**HEALTHCARE SYSTEM**

**A PROJECT REPORT**

**Submitted partial fulfilment of the**

**requirement for the award of the degree**

**of**

**MASTER OF COMPUTER APPLICATIONS(MCA)**

**By**

**Tanya Bodhwani**

**23FS20MCA0074**

**A close up of a logo

AI-generated content may be incorrect.**

COMPUTER APPLICATIONS

MANIPAL UNIVERSITY JAIPUR

JAIPUR-303007

RAJASTHAN, INDIA

MAY 2025

**DEPARTMENT OF COMPUTER APPLICATIONS**

MANIPAL UNIVERSITY JAIPUR, JAIPUR – 303 007 (RAJASTHAN), INDIA

Date: 16/05/2025

# CERTIFICATE

This is to certify that the project titled **HEALTHCARE SYSTEM** is a record of the bonafide work done by **TANYA BODHWANI** (23FS20MCA00074) submitted in partial fulfilment of the requirements for the award of the Degree of Master of Computer Applications (MCA) in **(Discipline) of Manipal University Jaipur, during the academic year 2018-19.**

## Dr. Vaibhav Bhatnagar

*Project Guide, Dept of Computer Applications*

*Manipal University Jaipur*

## Dr. Shilpa Sharma

*HOD, Dept of Computer Applications*

*Manipal University Jaipur*

(*On company letterhead)*

Date: 16/5/2025

# CERTIFICATE

This is to certify that the project entitled **HEALTHCARE SYSTEM** was carried out by **TANYA BODHWANI**(23FS20MCA00074) at **Manipal University, JAIPUR** under my guidance during **January 2025** to **May** **2025**.

## Dr. Vaibhav Bhatnagar

Project Guide,

Manipal University Jaipur

Acknowledgement

I am very grateful to all those who helped me during the development of this healthcare prediction system project. I am mainly grateful to my mentor/supervisor Dr. Vaibhav Bhatnagar for continually helping, encouraging, and giving me feedback, all of which made this project a success. What they shared allowed me to improve the technology and analytical features of the system. I am also grateful to my institution and to my teachers for helping me with the resources and learning setting needed for this project. Learning the relevant skills and information through the coursework made it possible to complete this project. I am grateful for my peers and friends for sharing their ideas, discussing questions, and supporting me during the project’s different phases. I also want to thank my family for always being there for me and pushing me forward in this process. Through this project, I have gained a lot and feel proud of the improvement my system offers in related to predictive analytics for healthcare.

.

**Tanya Bodhwani**

# Abstract

Health care professionals will be able to forecast diseases more effectively with the help of machine learning using SVM, linear regression and decision trees. The main goals are to improve learning in healthcare with machines, assess how accurate algorithms are, and develop a user-friendly system for predicting chronic diseases remotely. The strategy employs Python tools alongside application development. In conclusion, researchers found that using SVM, linear regression and decision trees increased the accuracy when predicting cardiac problems or cancer. It is clear that the Application is effective at promoting health through its excellent precision numbers.

See how effectively such a medical app works to control and manage diseases among patients. It has been determined that this application is useful for long-term care, forms part of healthcare plans, and that AI is an important component in the industry. There are also advancements in healthcare technology that help strengthen disease prediction accuracy and guide people towards making knowledgeable decisions, all from a user-friendly online platform. The adoption of such an active model in healthcare establishments is one of the best ways to enhance health results for patients.

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No** | **Figure Title** | **Page No** |
| 1 | Process diagram of machine learning model training | 18 |
| 2 | Site map of the Website | 19 |
| 3 | DFD Level 0 & Level 1 | 20 |
| 4 | Index Page1 | 23 |
| 5 | Code of index.html | 23 |
| 6 | Home page of the application1 | 24 |
| 7 | Home page of the application2 | 24 |
| 8 | Code of login.html | 25 |
| 9 | Login page | 25 |
| 10 | Signup.html code | 26 |
| 11 | Signup page | 26 |
| 12 | Cancer prediction form | 27 |
| 13 | Cancer prediction form page | 27 |
| 14 | Heart stroke prediction form.html | 28 |
| 15 | Heart stroke prediction form page | 28 |
| 16 | Heart prediction result | 29 |
| 17 | Cancer prediction result.html | 29 |
| 18 | Cancer prediction result page | 30 |
| 19 | User Profile Page | 30 |
| 20 | Admin panel of the Website | 31 |
| 21 | Cancer prediction model accuracy | 31 |
| 22 | Heart stroke model accuracy | 31 |

Contents

[Certificate I](#_Toc164768493)

[Certificate of Originality II](#_Toc164768494)

