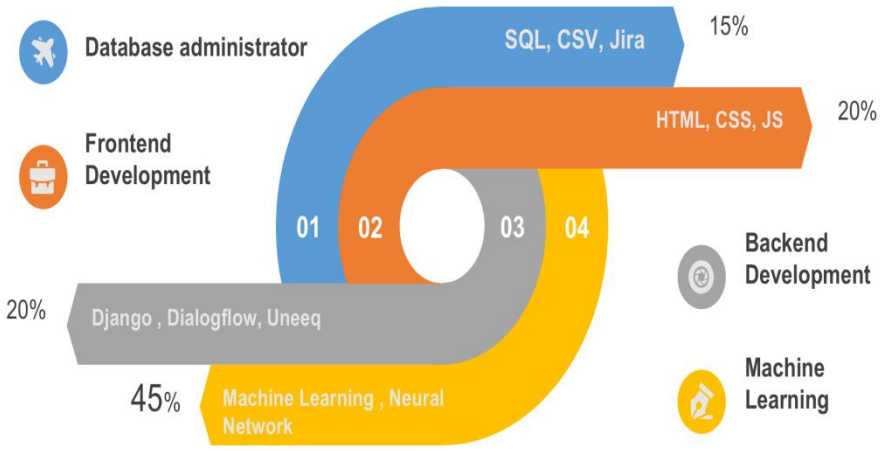


Figure 3 *Technologies of Me Diagnose*

2.1 FRONTEND TECHNOLOGIES

Front-end web development is the development of the graphical user interface of a website, through the use of HTML, CSS, and JavaScript, so that users can view and interact with that website. The website's front end is everything you see and can interact with using a browser. So,

creating this visual part is called front-end development. You could even say that designers creating user interfaces and planning experiences are also front-end developers, as they are working in collaboration on the same part of theproject.To create the front end, engineers use the combination of HTML (for basic page structure and content), CSS (for visual editing), and JavaScript (for making websites interactive). The same set of tools is used to create progressive web apps – mobile apps that look and feel like a native one but are created with the use of front-end technologies.

2.1.1 HTML

HTML stands for Hypertext Markup Language. It is used to design web pages using a markup language. HTML is the combination of Hypertext and Markup language. Hypertext defines the link between web pages. A markup language is used to definethe text document within the tag which defines the structure of web pages. This language is used to annotate (make notes for the computer) text so that a machine can understand it and manipulate text accordingly. Most markup languages (e.g., HTML) are human-readable. The language uses tags to define what manipulation has to be done on the text.

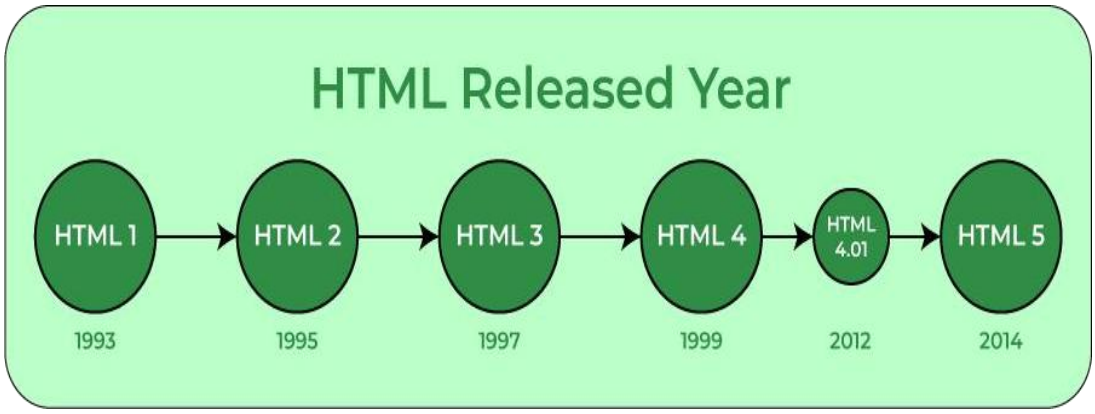


Figure 4 *HTML version histroy*

HTML is a markup language used by the browser to manipulate text, images, and other content, in order to display it in the required format. HTML was created by Tim Berners-Lee in 1991. The first-ever version of HTML was HTML 1.0, but the first standard version was HTML 2.0, published in 1995.

2.1.1.1 Elements and Tags: HTML uses predefined tags and elements which tell the browser how to properly display the content. Remember to include closing tags. If omitted, the browser applies the effect of the opening tag until the end of the page.

2.1.1.2 HTML page structure: The basic structure of an HTML page is laid out below. It contains the essential building-block elements (i.e., doctype declaration, HTML, head, title, and body elements) upon which all web pages are created.

2.1.1.3 <!DOCTYPE html>:

This is the document type declaration (not technically a tag). It declares a document as being an HTML document. The doctype declaration is not case-sensitive.

<html>: This is called the HTML root element. All other elements are contained within it.

<head>: The head tag contains the "behind the scenes" elements for a webpage. Elements within the head aren't visible on the front-end of a webpage. HTML elements used inside the <head> element include:

<style>-This html tag allows us to insert styling into our webpages and make them appealing to look at with the help of CSS.

<title>-The title is what is displayed on the top of your browser when you visit a website and contains title of the webpage that you are viewing.

<base>-It specifies the base URL for all relative URLs in a document.
<noscript>– Defines a section of HTML that is inserted when the scripting has been turned off in the user's browser.

<script>-This tag is used to add functionality in the website with the help of JavaScript.

<meta>-This tag encloses the meta data of the website that must be loaded every time the website is visited. For eg: - the metadata charset allows you to use the standard UTF-8 encoding in your website. This in turn allows the users to view your webpagein the language of their choice. It is a self-closing tag.

<body>: The body tag is used to enclose all the visible content of a webpage. In otherwords, the body content is what the browser will show on the front-end.

2.1.1.4 Features of HTML:

- It is easy to learn and easy to use.
- It is platform-independent.
- Images, videos, and audio can be added to a web page.

- Hypertext can be added to the text.
- It is a markup language.

2.1.1.5 Why learn HTML?

- It is a simple markup language. Its implementation is easy.
- It is used to create a website.
- Helps in developing fundamentals about web programming.
- Boost professional career.

2.1.1.6 Advantages:

- HTML is used to build websites.
- It is supported by all browsers.
- It can be integrated with other languages like CSS, JavaScript, etc.

2.1.1.7 Disadvantages:

- HTML can only create static web pages. For dynamic web pages, other languages have to be used.
- A large amount of code has to be written to create a simple web page.
- The security feature is not good.

2.1.2 CSS

- CSS saves time: You can write CSS once and reuse the same sheet in multiple HTML pages.
- Easy Maintenance: To make a global change simply change the style, and all elements in all the web pages will be updated automatically.
- Search Engines: CSS is considered a clean coding technique, which means search engines won't have to struggle to "read" its content.
- Superior styles to HTML: CSS have a much wider array of attributes than HTML, so you can give a far better look to your HTML page in comparison to HTML attributes.
- Offline Browsing: CSS can store web applications locally with the help of an offline cache.

Using this we can view offline websites.

- CSS versions release years:

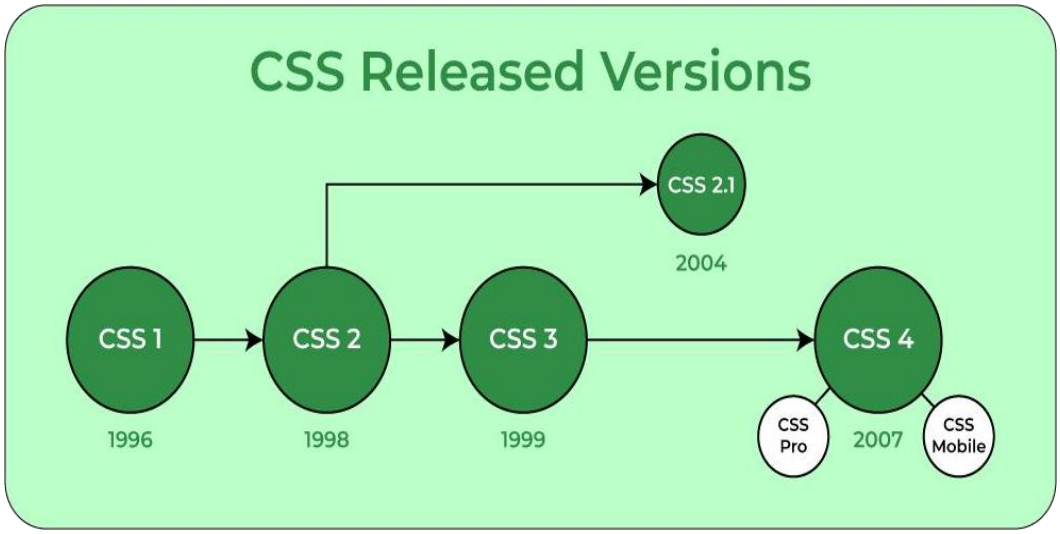


Figure 5 *CSS version topology*

- CSS comprises style rules that are interpreted by the browser and then applied to the corresponding elements in your document. A style rule set consists of a selector and declaration block.
- While html uses tags, CSS uses rule sets. CSS is easy to learn and understand, but it provides powerful control over the presentation of an HTML document.
- The selector points to the HTML element you want to style.
- The declaration block contains one or more declarations separated by semicolons.
- Each declaration includes a CSS property name and a value, separated by a colon
- CSS declaration always ends with a semicolon, and declaration blocks are surrounded by curly braces.

2.1.3 JavaScript

JavaScript is a lightweight, cross-platform, and interpreted compiled programming language which is also known as the scripting language for webpages. It is well- known for the development of web pages; many non-browser environments also use it.JavaScript can be used for Client-side developments as well as Server-side developments. JavaScript is both imperative and declarative type of language. JavaScript contains a standard library of objects,

like Array, Date, and Math, and acore set of language elements like operators, control structures, and statements.

- Client-side: It supplies objects to control a browser and its Document Object Model(DOM). Like if client-side extensions allow an application to place elements on an HTML form and respond to user events such as mouse clicks, form input, and page navigation. Useful libraries for the client-side are AngularJS, ReactJS, VueJS andso many others.

- Server-side: It supplies objects relevant to running JavaScript on a server. Like if the server-side extensions allow an application to communicate with a database, and provide continuity of information from one invocation to another of the application, or perform file manipulations on a server. The useful framework whichis the most famous these days is node.js.

Imperative language – In this type of language we are mostly concern about how it is to be done. It simply controls the flow of computation. The procedural programming approach, object, oriented approach comes under this like asyncawait we are thinking what it is to be done further after async call.

- Declarative programming – In this type of language we are concern about how it is to be done, basically here logical computation require. Here main goal is to describe the desired result without direct dictation on how to get it like arrow function do.

2.1.3.1 History of JavaScript

It was created in 1995 by Brendan Eich while he was an engineer at Netscape. It was originally going to be named Live Script but was renamed. Unlike most programming languages, the JavaScript language has no concept of input or output. It is designed to run as a scripting language in a host environment, and it is up to the host environment to provide mechanisms for communicating with the outside world. The most common host environment is the browser.

2.1.3.2 Features of JavaScript

According to a recent survey conducted by Stack Overflow, JavaScript is the most popular language on earth.

With advances in browser technology and JavaScript having moved into the server with Node.js and other frameworks, JavaScript is capable of so much more. Here are a few things that we can do with JavaScript:

- JavaScript was created in the first place for DOM manipulation. Earlier websiteswere mostly static, after JS was created dynamic Web sites were made.

- Functions in JS are objects. They may have properties and methods just likeanother object. They can be passed as arguments in other functions.

- Can handle date and time.

2.1.3.3 Applications of JavaScript

- Web Development: Adding interactivity and behavior to static sites JavaScript wasinvented to do this in 1995. By using AngularJS that can be achieved so easily.

- Web Applications: With technology, browsers have improved to the extent that a language was required to create robust web applications. When we explore a mapin Google Maps then we only need to click and drag the mouse. All detailed view isjust a click away, and this is possible only because of JavaScript. It uses Application Programming Interfaces(APIs) that provide extra power to the code. The Electron and React is helpful in this department.

- Server Applications: With the help of Node.js, JavaScript made its way from client to server and node.js is the most powerful on the server-side.

- Games: Not only in websites, but JavaScript also helps in creating games for leisure. The combination of JavaScript and HTML 5 makes JavaScript popular in game development as well. It provides the EaseJS library which provides solutions for working with rich graphics.

- Smartwatches: JavaScript is being used in all possible devices and applications. It provides a library PebbleJS which is used in smartwatch applications. This framework works for applications that require the internet for its functioning.

- Art: Artists and designers can create whatever they want using JavaScript to draw on HTML 5 canvas, and make the sound more effective also can be used p5.js library.

- Machine Learning: This JavaScript ml5.js library can be used in web development by using machine learning.

- Mobile Applications: JavaScript can also be used to build an application for non- web contexts. The features and uses of JavaScript make it a powerful tool for creating mobile applications. This is a Framework for building web and mobile apps using JavaScrip

2.1.3.4 Limitations of JavaScript:

- Security risks: JavaScript can be used to fetch data using AJAX or by manipulating tags that load data such as <img>, <object>, <script>. These attacks are called cross site script attacks. They inject JS that is not the part of the site into the visitor's browser thus fetching the details.