[Acknowledgement III](#_Toc164768495)

[Abstract IV](#_Toc164768496)

[Chapter 1 – Introduction 2](#_Toc164768497)

[1.1. Background 2](#_Toc164768498)

[1.2. Problem statement 2](#_Toc164768499)

[1.3. Aim 3](#_Toc164768500)

[1.4. Objectives 3](#_Toc164768501)

[1.5. Research questions 3](#_Toc164768502)

[1.6. Solution 4](#_Toc164768503)

[1.7. Overview of the methodology 4](#_Toc164768504)

[1.8. The organization of the report 4](#_Toc164768505)

[1.9 Summary 4](#_Toc164768506)

[Chapter 2 – Literature Review 6](#_Toc164768507)

[2.1. Introduction 6](#_Toc164768508)

[2.2. Healthcare applications and prediction of chronic diseases 6](#_Toc164768509)

[2.3. Healthcare applications and prediction of heart diseases 7](#_Toc164768510)

[2.4. Healthcare applications and use of Support Vector Machine (SVM) 8](#_Toc164768511)

[2.5. Healthcare and use of Logistic Regression (L.R.) 8](#_Toc164768512)

[2.6. Healthcare application and use of Decision Tree algorithm 9](#_Toc164768513)

[2.7. Healthcare applications in cancer prediction 9](#_Toc164768514)

[2.8. Compare and contrast themes. 10](#_Toc164768515)

[2.9. Limitations of literature 10](#_Toc164768516)

[2.10. Conceptual framework 11](#_Toc164768517)

[2.11. Summary of sections 11](#_Toc164768518)

[Chapter 3 – Methodology 12](#_Toc164768519)

[3.1. Introduction 12](#_Toc164768520)

[3.2. The rationale for the research approach 12](#_Toc164768521)

[3.3. Research sample and data sources 12](#_Toc164768522)

[3.4. Data Collection Method 13](#_Toc164768523)

[3.5. Data analysis method 14](#_Toc164768524)

[3.6. Model Implementation 15](#_Toc164768525)

[3.7. Results 16](#_Toc164768526)

[3.8. Exports to web application 16](#_Toc164768527)

[3.9 Summary 17](#_Toc164768528)

[Chapter 4 – Implementation 18](#_Toc164768529)

[Introduction 18](#_Toc164768530)

[Methodology 18](#_Toc164768531)

[Machine Learning Model training steps 18](#_Toc164768532)

[Development and Integration of Websites 19](#_Toc164768533)

[Key Features 20](#_Toc164768534)

[Pages of Website (Site map) 21](#_Toc164768535)

[Results and Outcomes 29](#_Toc164768536)

[Conclusion 29](#_Toc164768537)

[Summary 29](#_Toc164768538)

[Chapter 5 – Discussion and Analysis 31](#_Toc164768539)

[Summary 33](#_Toc164768540)

[Chapter 6 – Conclusion and Future Scope 35](#_Toc164768541)

[Conclusion 35](#_Toc164768542)

Future Scope………………………………………………………………………………36

[References 37](#_Toc164768543)

[Appendix 44](#_Toc164768544)

[Survey questions 44](#_Toc164768545)

# 

# Chapter 1 – Introduction

## Background

Many people globally experience a range of illnesses caused by both their habits and the environment. Knowing what diseases are likely to spread helps keep people healthy. Still, healthcare experts and others find it very difficult to predict any illness with completely accurate results. This work attempts to discuss various applications that implement different techniques and algorithms to solve the difficulty. As a result, the apps aim to build software that can effectively predict illnesses. Thanks to the vast data available in healthcare and medical sciences, it will be uncomplicated to pick out the fundamental info to be used for prediction applications. The health department’s data will be used to perform data mining, where it becomes possible to see patterns in a vast quantity of disease data. The tool uses typical ways to predict disease by considering the symptoms reported by patients (Dahiwade et al. 2019). Most of its work is focused on diseases such as cancer and heart attack. SVM, decision tree, and linear regression will be applied by the project to predict all types of diseases completely. In order to predict a disease, one must use the symptoms from the disease dataset, while for typical disease prediction, analyzing the lifestyle and considered information for accurate results is important. A disease prediction using Linear regression has an accuracy of 95%, SVM has 88%, and the Decision tree has 86.7% accuracy, according to T. et al. in 2013. Healthcare professionals benefit from using and recognizing machine learning applications. By using both powerful computation and big data, along with machine learning algorithms, there will be more benefits for improving healthcare. In this context, we use supervised learning with support vector machines, decision trees, and logistic regression as algorithms to predict labelled data. These are the models that are helpful in finding anomalies and fraud (Alanazi 2022).