- Performance: JavaScript does not provide the same level of performance as offered by many traditional languages as a complex program written in JavaScript would be comparatively slow. But as JavaScript is used to perform simple tasks in a browser, so performance is not considered a big restriction in its use.

- Complexity: To master a scripting language, programmers must have a thorough knowledge of all the programming concepts, core language objects, client and server-side objects otherwise it would be difficult for them to write advanced scripts using JavaScript.

- Weak error handling and type checking facilities: It is weakly typed language as there is no need to specify the data type of the variable. So wrong type checking is not performed by compile.

JavaScript can be added to your HTML file in two ways:

- Internal JS: We can add JavaScript directly to our HTML file by writing the code inside the <script> tag. The <script> tag can either be placed inside the <head> or the <body> tag according to the requirement.

- External JS: We can write JavaScript code in other file having an extension.js and then link this file inside the <head> tag of the HTML file in which we want to add this code.

2.1.4 BOOTSTRAP

Bootstrap is a free, open-source front-end development framework for the creation of websites and web apps. Designed to enable responsive development of mobile-first websites, Bootstrap provides a collection of syntax for template designs. Bootstrap is an HTML, CSS & JS Library that focuses on simplifying the development of informative web pages (as opposed to web apps). The primary purpose of adding it to a web project is to apply Bootstrap's choices of color, size, font and layout to that project. As such, the primary factor is whether the developers in charge find those choices to their liking.

Once added to a project, Bootstrap provides basic style definitions for all HTML elements. The

result is a uniform appearance for prose, tables and form elements across web browsers. In addition, developers can take advantage of CSS classes defined in Bootstrap to further customize the appearance of their contents. For example, Bootstrap has provisioned for light- and dark-coloured tables, page headings, more prominent pull quotes, and text with a highlight.

Bootstrap also comes with several JavaScript components which do not require other libraries like jQuery. They provide additional user interface elements such as dialog boxes, tooltips, progress bars, navigation drop-downs, and carousels. Each Bootstrap component consists of an HTML structure, CSS declarations, and in some cases accompanying JavaScript code. They also extend the functionality of some existing interface elements, including for example an auto-complete function for input fields.

2.1.5 jQuery

jQuery is an open source JavaScript library that simplifies the interactions between an HTML/CSS document, or more precisely the Document Object Model (DOM), and JavaScript. Elaborating the terms, jQuery simplifies HTML document traversing and manipulation, browser event handling, DOM animations, Ajax interactions, and cross-browser JavaScript development.

jQuery is widely famous with its philosophy of "Write less, do more." This philosophy can be further elaborated as three concepts:

- Finding some elements (via CSS selectors) and doing something with them (via jQuery methods) i.e., locate a set of elements in the DOM, and then do something with that set of elements.
- Chaining multiple jQuery methods on a set of elements.
- Using the jQuery wrapper and implicit iteration.
- Using jQuery (JS) library on HTML page.
- There are several ways to start using jQuery on your web site.
- Use the Google-hosted/ Microsoft-hosted content delivery network (CDN) toinclude a version of jQuery.
- Download own version of jQuery from jQuery.com and host it on own server orlocal filesystem.

2.1.5.1 Why jQuery?

Some of the key points which support the answer for why to use jQuery:

- It is incredibly popular, which is to say it has a large community of users and ahealthy number of contributors who participate as developers and evangelists.
- It normalizes the differences between web browsers so that you don't have to.
- It is intentionally a lightweight footprint with a simple yet clever pluginarchitecture.
- Its repository of plugins is vast and has seen steady growth since jQuery's release.
- Its API is fully documented, including inline code examples, which in the world of JavaScript libraries is a luxury. Heck, any documentation at all was a luxury for years.
- It is friendly, which is to say it provides helpful ways to avoid conflicts with other JavaScript libraries.

2.1.5.2 Advantages

- Wide range of plug-ins. jQuery allows developers to create plug-ins on top of the JavaScript library.
- Large development community
- It has a good and comprehensive documentation
- It is a lot easier to use compared to standard java script and other JavaScript libraries.
- jQuery lets users develop Ajax templates with ease,
- Ajax enables a sleeker interface where actions can be performed on pages without requiring the entire page to be reloaded.
- Being Light weight and a powerful chaining capability makes jQuery stronger.

2.1.5.3 Disadvantages:

- While JQuery has an impressive library in terms of quantity, depending on how much customization you require on your website, the functionality may be limited thus using raw JavaScript may be inevitable in some cases.
- The jQuery java script file is required to run jQuery commands, while the size of this file is relatively small (25-100KB depending on the server), it is still a strain onthe client computer and

maybe your web server as well if you intend to host the jQuery script on your own web server.

2.1.6 AJAX

AJAX is a web development technique for creating interactive web applications. Ifyou know JavaScript, HTML, CSS, and XML, then you need to spend just one hour tostart with AJAX

AJAX stands for Asynchronous JavaScript and XML. AJAX is a new technique for creating better, faster, and more interactive web applications with the help of XML, HTML, CSS, and Java Script.

Ajax uses XHTML for content, CSS for presentation, along with Document Object Model and JavaScript for dynamic content display. Conventional web applications transmit information to and from the sever using synchronous requests. It means you fill out a form, hit submit, and get directed to a new page with new information from the server. With AJAX, when you hit submit, JavaScript will make a request to the server, interpret the results, and update the current screen. In the purest sense, the user would never know that anything was even transmitted to the server.XML is commonly used as the format for receiving server data, although any format, including plain text, can be used. AJAX is a web browser technology independent of web server software. A user can continue to use the application while the client program requests information from the server in the background. Intuitive and natural user interaction. Clicking is not required, mouse movement is a sufficient event trigger. Data-driven as opposed to page-driven.

AJAX is the most viable Rich Internet Application (RIA) technology so far. It is getting tremendous industry momentum and several tool kit and frameworks are emerging. But at the same time, AJAX has browser incompatibility and it is supported by JavaScript, which is hard to maintain and debug.

AJAX is based on the following open standards −

- Browser-based presentation using HTML and Cascading Style Sheets (CSS).
- Data is stored in XML format and fetched from the server.
- Behind-the-scenes data fetches using XML HttpRequest objects in the browser.
- JavaScript to make everything happen.

2.2 BACKEND TECHNOLOGIES

Back-end development. It focuses on databases, scripting, website architecture. Itcontains behind-the-scene activities that occur when performing any action on a website. It can be an account login or making a purchase from an online store. Code written by back-end developers helps browsers to communicate with database information.

Backend is the server-side of the software that stores and analyzes data, as well as ensuring smooth application performance. Backend developers take on a range of duties, such as writing APIs, libraries and working with system components, business processes, and data architecture.

Being a process that stays invisible for users, backend development sends and receives information, communicates with the frontend, and displays the data as a web page.

2.2.1 MACHINE LEARNING

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. IBM has a rich history with machine learning. One of its own, Arthur Samuel, is credited for coining the term, "machine learning" with his research around the game of checkers. Robert Nealey, the self-proclaimed checkers master, played the game on an IBM 7094 computer in 1962, and he lost to the computer. Compared to what can be done today, this feat almost seems trivial, but it's considered a major milestone within the field of artificial intelligence. Over the next couple ofdecades, the technological developments around storage and processing power will enable some innovative products that we know and love today, such as Netflix's recommendation engine or self-driving cars. Machine learning is an important component of the growing field of data science.

Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

When we talk about machine learning, we're referring to a specific technique that allows a computer to "learn" from examples without having been explicitly programmed with step-by-step

instructions. Currently, machine learning algorithms are geared towardanswering a single type of question well. For that reason, machine learning algorithms are at the forefront of efforts to diagnose diseases, predict stock market trends, and recommend music.

Machine Learning is getting computers to program themselves. If programming is automation, then machine learning is automating the process of automation. Writing software is the bottleneck, we don't have enough good developers. Let the data do the work instead of people. Machine learning is the way to make programming scalable.

Sample applications of machine learning:

- **Web search:** ranking page based on what you are most likely to click on.
- **Computational biology:** rational design drugs in the computer based on pastexperiments.
- **Finance:** decide who to send what credit card offers to. Evaluation of risk on creditoffers. How to decide where to invest money.
- **E-commerce:** Predicting customer churn. Whether or not a transaction is fraudulent.
- **Space exploration:** space probes and radio astronomy.
- **Robotics:** how to handle uncertainty in new environments. Autonomous. Self-driving car.
- **Information extraction:** Ask questions over databases across the web.
- **Social networks:** Data on relationships and preferences. Machine learning to extractvalue from data.
- **Debugging:** Use in computer science problems like debugging. Labour intensiveprocess. Could suggest where the bug could be.
- There are tens of thousands of machine learning algorithms and hundreds of new algorithms are developed every year.

Every machine learning algorithm has three components:

- **Representation:** how to represent knowledge. Examples include decision trees, sets of rules, instances, graphical models, neural networks, support vector machines, model ensembles and others.
- **Evaluation:** the way to evaluate candidate programs (hypotheses). Examples include

accuracy, prediction and recall, squared error, likelihood, posterior probability, cost, margin, entropy k-L divergence and others.

- **Optimization:** the way candidate programs are generated known as the search process. For example, combinatorial optimization, convex optimization, constrained optimization.

Types of Learning There are three types of machine learning:

- **Supervised learning:** (also called inductive learning) Training data includes desiredoutputs. This is spam this is not, learning is supervised.

- **Unsupervised learning:** Training data does not include desired outputs. Example is clustering. It is hard to tell what good learning is and what is not.

- **Reinforcement learning:** Rewards from a sequence of actions. AI types like it, it isthe most ambitious type of learning.

2.2.1.1 Supervised Learning

Supervised learning is one of the most basic types of machine learning. In this type, the machine learning algorithm is trained on labeled data. Even though the data needs to be labeled accurately for this method to work, supervised learning is extremely powerful when used in the right circumstances.

In supervised learning, the ML algorithm is given a small training dataset to work with. This training dataset is a smaller part of the bigger dataset and serves to give the algorithm a basic idea of the problem, solution, and data points to be dealt with. The training dataset is also very similar to the final dataset in its characteristics and provides the algorithm with the labeled parameters required for the problem.

The algorithm then finds relationships between the parameters given, essentially establishing a cause-and-effect relationship between the variables in the dataset. At the end of the training, the algorithm has an idea of how the data works and the relationship between the input and the output.

2.2.1.2 Unsupervised Learning

Unsupervised machine learning holds the advantage of being able to work with unlabeled data. This means that human labor is not required to make the dataset machine-readable, allowing much larger datasets to be worked on by the program.

In supervised learning, the labels allow the algorithm to find the exact nature of the relationship

between any two data points. However, unsupervised learning does not have labels to work off of, resulting in the creation of hidden structures. Relationships between data points are perceived by the algorithm in an abstract manner, with no input required from human beings.

The creation of these hidden structures is what makes unsupervised learning algorithms versatile. Instead of a defined and set problem statement, unsupervised learning algorithms can adapt to the data by dynamically changing hidden structures. This offers more post-deployment development than supervised learning algorithms.

2.2.1.3 Reinforcement Learning

Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method. Favourable outputs are encouraged or 'reinforced', and non-favourable outputs are discouraged or 'punished'.

Based on the psychological concept of conditioning, reinforcement learning works by putting the algorithm in a work environment with an interpreter and a reward system. In every iteration of the algorithm, the output result is given to the interpreter, which decides whether the outcome is favourable or not.

In case of the program finding the correct solution, the interpreter reinforces the solution by providing a reward to the algorithm. If the outcome is not favourable, the algorithm is forced to reiterate until it finds a better result. In most cases, the reward system is directly tied to the effectiveness of the result.

In typical reinforcement learning use-cases, such as finding the shortest route between two points on a map, the solution is not an absolute value. Instead, it takes on a score of effectiveness, expressed in a percentage value.

2.2.2 PYTHON

Python has become one of the most popular programming languages in the world in recent years. It's used in everything from machine learning to building websites and

software testing. It can be used by developers and non-developers alike. Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems. This versatility, along with its

beginner-friendliness, has made it one of the most-used programming languages today. Python is commonly used for developing websites and software, task automation, data analysis, and data visualization. Since it's relatively easy to learn, Python has been adopted by many non-programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.

2.2.3 DJANGO

Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needingto reinvent the wheel.

Here are a few reasons why Django is the first choice when it comes to web development.

- **Backward compatible**

Django offers the provision of working with its older versions and makes use of its older formats and features

- **Community Support**

Django enjoys the support of a huge and very professional community of developers who have inside out knowledge of Django and are always ready to help. Having a large community has benefits of its own. It makes finding answers to problems much easier as there's a great probability that the problem faced by you now has occurred to someone else too and now probably has an answer to it on one or the other forums. The community is very quick in answering the issues and fixing the bugs of fellow developers.

- **DevOps Compatible**

DevOps is the blend of cultural philosophies, tools, and practices that increases anorganization's ability to deliver services and applications at high speed: evolving and upgrading products at a faster rate than organizations using conventional software development and infrastructure management processes. Incorporating DevOps in Django is great as it fixes issues faster with

enhanced operational support. It uses continuous delivery methodology to increase the efficiency of the system.

- **DRY and KISS compliant**

Django religiously follows the 'KISS' principle which is "Keep It Short and Simple". In Django, it simply means that the code must be brief, easily understandable, and methods should not exceed more than 50-60 lines.

- **Infrastructure**

Django is independent and a complete set in itself. It means that it does not require any other external solution. It is everything, from an ORM to a web server. This enables it to use various databases and switch them accordingly.

- **REST framework**

REST stands for Representational State Transfer framework which is a renowned toolkit for creating web APIs. This is an added benefit with Django as it is powerful enough to build a full-fledged API in just two or three lines of code. An additional benefit to it is that REST is immensely flexible. Therefore, data is not bound to any protocol and can return various data formats and manage several types of calls.

- **Secure & up-to-date**

Django is constantly kept up to an elevated standard, following the most recent patterns in site development and security. That certainly addresses the inquiry "Is Django useful for web development?" — As security is a 'number one' in any to-do list. Django is customarily updated with security fixes, and regardless of whether you're utilizing an older rendition of the system, its security is as intact as the new one. It's no big surprise as Django has an LTS (Long-term Support) variant.

- **Simple**

Django's documentation is excellent. It was introduced with excellent documentation, andthey are as yet kept intact in the same way, which makes it simple to utilize. Also, one of Django's principal reasons for existing is to disentangle the development procedure: it covers the fundamentals, so you can concentrate on the details of your project.

- **Suits any kind of project**

Unlike C# or Java, Django is no business solution. Still, it is appropriate for all kinds of projects regardless of their size. A Django application can be a social media application with enormous traffic and heavy volumes of data or it could be a simple web application for handling logbooks. e.

- **Time effective**

Django is ridiculously fast. It was built to slide the applications from imagination to reality in a blink. Django applications are both economical and efficient. Thus it is the right choice for developers who have major stakes placed on the due dates.

2.3 SOFTWARE ENGINEERING

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS). Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database.

Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized. Most databases use structured query language (SQL) for writing and querying data.

2.3.1 AGILE

Agile is an iterative approach to project management and software development that helps teams deliver value to their customers faster and with fewer headaches. Instead ofbetting everything on a "big bang" launch, an agile team delivers work in small, but consumable, increments. Requirements, plans, and results are evaluated continuously soteams have a natural mechanism for responding to change quickly.

Agile project management is an iterative approach to managing software development projects that focuses on continuous releases and incorporating customer feedback with every iteration.

Software teams that embrace agile project management methodologies increase their development speed, expand collaboration, and foster the ability to better respond to market trends.

The Agile methodology is a way to manage a project by breaking it up into several phases. It

involves constant collaboration with stakeholders and continuous improvement at every stage. Once the work begins, teams' cycle through a process of planning, executing, and evaluating. Continuous collaboration is vital, both with team members and project stakeholders.

The Agile Manifesto of Software Development put forth a groundbreaking mindset on delivering value and collaborating with customers when it was created in 2001. Agile's four main values are:

- Individuals and interactions over processes and tools
- Working software over comprehensive documentation
- Customer collaboration over contract negotiation
- Responding to change over following a plan

So what is Agile methodology in project management? It's a process for managing a project that involves constant collaboration and working in iterations. Agile project management works off the basis that a project can be continuously improved upon throughout its life cycle, with changes being made quickly and responsively.

2.3.1.1 ADVANTAGES OF AGILE MANAGEMENT SOFTWARE

- Adaptability –Agile management software is adaptable to change
- Progressive improvement—by assessing the team's performance in past sprints, thesoftware assists them in improving their efficiency.
- Collaboration between members improves, allowing them to become more self-organized and cross-functional.
- Motivation-Setting clear and attainable goals helps to enhance motivation.
- Metrics that matter for data-driven decisions
- Team Performance-Through agile management tools, an emphasis is placed on outcomes such as enhancing team performance and developing better software.
- Predictability and control in project planning and execution—estimation of budgets,deadlines, hazards, and more.
- The Product Quality – agile management tool that breaks the project into manageable chunks, allowing the team to focus more easily on quality developmentand testing.

- Consumer satisfaction-centric concepts and principles necessitate active customer participation in all phases of development.

- Faster Return on Investment (ROI)-using agile means faster software release, as you obtain a working software after the first sprint.

2.3.1.2 WHY AGILE MANAGEMENT TOOL?

- Reach Business Objectives of Any Complexity: Decompose your Objective into Key Results of numerous sublevels, and duties. Create an Objective Map of any complexity to achieve the hardest Objectives

- Create Clear Plans to Achieve Your Objectives Effectively: Any movement is associated with a selected Objective, and every Objective has its motion plan

- Have a Helicopter View on Your Objectives: Benefit from all of the needed statistics accessed at the proper time and located within the proper area

- Manage Your Team to Achieve Your Objectives: Analyze the consequences and performance of your team in standard and every crew member especially in accordance together with your Objective.

- Control Your Progress with Key Success Factors: Using measurable achievement standards, you obtain your Objective with a particular KPI, on the planned time, and with the needed budget

- Analyze Risks and Adjust Your Plans if Necessary: Always see your weaknesses and easily exchange your modern-day plans to reap your Objectives greater efficaciously.

- The right goal alignment doubles the productivity of any enterprise: Make changes to your plans and constantly keep them up to date. The Sprints functionality helps you to cope with the toughest responsibilities

2.3.2 KANBAN

Kanban is a popular framework used to implement agile and DevOps software development. It requires real-time communication of capacity and full transparency of work. Work items are represented visually on a kanban board, allowing team membersto see the state of every piece of work at any time. Kanban is a process management toolthat visualises the status of each job on a company's radar, and controls the flow of production from customer requests back to the warehouse. Kanban is a process management tool that visualises the status of each job on a

company's radar, and controls the flow of production from customer requests back to the warehouse.

Principles : -

- Start with what you do now.
- Agree to pursue incremental, evolutionary change.
- Respect the current process, roles, responsibilities, and titles
- Encourage acts of leadership at all levels in your organization

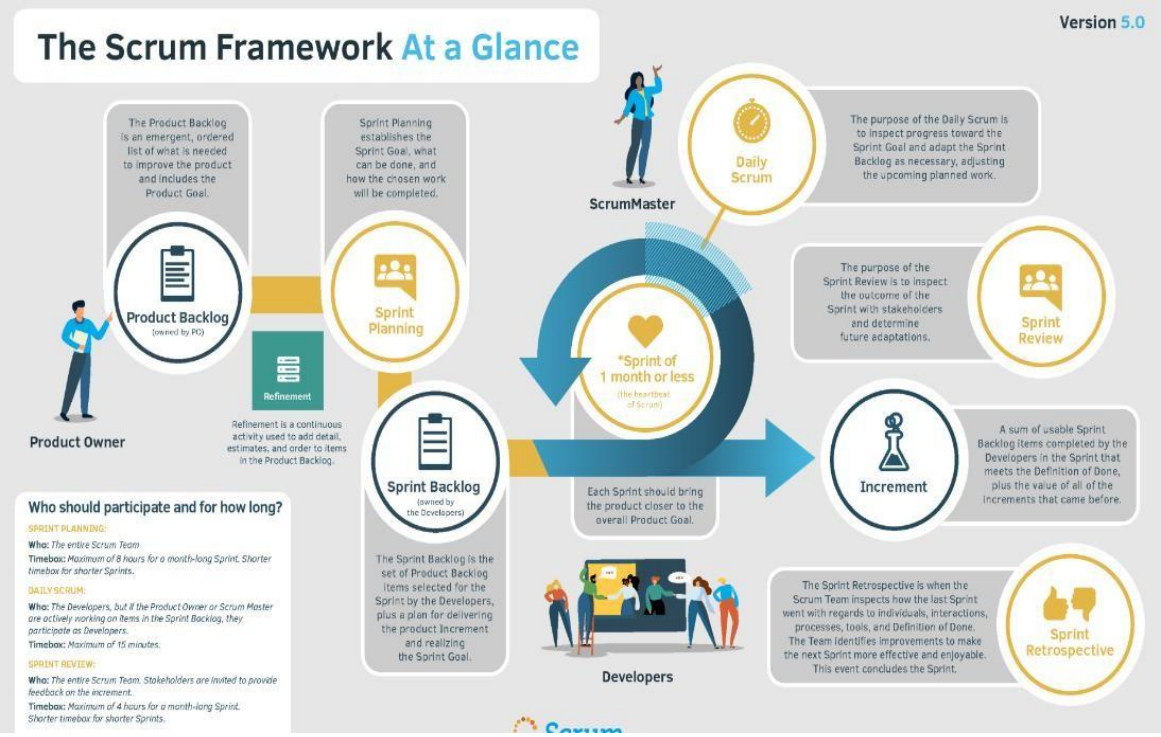


Figure 6 *Scrum Framework*

2.3.3 SCRUM

Scrum is a framework for project management that emphasizes teamwork, accountability and iterative progress toward a well-defined goal. The framework begins with a simple premise: Start with what can be seen or known.

The Six Scrum Principles are:

1. Control over the empirical process

2. Self-organization

3. Collaboration

4. Value-based prioritization

5. Time-boxing

6. Iterative development

2.4 DATABASE

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS). Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database.

Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized. Most databases use structured query language (SQL) for writing and querying data.

2.4.1 SQL

SQL is a language to operate databases; it includes database creation, deletion, fetching rows, modifying rows, etc. SQL is an ANSI (American National Standards Institute) standard language, but there are many different versions of the SQL language. SQL is Structured Query Language, which is a computer language for storing, manipulating and retrieving data stored in a relational database. SQL is the standard language for RelationalDatabase System.

All the Relational Database Management Systems (RDMS) like MySQL, MS Access, Oracle, Sybase, Informix, Postgres and SQL Server use SQL as their standard database language.

Also, they are using different dialects, such as:

- MS SQL Server using T-SQL,
- Oracle using PL/SQL,
- MS Access version of SQL is called JET SQL (native format) etc.

SQL is widely popular because it offers the following advantages:

- Allows users to access data in the relational database management systems.

- Allows users to describe the data.
- Allows users to define the data in a database and manipulate that data.
- Allows to embed within other languages using SQL modules, libraries & pre-compilers.
- Allows users to create and drop databases and tables.
- Allows users to create view, stored procedure, functions in a database.
- Allows users to set permissions on tables, procedures and views.

2.5 Hardware Requirements

Name of Equipment	Quantity
PC with minimum 8GB/1TBConfiguration	3

Table 1 Hardware Requirements

2.6 Software Requirements

Name of Technology	Platform	Used For
Python (Tkinter, Django) ,HTML, CSS, JS	Anaconda	Front End
Python (OpenCV, Machine Learning)	Anaconda	Back End
CSV,SQL	CSV,SQL	Database
Figma	Figma	Designing
Kanban Board	Jira	Project Management

Table 2 Software Requirements

Chapter 3

Development Methodology

The software development methodology framework didn't emerge until the 1960s. According to Elliott the systems development life cycle (SDLC) can be considered to be the oldest formalized methodology framework for building information systems. The main idea of the SDLC has been "to pursue the development of information systems in a very deliberate, structured and methodical way, requiring each stage of the life cycle from inception of the idea to delivery of the final system, to be carried out in rigidly and sequentially".within the context of the framework being applied. The main target of this methodology framework in the 1960s was "to develop large scale functional business systems in an age of large scale business conglomerates. Information systems activities revolved around heavy data processing and number crunching routines"

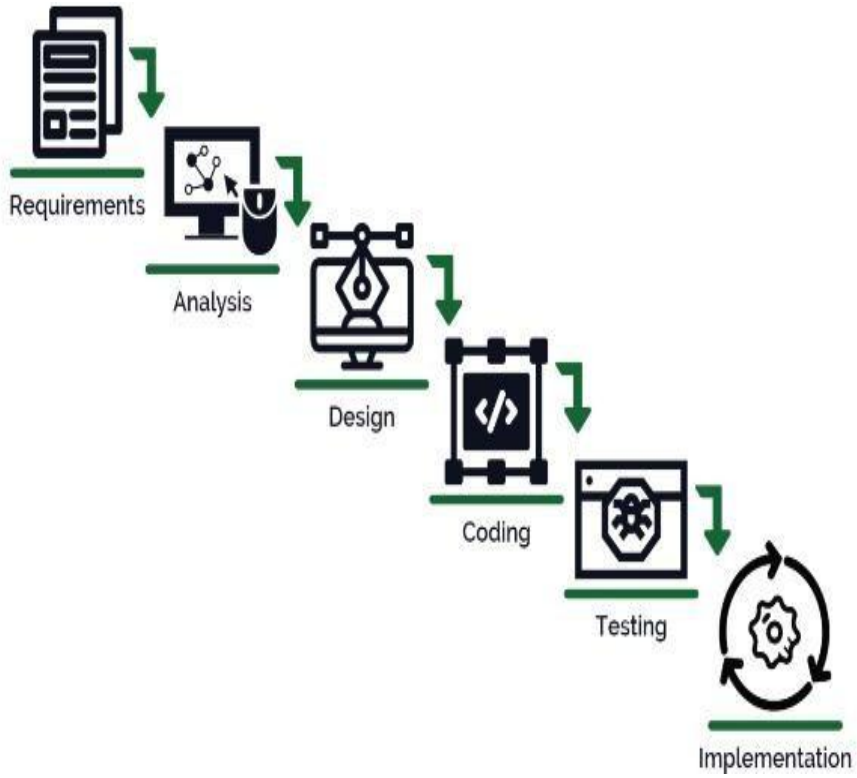


Figure 7 *Agile Methodology*

STEPS :-

Step 1: Analysis of all the requirements. **Step 2:** Gathering of data for the project. **Step 3:** Analysing the Data for training.