## Problem statement

Doctors are now encountering greater difficulty in predicting illnesses. Unhealthy habits and the conditions around us are causing people to face many different health issues. Early recognition of illness can bring down the seriousness of many diseases. It is becoming more difficult for doctors and nurses to correctly predict diseases simply by hand. Outlooks regarding heart disease and cancer are very hard to foresee. As a result, there is a push to make an application that detects diseases so doctors can treat their patients on schedule.

## Aim

This research intends to build an app that can correctly predict heart diseases and cancer. The work also demonstrates that SVM, linear regression, and decision tree models can help achieve the highest level of accuracy in prediction. This article explains how an application is built and the information gathering methods that must be used to do so.

## Objectives

SVM, decision trees, and linear regression are the main algorithms being emphasized for understanding and using throughout this work. Several of the primary aims of the report are –

* Discovering how machine learning is being used in healthcare development.
* SVM, linear regression, and decision trees are expected to assist the health industry in growing new technologies in medicine.
* Machine learning will help boost different health-related strategies and methods.
* What accuracy do the algorithms have when working on the survey?
* A platform where end users can estimate the chances of developing chronic diseases, without having to see the patient in person.
* To make sure the prediction application is accurate, we look into health issues directly.
* Machine learning methods will make the prediction more accurate.

## Research questions

1. How important is healthcare in helping public and community health?
2. Is it possible that the use of this healthcare app helped the user lower the number of hospital stays and keep their disease under control?
3. Does applying healthcare through the app improve both treatment and health outcomes for its users?

## Solution

In the present study, some important algorithms and techniques for healthcare application development are highlighted. It will show how health care uses machine learning technology. Using machine learning in healthcare helps make it easier for people to be diagnosed and treated early. Approaches based on imaging help bring down the cost of patient treatment and also lower the chances of late diagnosis (Kaul, Raju and Tripathy, 2021). Research revealed that by using machine learning techniques, the health sector can achieve better predictions and major changes that lead to easier daily living. Thanks to machine learning, the data used to decide on heart disease diagnoses is often more optimal (Javaid et al., 2022).

## Overview of the methodology

This research makes use of a mix of common assembly techniques. It uses several techniques to gather data from earlier reports and surveys in the most efficient way. Python, Jupyter, pickle, and Django are applied in application development to boost disease prediction and give complete details about common diseases. In order to collect data for health prediction applications, a mixed approach that combines machine learning and supervised learning methods is used (Sodhi, Awasthi and Sharma, 2019).

## The organization of the report

The report includes the Healthcare applications and prediction of chronic diseases as well as Healthcare applications and prediction of heart diseases. Among the types of points in the report are Healthcare applications and use of Support Vector Machine (SVM), Healthcare and use of Logistic Regression (L.R.), Healthcare application and use of Decision Tree algorithm, the limitation and delimitation in the research literature, research methodology and the plan for the practical work, data analysis and the report’s main conclusions, testing, the way the methods were applied, the evaluation of results, and the recommendation (G et al., 2022).

## 1.9 Summary

In this chapter, a summary of the research being done in healthcare for disease forecasting is given. The background discusses the worldwide challenges to health that come from lifestyle and environmental issues. With logistic regression, support vector machine SVM, and decision trees, targets can forecast diseases, mainly cancer and heart attacks. It points out that predicting illnesses manually is a challenge doctors encounter. Instead, the focus of this study is on building an application to support early disease detection. Examples include getting to know how ML is used in healthcare, bettering health strategies, and looking at how effective prediction algorithms are. The main purpose of the research questions is to see how this application contributes to community health and evaluate the outcomes achieved so far. Implementing the solution will help machine learning boost healthcare contacts and make treatments less costly. This methodology combines Python technologies to collect and review data.

# Chapter 2 – Literature Review

## Introduction

The proposed literature review is mainly focused on handling the three research questions that are part of the healthcare application proposal. To start, the research focus has been predetermined as “Healthcare App for Disease Prediction With a Doctor’s Database Providing Probable and Frequent Disease Information”. In order to do this, we take the previously explored research questions and they become the basis for justifying this study’s scope and doing a literature review. The 3 questions and the algorithms (SVM, L.R., and decision tree) under study will define the main scope of the research literature, applied to healthcare applications of mobile apps. The research incorporated in this thesis will mainly cover the present-day applications tied to the project’s topic and also consider proposed algorithms. Besides, the search methodology and the process of extracting data from the studies reviewed various previous works that added the keywords “SVM and Healthcare” and “Healthcare apps”, along with “L.R.” and “Healthcare applications” to their titles. Data was sourced by looking through the Google Scholar website, ScienceDirect portal, and filtering for results that referred only to 2018 through 2023..