Step 4: Training the data and calculating efficiency.

Step 5: Designing of the layout of different modules and front end.

Step 6: Making every module functional.

Step 7: Make a single unit by combining all modules.

Step 8: Proceed for the testing of the designed system.

Step 9: Check whether the system is compatible with every focused environment ornot.

Step 10: Make changes if required.

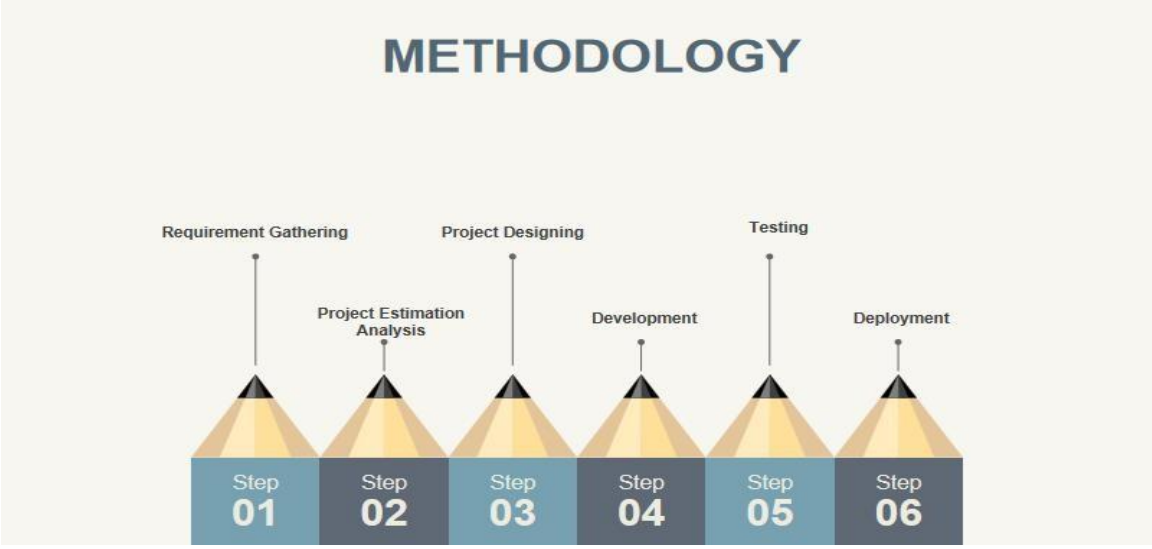


Figure 8 *Product Methodology*

3.1 Analysis of all the requirements

Requirements analysis or requirements engineering is a process used to determine the needs and expectations of a new product. It involves frequent communication with the stakeholders and end-users of the product to define expectations, resolve conflicts, and document all the key requirements. One of the greatest challenges faced by any organization is to share the vision of the final product with the customers. Hence, a business requirements analysis involves a team

effort of all the key stakeholders, software developers, end-users, and customer managers to achieve a shared understanding of what the product should do. This is always done in the early phase of any project to ensure that the final product conforms to all the requirements.

3.2 Project Estimation

Project estimation is the process of forecasting the time, cost, and resources needed to deliver a project. It typically happens during project initiation and or planning and takes the project's scope, deadlines, and potential risks into account. A project estimate gives you and your stakeholders a general idea of how much time, effort, and money it'll take to get the job done. That makes it easier to build a feasible project budget and plan so you can set your team and organization up for success.

3.3 Project Designing

During the Design Phase, the system is designed to satisfy the requirements identified in the previous phases. The requirements identified in the Requirements Analysis Phase are transformed into a System Design Document that accurately describes the design of the system and that can be used as an input to system development in the next phase. The purpose of the Design Phase is to transform the requirements into complete and detailed system design specifications. Once the design is approved, the Development Team beginsthe Development Phase.

3.4 Project Development

Project performance and monitoring ensures that project results align with the management plan.

Project Objectives: Measuring if a project is on schedule and budget is an indication if the project will meet stakeholder objectives.

Quality Deliverables: This determines if specific task deliverables are being met.

Project Performance: This monitors changes in the project. It takes into consideration the amount and types of issues that arise and how quickly they are addressed. These can occur from unforeseen hurdles and scope changes.

3.5 Testing

During the testing phase, we find out whether their code and programming work according to

requirements. And while it's not possible to solve all the failures you might find during the testing phase; it is possible to use the results from this phase to reduce the number of errors within the software program.

Before testing can begin, we developed a test plan. The test plan includes the types of testing we'll be using, resources for testing, how the software will be tested. Test scripts ensure consistency while testing

- White Box
- B;lack Box

3.6 Deployment

We Successfully implemented the programming and coding to each system location. The deployment phase includes pushing the program and coding to each regional site and each computer system.

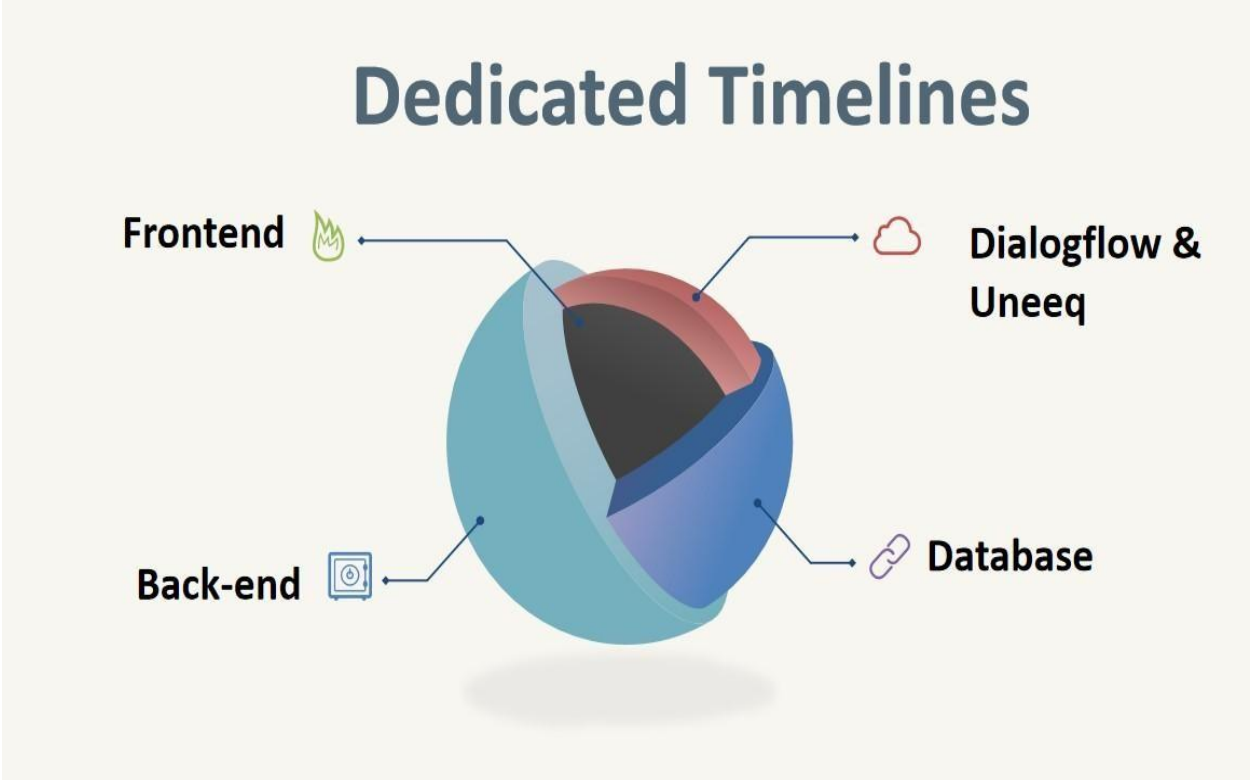


Figure 9 *Dedicated Timelines*

S.No.	Work to be Done
1.	Mission Planning
2.	Concept Development
3.	Documentation
4.	Detailed Design
5.	Project System Design
6.	Implementing functionality into modules
7.	Combining all functionality into single unit
8.	Testing Phase I
9.	Testing Phase II
10.	Deployment

Table 3 Workflow planing

Chapter 4

Disease Analysis

The intelligent diagnosis problem to be solved in this project includes two aspects. The first one is seeking diagnosis experience according to the relationships between symptoms and diseases. Here, supervised machine learning methods are adopted. The second one is disease prediction based on the symptoms provided by visitors. The first one is the main problem.

In our Project, symptoms have been extracted in data preprocessing. Consider that samples with respect to the same disease type are in a hyperplane and linearly separable, and a different symptom may make two similar samples refer to different disease types; the support vector machine (SVM) algorithm is an appropriate method.

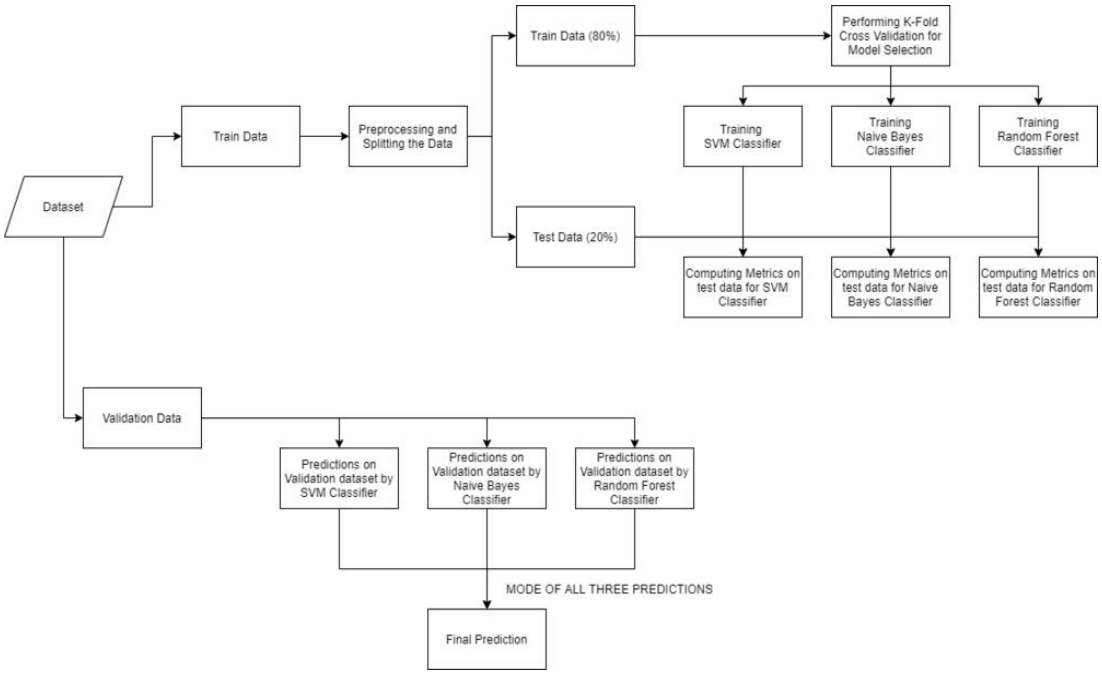


Figure 10 *Workflow of Model selection*

From an open-source dataset, an excel sheet was created where we listed down all the symptoms for the respective diseases. After which depending on the diseases, age and gender were specified as a part of the dataset. We listed down around 230 diseases with more than 1000 unique symptoms in all. The symptoms, age, and gender of an individual were used as input to various

machine learning algorithms.

Naive Bayes It is a machine learning algorithm for classification problems and is based on Bayes' probability theorem. The primary use of this is to do text classification which involves high dimensional training data sets. We used the Bayes theorem that can be defined as:

$$P(h|d) = \frac{P(d|h) \cdot P(h)}{P(d)}$$

Where P(h|d) is the probability of hypothesis h given the data d. This is called the posterior probability. P(d|h) is the probability of data d given that the hypothesis h was true. P(h) is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h. P(d) is the probability of the data (regardless of the hypothesis).

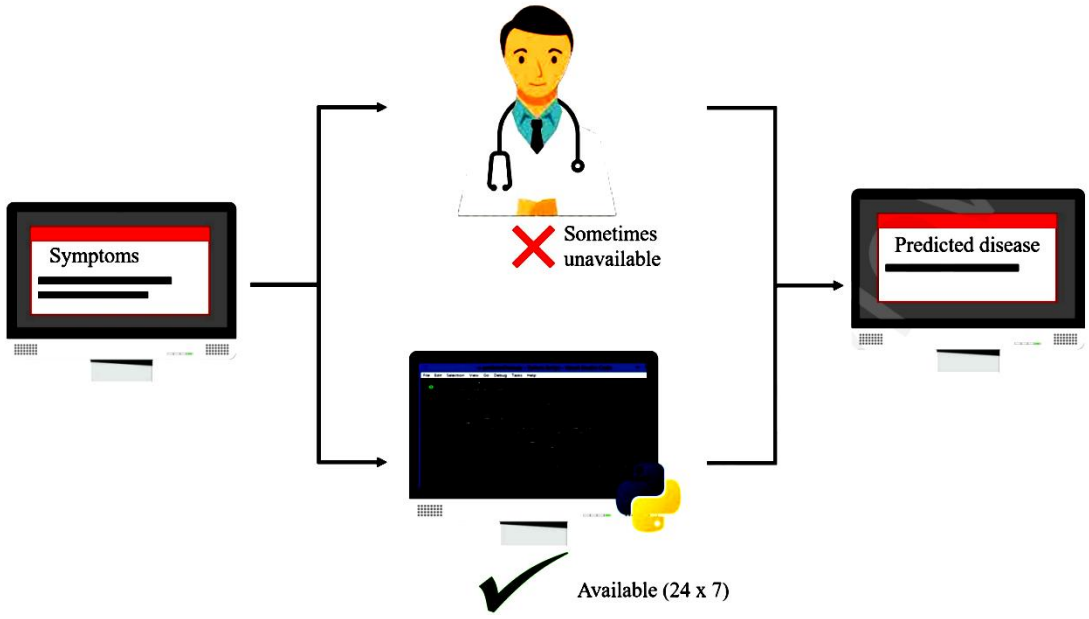


Figure 11 *Workflow of Model*

4.1. Knowledge Base

The knowledge base may include FAQs, manuals, troubleshooting guides, runbooks,and other information your team may want or need to know. Many knowledge bases are structured around artificial intelligence that can interact and respond to user input. Building a knowledge base can help you scale out your customer support and customer success efforts — without having to add new customer support staff. As such, it can help you reduce costs, improve customer satisfaction.

Clarity: Use action-based headlines to help your audience know at-a-glance what the content will

explain. Explain all jargon or technical terms in clear language.

Readability: Readers prefer content that provides information at-a-glance. That means paragraphs should be short and blocks of text should be broken up with headings, subheadings, bullets or numbered lists where applicable.

Engagement: Include images, charts, infographics, or videos within content or as stand- alone content to increase engagement.

Utility: Be sure to link to related articles within your knowledge base. This will help your audience easily find other resources that fully answer their questions.

Value: Every piece of content in your knowledge base should offer a valuable solution or insight.

4.2 Gathering of data for the project.

Data collection is the process of gathering quantitative and qualitative information on specific variables with the aim of evaluating outcomes or gleaning actionable insights. Good data collection requires a clear process to ensure the data you collect is clean, consistent, and reliable. Establishing that process, however, can be tricky. It involves taking stock of your objectives, identifying your data requirements, deciding on amethod of data collection, and finally organizing a data collection plan that synthesizes the most important aspects of your program. Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

Firstly loading the dataset from the folders using the pandas library. While reading the dataset we will be dropping the null column. This dataset is a clean dataset with no null values and all the features consist of 0's and 1s. Whenever solving a classification task it is necessary to check whether our target column is balanced or not. using a bar plot, to check whether the dataset is balanced or not.

From the above plot, observe that the dataset is a balanced dataset i.e. there are exactly 120 samples for each disease, and no further balancing is required. Notice that our target column i.e. prognosis column is of object datatype, this format is not suitable to train a machine learning model. So, will be using a label encoder to convert the prognosis column to the numerical datatype. Label Encoder converts the labels into numerical form by assigning a unique index to the labels. If the total number of labels is n, then the numbers assigned to each label will be

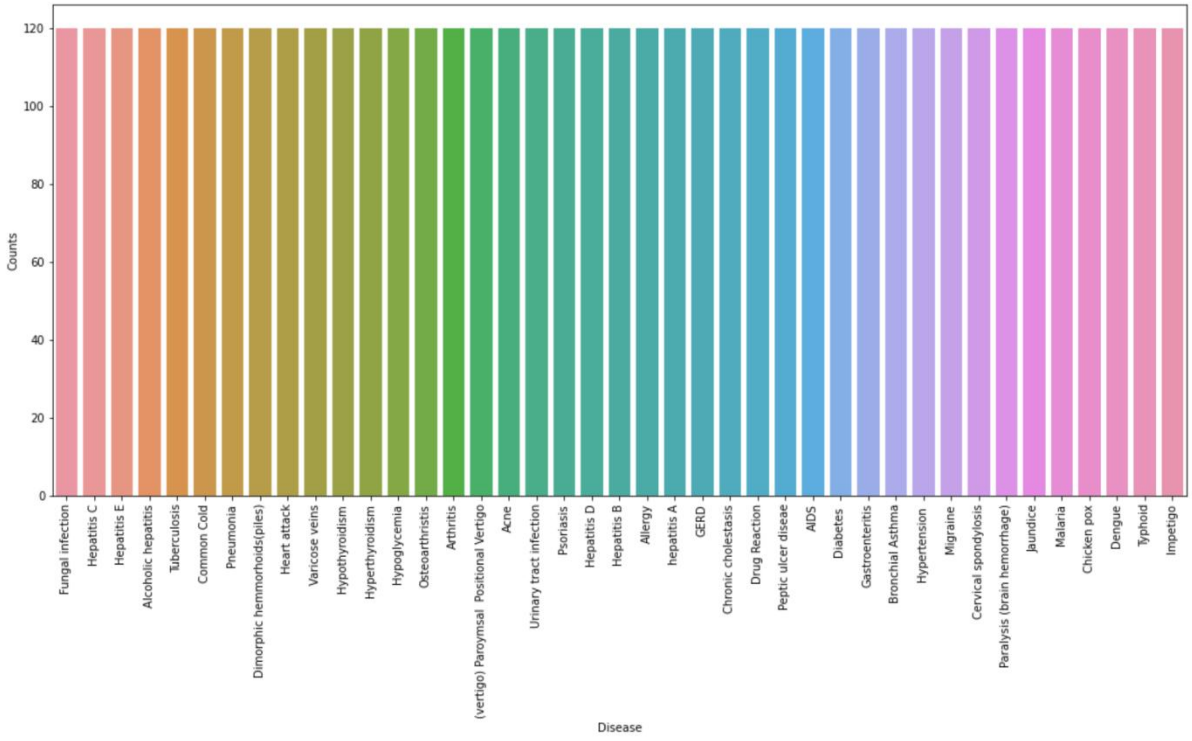
between 0 to n-1.

```
DATA_PATH = "dataset/Training.csv"
data = pd.read_csv(DATA_PATH).dropna(axis = 1)

# Checking whether the dataset is balanced or not
disease_counts = data["prognosis"].value_counts()
temp_df = pd.DataFrame({
    "Disease": disease_counts.index,
    "Counts": disease_counts.values
})

plt.figure(figsize = (18,8))
sns.barplot(x = "Disease", y = "Counts", data = temp_df)
plt.xticks(rotation=90)
plt.show()
encoder = LabelEncoder()
data["prognosis"] = encoder.fit_transform(data["prognosis"])
```

Figure 12 *Reading of data Set*



Graph 1 *Reading of data Set*

4.3 Analysing the Data for training.

Data analysis is whenever we take any decision in our day-to-day life is by thinking about what happened last time or what will happen by choosing that particular decision. This is nothing but

analyzing our past or future and making decisions based on it.

- **Identify training programme problem areas.** This can be done through analysis of the data you collect.

- **Identify similar training programmes.** These could be programmes with, for example, similar learner profiles, content, training methods etc.

- **Identify 'best practice' training programmes.** Using comparable evaluation methods, look for the best results achieved from the training programmes you have identified.

- **Establish the differences.** Identify which of the best practice processes could be adopted by future programmes.

- **Develop plans and targets for future performance.** Enable inclusion of best practices into future training development and delivery, and set targets for performance based on best practice outcomes.

- **Communicate.** Ensure everyone involved in the relevant training development and delivery processes is aware of the new implementation plans and targets.

- **Implement.** Put the implementation plans into practice when developing and delivering new training programmes.

- **Review and recalibrate.** Review how successful the new processes have been and whether targets have been met. If the results fall short of expectations, reviewprocesses, identify problem areas and potential causes, and set actions and targets foraddressing these.

```
X = data.iloc[:,:-1]
y = data.iloc[:, -1]
X_train, X_test, y_train, y_test =train_test_split(
  X, y, test_size = 0.2, random_state = 24)

print(f"Train: {X_train.shape}, {y_train.shape}")
print(f"Test: {X_test.shape}, {y_test.shape}")

Output :

Train: (3936, 132), (3936,)
Test: (984, 132), (984,)
```

Figure 13 *Reading of size of Data set*

4.4. Training the data and calculating efficiency.

- **Level 1: Reaction –** The first step is to evaluate the learners' reactions and responses to the training.
- **Level 2: Learning –** The second step is to measure the knowledge and skills learned during the training.
- **Level 3: Behaviour –** Step three assesses the behavioural change (if any and to what extent) due to the training.
- **Level 4: Impact –** The final step is to measure the training's impact on business goals and results
-

After splitting the data, will be now working on the modeling part. Using K-Fold cross-validation to evaluate the machine-learning models. Using Support Vector Classifier, Gaussian Naive Bayes Classifier, and Random Forest Classifier for cross-validation. Before moving into the implementation part let us get familiar with k-fold cross-validation and the machine learning models.

- K-Fold Cross-Validation: K-Fold cross-validation is one of the cross-validation techniques in which the whole dataset is split into k number of subsets, also known as folds, then training of the model is performed on the k-1 subsets and the remaining one subset is used to evaluate the model performance.
- Support Vector Classifier: Support Vector Classifier is a discriminative classifier i.e. when given a labeled training data, the algorithm tries to find an optimal hyperplane that accurately separates the samples into different categories in hyperspace.
- Gaussian Naive Bayes Classifier: It is a probabilistic machine learning algorithm that internally uses Bayes Theorem to classify the data points.
- Random Forest Classifier: Random Forest is an ensemble learning-based supervised machine learning classification algorithm that internally uses multiple decision trees to make the classification. In a random forest classifier, all the internal decision trees are weak learners, and the outputs of these weak decision trees are combined i.e. mode of all the predictions is

as the final prediction.

```
============================================================
SVC
Scores: [1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
Mean Score: 1.0
============================================================
Gaussian NB
Scores: [1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
Mean Score: 1.0
============================================================
Random Forest
Scores: [1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
Mean Score: 1.0
```

Figure 13 *Mean Score of model*

```
Accuracy on train data by Random Forest Classifier: 100.0
Accuracy on test data by Random Forest Classifier: 100.0
```

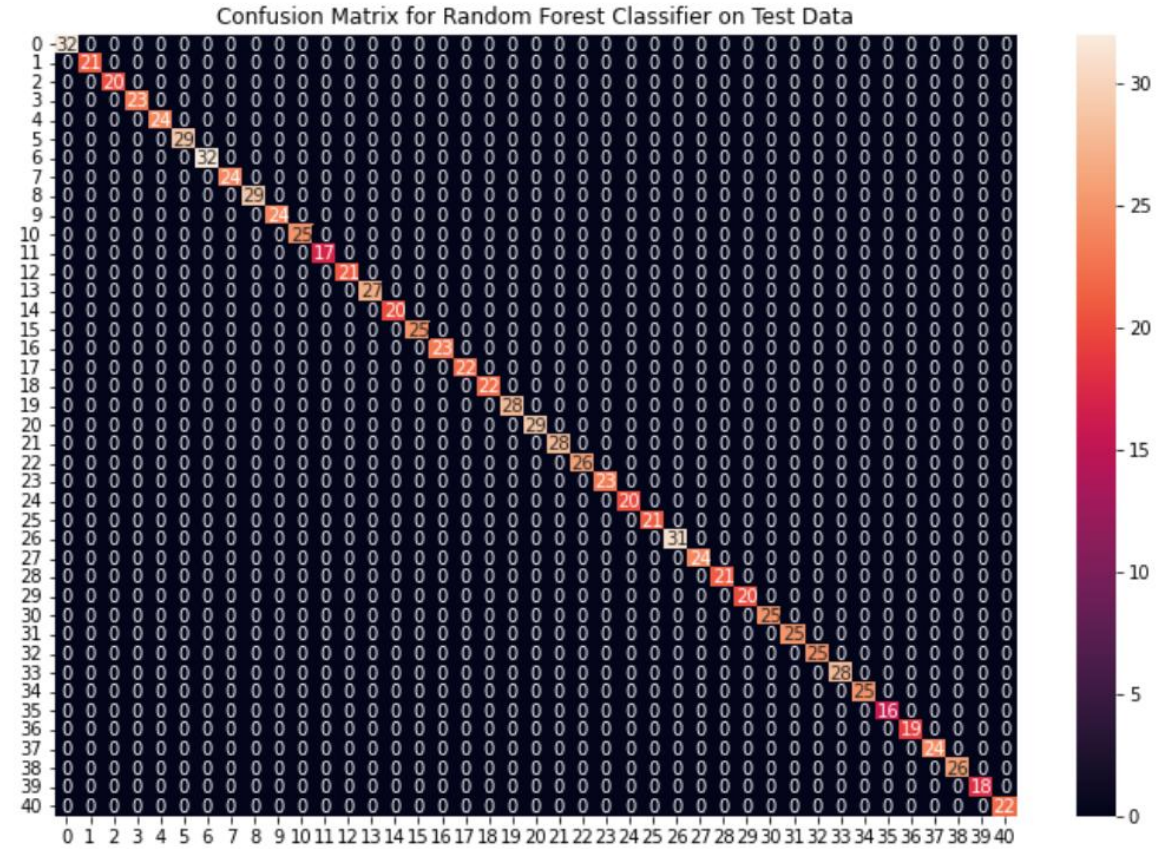
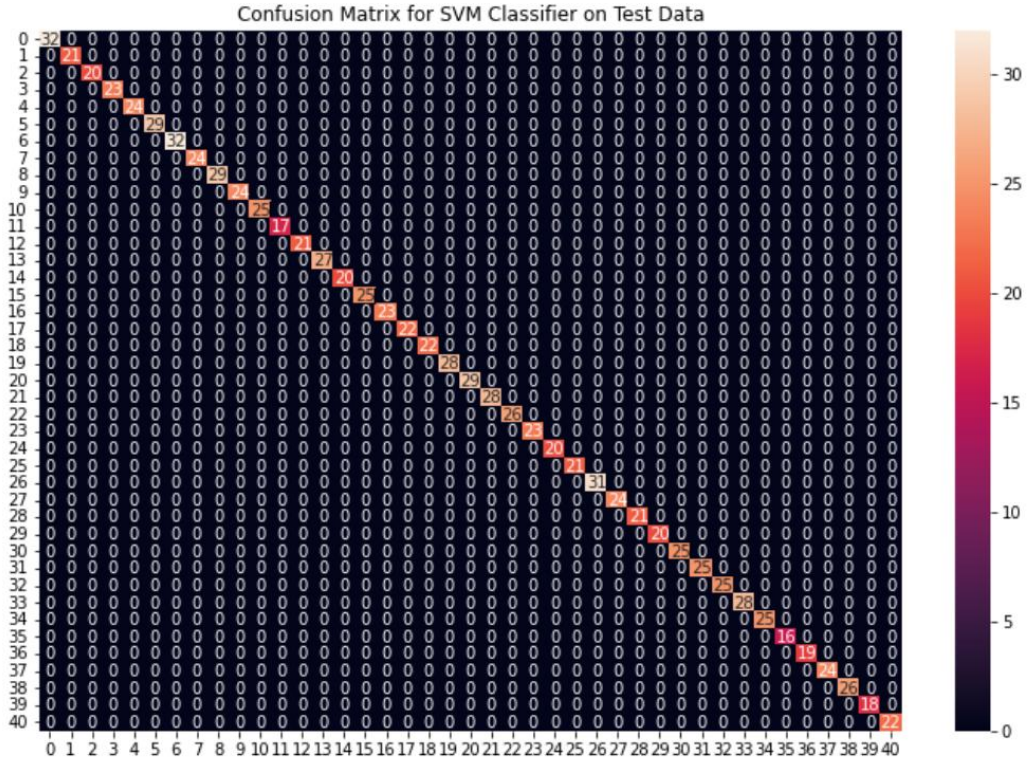