## Healthcare applications and prediction of chronic diseases

Considerable work has been done on how healthcare applications, such as the one that has been suggested, use various models for predicting healthcare challenges. For example, Bansal, Ahirwar, and Shukla (2018) have studied how different healthcare applications predict various diseases using various contexts. Also, the study explained that when using the IoT and the decision tree algorithm, healthcare applications can predict chronic diseases. Communication-based approaches, in addition to using other algorithms such as the decision tree, are used to predict chronic diseases including cancer and bone marrow illnesses. A similar study was performed by (Tumpa and Dey, 2022), which similarly supported the findings from before. This research confirmed that machine learning as a whole works effectively for chronic disease prediction. Additionally, the author underlined that medical data is being implemented together with machine learning in healthcare applications to predict certain diseases. How to use machine learning in healthcare applications and how effective it is were the main points studied here. In addition, (Alharthi, 2018) discussed how predictive health analytics help Saudi Arabia’s healthcare and mentioned that even though there are some operational difficulties in disease prediction, there are some upsides to using prediction analysis. A major issue pointed out by this research involves successfully rolling out human resources, technology, and information systems required for disease-related predictive analytics that use machine learning in healthcare

## Healthcare applications and prediction of heart diseases

In addition, research has been conducted to identify methods for predicting heart diseases to lower chances of certain diseases in healthcare settings and similar applications. For instance, Yahaya, David Oye, and Joshua Garba’s study (2020) point out that there have been several versions of these algorithms and programs created for the detection as well as prediction of heart disease in patients. Some of these things are the ANN (Artificial Neural Network), the D.T. (Decision Tree), and the Naive Bayes algorithms. Each of these algorithms has been applied to predict heart health in addition to other areas of healthcare. Furthermore, the study discussed issues involving heart disease technologies and observed significant flaws in predicting the diseases. In addition, the study discussed broad topics such as machine learning and data mining.

A further study (Ahsan and Siddique, 2022) backed this work by mentioning the use of technology for preventing heart disease. An ECG machine is used to collect data and information from patients at any given time to confirm or exclude the chance of heart disease. Also, the main points of the study came up in the middle sections. It also included information on the situations that can cause data imbalances, as well as the accuracy and validity of the data, which are among the main problems that occur when ECG machines are used to predict heart disease. Besides, according to the research, using data mining and machine learning helps to predict complications more accurately and make fewer mistakes when dealing with heart diseases.

## Healthcare applications and use of Support Vector Machine (SVM)

SVM is used to forecast various healthcare procedures. For instance, previous works by (Bansal, Ahirwar, and Shukla, 2018) studied how accurate data predictive analytics can be when SVM is teamed with N.B. and KNN algorithms. It was found that, using devices and apps with SVM technology, patients succeeded in being diagnosed for the specific diseases with about 85.19% effectiveness. This study also points out that it has been possible to predict heart diseases as well as other major diseases that often affect specific regions.

Besides this, other studies aiming to confirm this report have also researched the use of SVM in detecting or preventing diseases in advance. (Battineni et al., 2020) revealed that SVM, as well as models and algorithms based on KNN, are useful for predicting diseases related to the lungs and breathing, namely those of COPD. Although other methods, for example the N.B. and the Decision Tree, make use of other capabilities, the SVM was the best in predicting COPD cases that have developed in recent years. This study also makes use of R.F. and Logistic Regression models and algorithms.

## Healthcare and use of Logistic Regression (L.R.)

L.R., or logistics regression, is still another tool for examining how healthcare applications use these models to predict diseases. This study by (Islam et al., 2018) considered the ability of L.R. algorithms to predict outcomes. It made sure that the estimated effectiveness of the algorithm for predicting the disease is about 89%. L.R. is also used when it comes to finding and monitoring healthcare-related diseases. The algorithm lay out in the study has shown that it can successfully predict coronary heart diseases 89% of the time, providing solid support for what the L.R. application and algorithm are capable of. Besides, the analysis highlighted certain aspects of the L.R., for example, challenges dealing with memory and the processing of data storage units (Zabor et al., 2022).

Besides, the research undertaken by (Straw and Wu, 2022) also included looking at the current aspects of whether using L.Rs helps rule out the chances of certain healthcare and disease outcomes. As a result of this similar study and the study by Alanazi, it was shown that L.R. helped to predict women’s long-term chronic diseases in 71.5% of patients, further confirming the previous study. Moreover, separate research done by (M. and Chattu, 2021) has also established the use of the L.R. model to predict the chances of diseases involving the lungs and liver.