Figure 14 *Confusion Matrix for Random Forest Classifier*

```
Accuracy on train data by SVM Classifier: 100.0
Accuracy on test data by SVM Classifier: 100.0
```




```
Accuracy on train data by Naive Bayes Classifier: 100.0
Accuracy on test data by Naive Bayes Classifier: 100.0
```

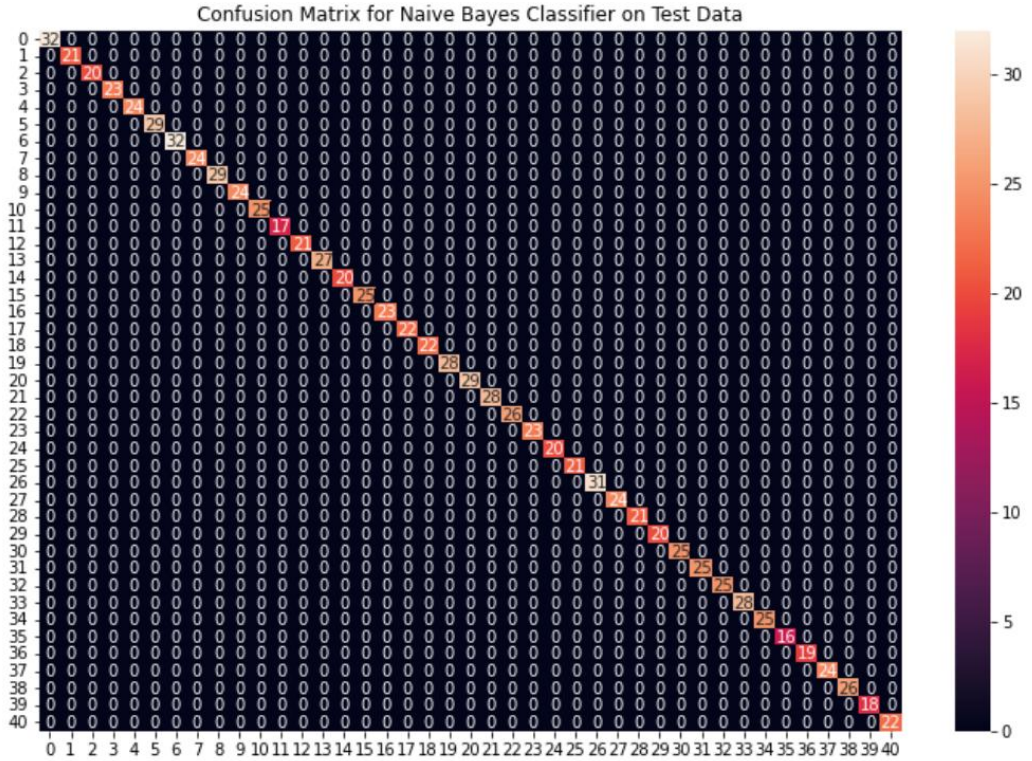


Figure 15 *Confusion Matrix for SVM and Naive Bayes*

4.5 Disease Recognition

Suppose that a preprocess step has been implemented on existing electronic medical records. Disease symptoms, disease types, and relations between the two are known clearly. Number the N disease symptoms in the database, and the symptoms are numbered as 1,2,3,...,N, respectively. Considering that the same symptoms in different gender patients are often with regard to different common diseases or disease types, gender is deemed as a default "symptom," which is labelled 1. Diseases in the database are divided into main categories, Each main disease category is further divided into several subclasses and numbered. Establish a data relationship list, in which the data structure is (Symptom 1, Symptom 2, Symptom 3, …, Symptom N, Disease). Each entry contains N symptoms. If symptom does exist in the item of a disease, the value below "Symptom " is 1, or else, the value is 0.

The symptoms, which are provided by a patient, are k then Identification of Subclass Disease Types, Identification of Specific Diseases are done.

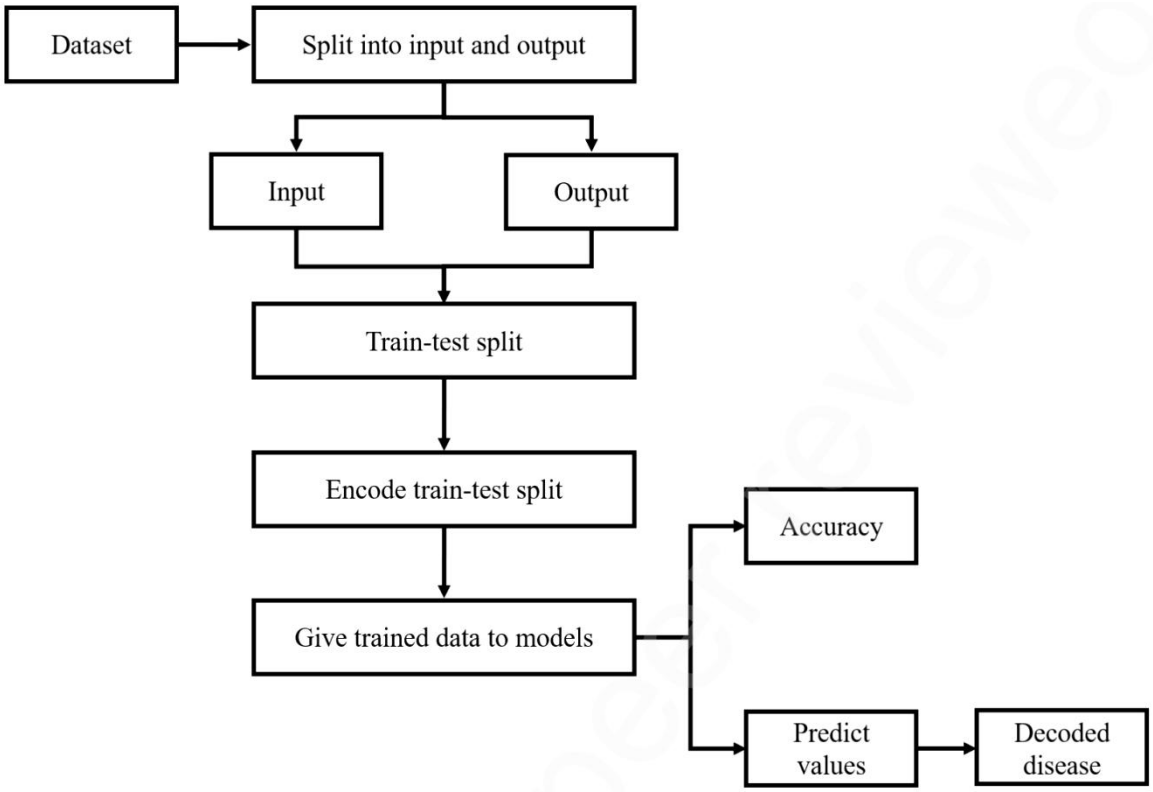


Figure 16 *Working of Algorithm*

Chapter 5

Glance at MeDiagnose

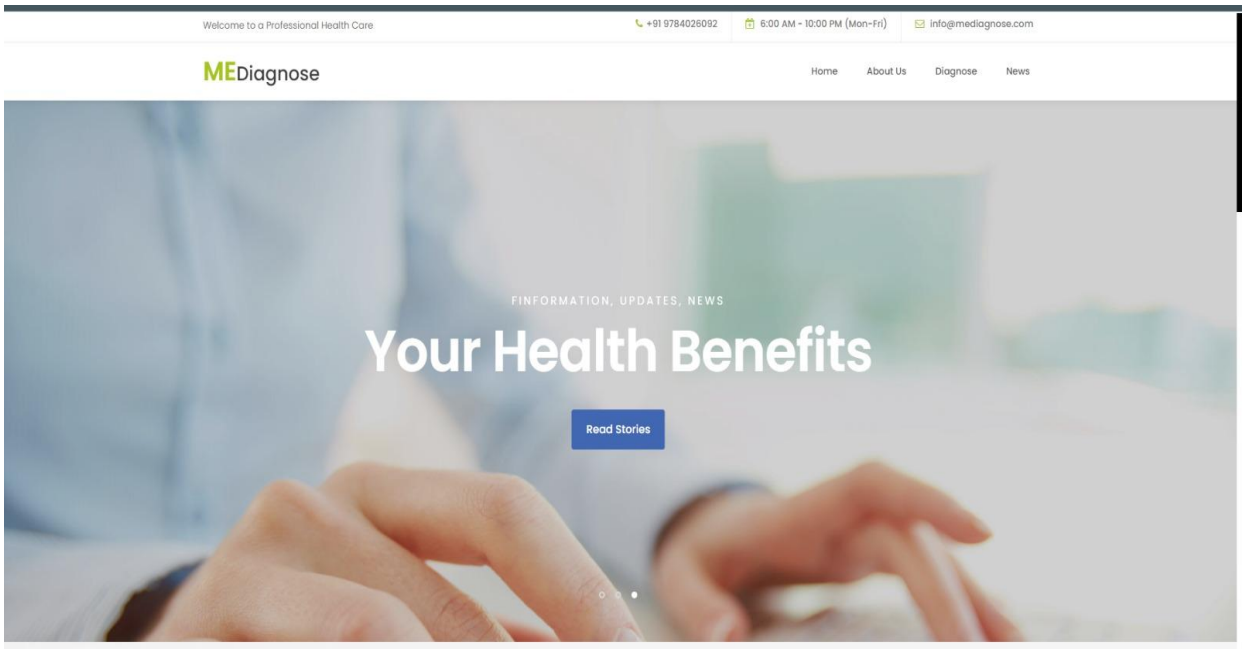


Figure 17 *Slide 1for MeDiagnose Home*

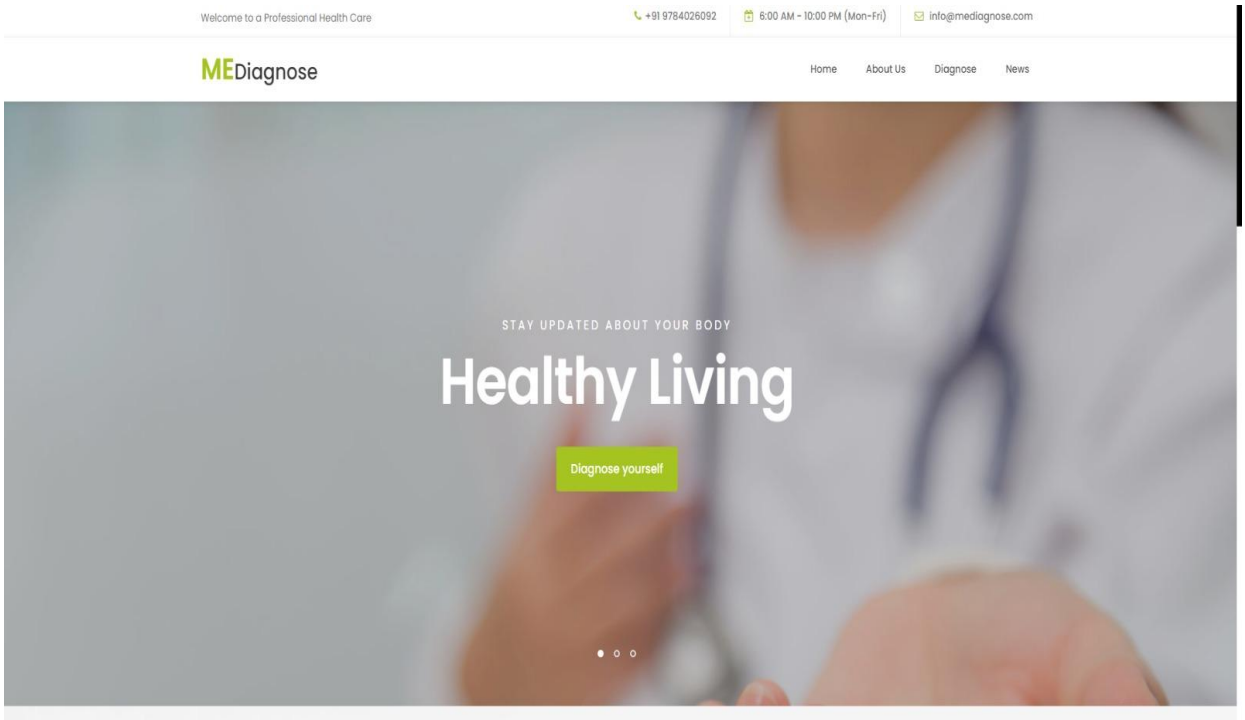


Figure 18 *Slide2 for MeDiagnose Home*

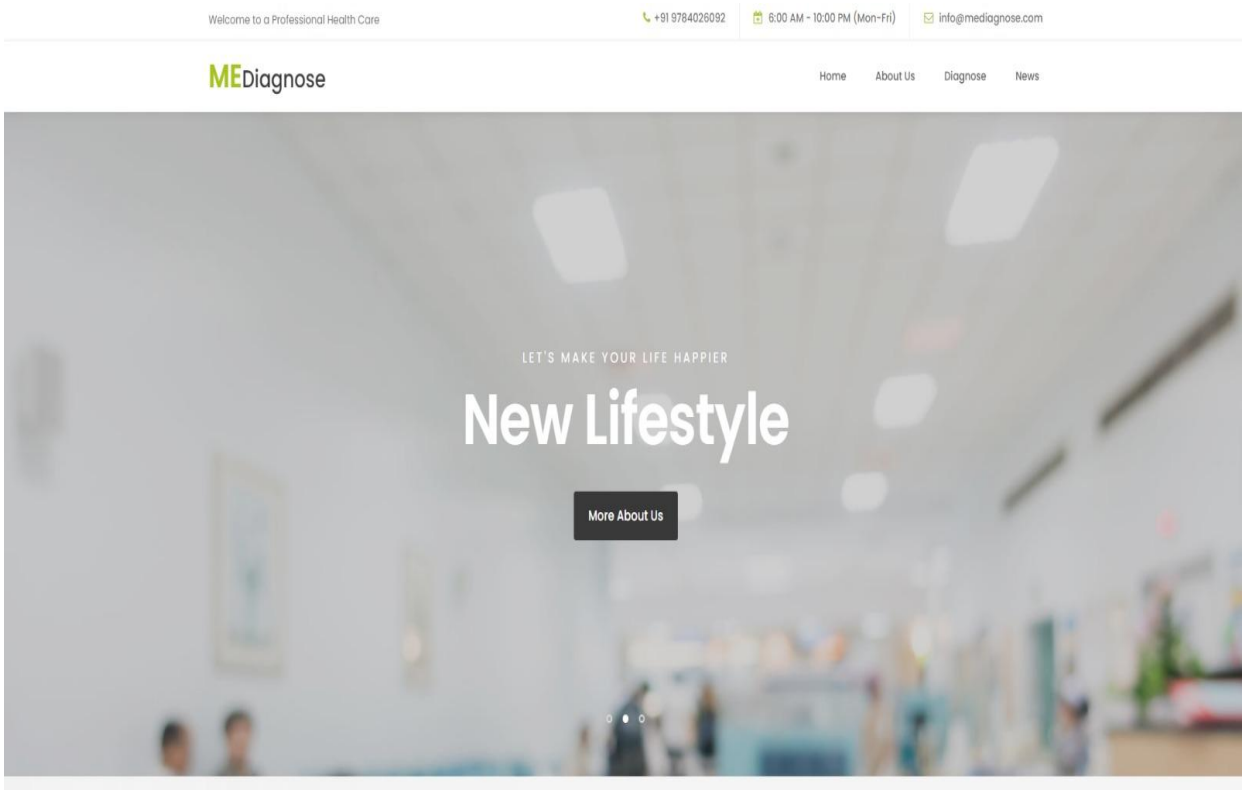


Figure 19 *Slide3 for MeDiagnose Home*

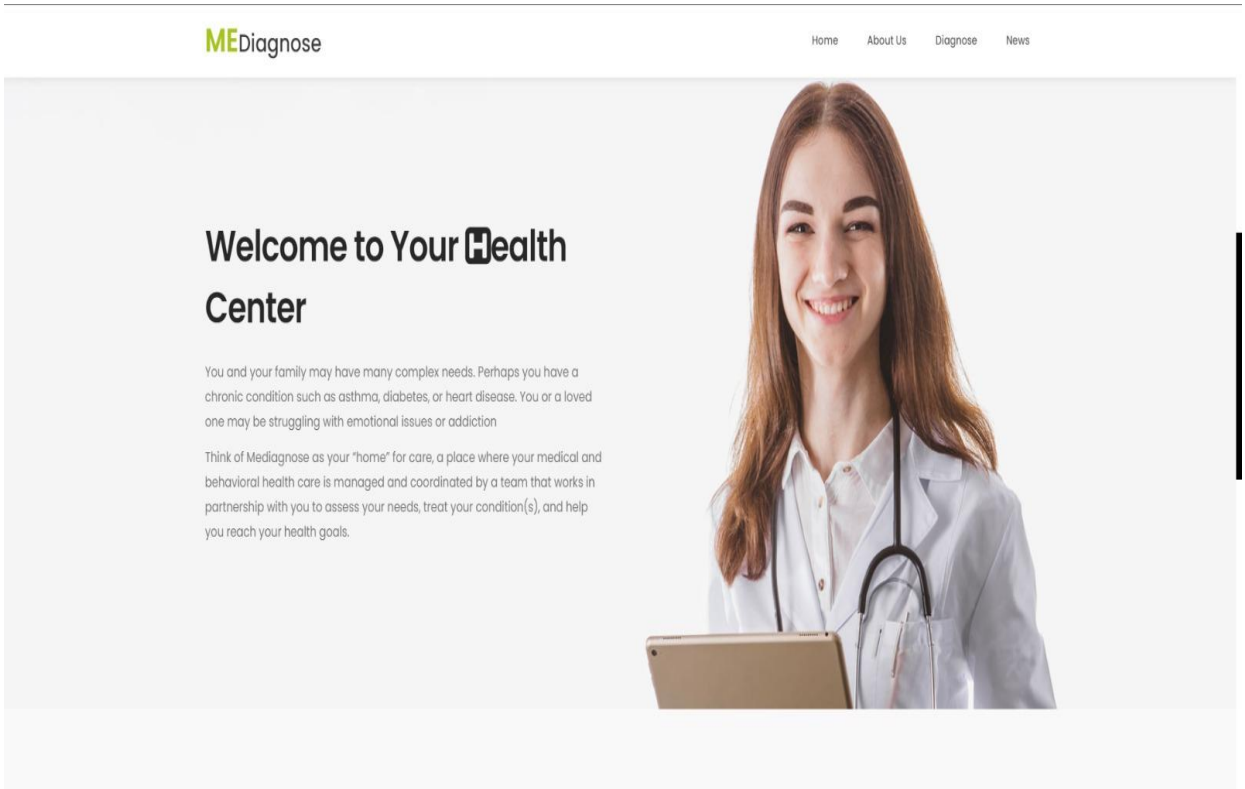


Figure 20 *MeDiagnose About us*

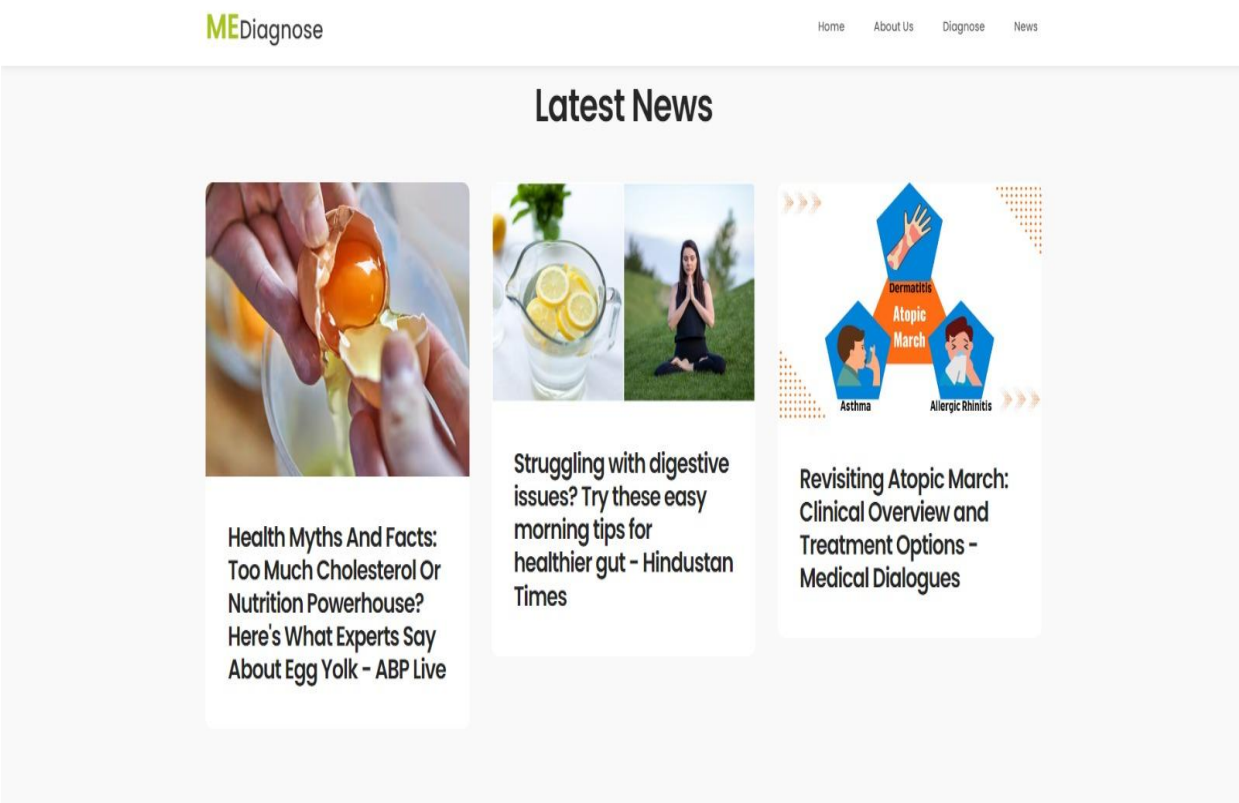


Figure 21 *MeDiagnose News*

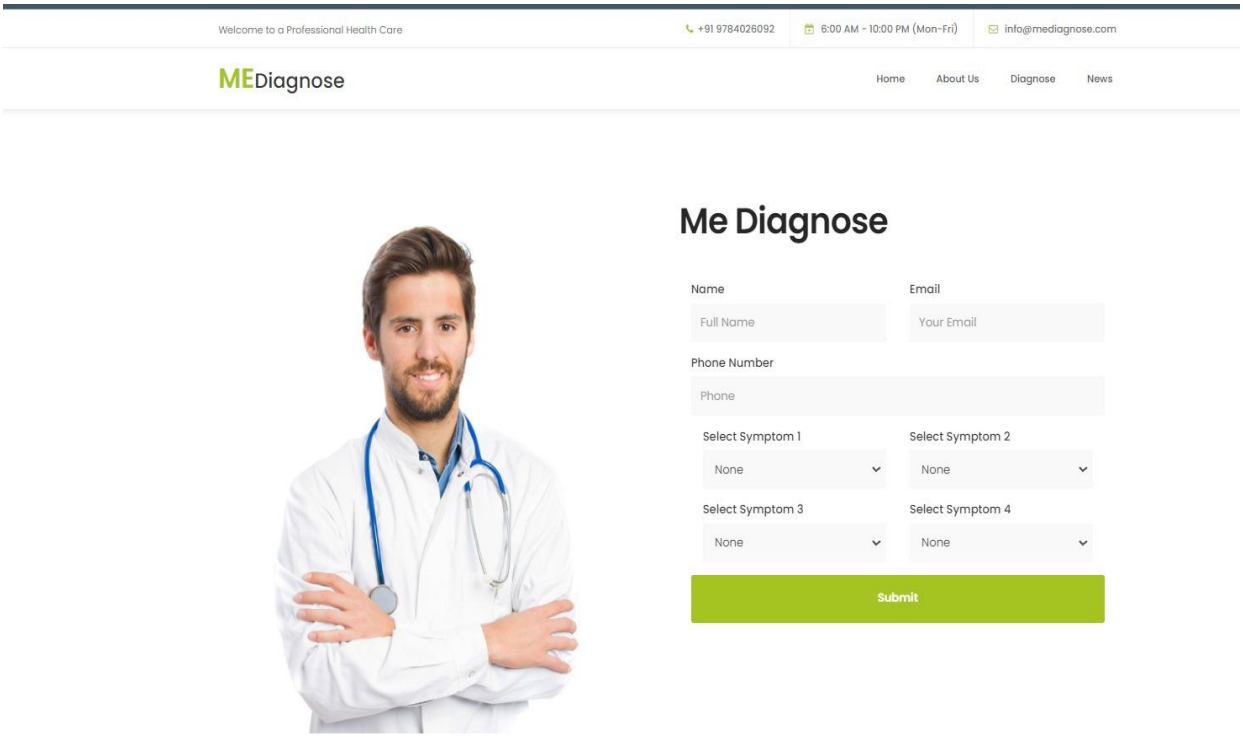


Figure 22 *MeDiagnose Diagnose Page*

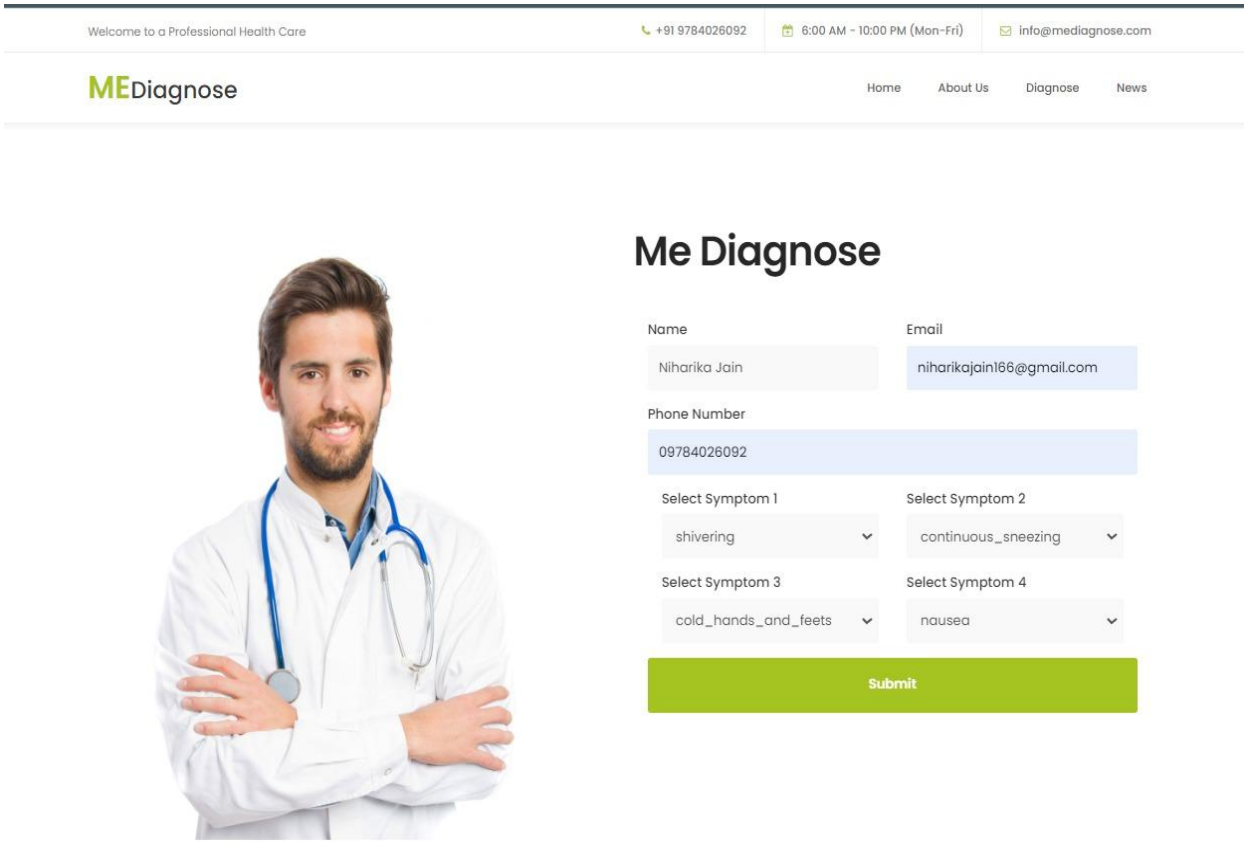


Figure 23 *MeDiagnose Diagnose Form*

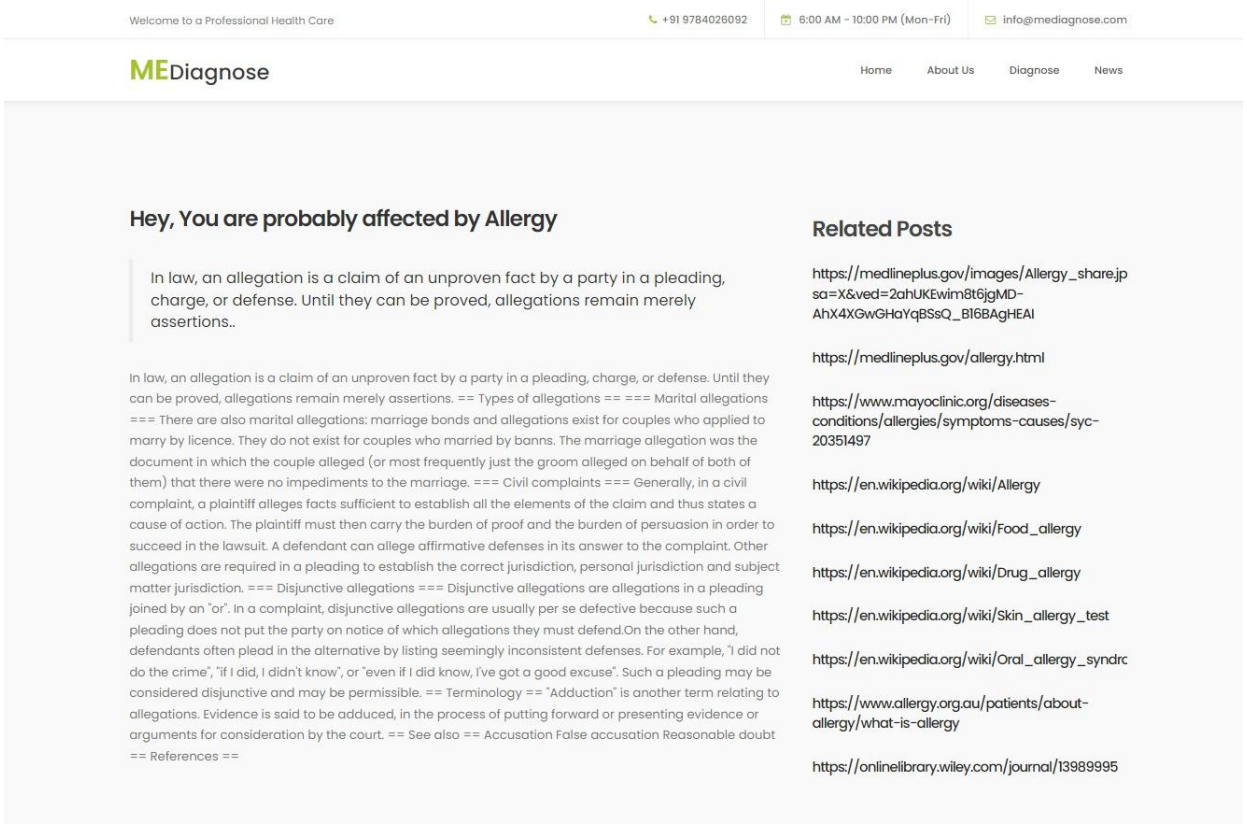


Figure 24 *MeDiagnose Diagnose Result*

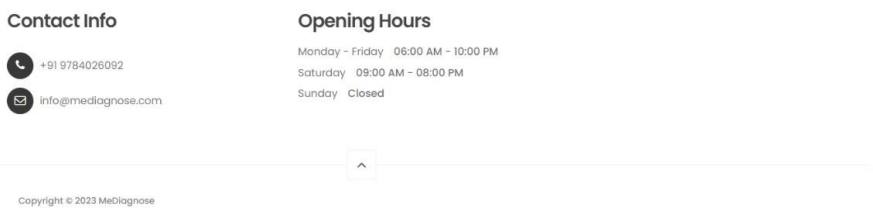


Figure 25 *MeDiagnose Footer*

Chapter 6
Conclusion

In this project, neural network and SVM machine learning methods are given to solve the automatic disease diagnosis problem only based on symptoms. In our methods, each symptom is a feature. The methods work in three layers, which are main disease category identification, subclass disease type identification, and specific disease identification. The methods are suitable for the diagnosis of common diseases and disease triage for specialized diseases. The availability in practice is proved and analyzed in the experiments of this paper. In addition, future research is also required to investigate automatic symptom extraction and discuss the maximum number size of symptoms.

The Global Healthcare Predictive Analytics Market Size was valued at USD 9.5 Billion in 2021 and is estimated to reach the market value of USD 87.5 Billion by 2030, growing at a CAGR of 28.2% from 2022 to 2030. The growing prevalence of chronic disorders and the increasing expenditures on these diseases has been troublesome for people with low and middle income. This factor has exploded the healthcare predictive analytics market revenue throughout the forecast timeframe from 2022 to 2030. According to the Centers for Disease Control and Prevention (CDC), people with chronic and mental health conditions account for 90% of the United States' USD 3.8 trillion in yearly healthcare expenditure. As a result, the potential cost savings from lowering chronic disorder treatment are substantial. For instance, more than 877,500 Americans die each year from heart disease or stroke, accounting for one-third of all deaths. These diseases also have an economic impact, costing the US healthcare system USD 216 billion per year.

Chapter 7

References

[1] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," *International Journal of Computer Application*, vol. 19, no. 6, pp. 6–12, 2011.

[2] J. Shiraishi, Q. Li, D. Appelbaum, and K. Doi, "Computer-aided diagnosis and artificial intelligence in clinical imaging," *Seminars in Nuclear Medicine*, vol. 41, no. 6, pp. 449–462, 2011.

[3] W. Gao, H. Liang, W. Zhong, and J. Lv, "Differential diagnosis of neonatal necrotizing enter colitis based on machine learning," *China Digital Medicine*, vol. 14, no. 3, pp. 50–52, 2019.

[4] K. Matjaz, K. Igor, G. Ciril, K. Katarina, and F. Jure, "Analyzing and improving the diagnosis of ischaemic heart disease with machine learning," *Artificial Intelligence in Medicine*, vol. 16, no. 1, pp. 25–50, 1999.

[5] K. Igor, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001.

[6] M. Manish, D. Damini, S. Daniel et al., "Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicenter prospective registry analysis," *European Heart Journal*, vol. 38, no. 7, pp. 500–507, 2017.

[7] D. lin, A. V. Vasilakos, Y. Tang, and Y. Yao, "Neural networks for computer-aided diagnosis in medicine: a review," *Neurocomputing*, vol. 216, no. 5, pp. 700–708, 2016.

[8] J. Wang, C. Su, and T. Ren, "Design and realization of the intelligent hospital guide," *Journal of Medical Informatics*, vol. 39, no. 8, pp. 29–32, 2018.

[9] X. Luo, Y. Chang, and J. Yang, "An automatic disease diagnosis method based on big medical data," in *Proceedings Of the 2015 International Conference On Information Science And Security*, pp. 252–254, IEEE, Seoul, Korea(South), December 2015.

[10] R. Xu, *"The Research of Intelligence Auxiliary Disease Guidance Based on Text Mining Technology," Master Dissertation*, Beijing University of Posts and Telecommunications, Beijing, China, 2015.

[11] C. Li, *"Research and application on intelligent disease guisdance and medical question answering method"*, Dalian University of Technology, Dalian, China, 2016.

[12] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.

[13] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.

[14] A. Onan, "Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering," IEEE Access, vol. 7, pp. 145614–145633, 2019.

[15] Z. Zhou, Machine Learning, Tsinghua University Press, Beijing, China, 2016.

[16] Z. Liu, D. Li, Q. Qin, and W. Shi, "An anlytical overview of methods for multi-category support vector machine," Computer Engineering and Applications, vol. 46, no. 7, pp. 10–13+65, 2004.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, no. 85, pp. 2825–2830, 2011.

[18] M. Maniruzzaman, M.J. Rahman, B. Ahammed, M.M. Abedin, Classification and prediction of diabetes disease using machine learning paradigm, Health Information Science and Systems 8(1), 7 (2020)