## Healthcare application and use of Decision Tree algorithm

Another use of decision tree in healthcare is to help make predictions more accurate when applied within healthcare applications that have rigid integration. Therefore, the research carried out by (Mienye, Sun, and Wang, 2019) revealed that ID 3 has been commonly used among different decision tree models for healthcare-related disease prediction reasons (Bae, 2014). This paper reports that when the Decision Tree model and algorithm were used, about 80% success was accomplished in this context (Mo et al., 2023). Other studies have also been conducted that focus on this way of using models. However, this study mainly states that the ID3 algorithm treats current medical information in an interface or application to accurately predict potential illnesses and results. Besides, another research (Kumar, 2022) used the Decision Tree approach and detected that the success rate for predicting coronary heart diseases was about 96.5% when Decision tree methods were involved. The same similarity has been reported in the studies of (Javeed et al., 2022) and (Khan et al., 2019).

## Healthcare applications in cancer prediction

Machine learning is used in diverse situations to help predict cancer. Zhang, Shi, and Wang (2023) pointed out that using machine learning in screening and other healthcare devices can help diagnose as well as predict cases of skin, blood, and breast-related cancers (Shaikh and Rao, 2021). Brain cancer and lung cancer can also be anticipated ahead of time with the aid of machine learning and Artificial intelligence used in several healthcare systems. It is worth noting that other studies (by Rafique, Islam, and Kazi in 2021 and Fatima et al. in 2020) also confirm this, suggesting that with machine learning, specific types of genetic abnormalities can be successfully predicted and identified before certain cancers are formed in a particular setting. Additionally, another study conducted by (Li et al., 2021) predicted that machine learning can accurately determine whether or not breast cancer is present in older women patients. A large number of these studies have had about an 80% rate of success (Rong et al., 2020).

## Compare and contrast themes.

Machine learning, the decision tree, and other two algorithms have all been the main subjects explored in literature on the topic. It could be compared to other examples by using the other theme discovered in this context, and it was the use of machine learning to forecast diseases related to the heart and cancers. Acording to the study, machine learning plays a significant role and is most effective at detecting coronary heart diseases. In such cases, radiologists are slightly less accurate when trying to predict cancer. Though the predictions are not very different, some areas of application have success in both areas. Besides, the use of the L.R., SVM, and Decision Tree algorithms is compared in this case study. An example is that an ordinary decision tree is only used for lung disease prediction, but L.R. and SVM are more suitable for predicting diseases expressed in inflammatory and in the lungs.

## Limitations of literature

Most of the findings revealed that the literature has significant limitations. There were three key problems that affected the literature. In the beginning, not as many examinations center on the two algorithms that were mainly noticed. There were many more studies on machine learning and similar fields compared to the ones on Logistics regression and Decision Tree predictions of illnesses. Thus, it is evident that there were more research opportunities in this area. One more problem with this research was that many of the data sets were not recent. Most of the literature looked at data fitting for that context, but the results obtained were old and unsuitable for current use. This also turned out to be a serious problem and a constraint in the course of the research.

## Conceptual framework

Once the ideas from the studies were combined, it was found that the chance of accurately predicting diseases with healthcare technology was likely to be smaller. In addition, the prediction of diseases involving heart and cancer has a high chance of accuracy, ranging from 80 to 90%, so essential parts of the predictions in predictive healthcare can be addressed. Within the context of prediction, machine learning has also been studied by looking at the identified literature. It is still possible that success rates in healthcare will vary, despite the use of numerous predictive analytics.

## Summary of sections

It was then noted that the authors put forward important directions for creating a healthcare application that could help in the prediction of heart-related and cancer-related illnesses. It is difficult to accurately predict both of them using health applications, so the suggested model applied the versions of Decision Tree, L.R., and SVM algorithms. According to the literature, all three algorithms were very successful, with estimates showing that L.R. prediction and Decision Tree prediction can be around 80% and 71%, and SVM around 85%. All the literature presented demonstrated that skin cancers, chronic obstructive pulmonary diseases, the range of inflammatory diseases, breast cancers, and lung-related diseases were successfully treated with these algorithms.

# Chapter 3 – Methodology

## Introduction

By conducting the research, it is aimed to produce a model or useful healthcare app capable of helping predict, manage, and present key details about heart and cancer diseases. The goal of the proposed model or web application is to improve health for individuals through advanced technology and current expertise in managing different diseases. Chapter Five, “Methodology”, details the research methods, procedures, and others aspects that are relevant. Different sections explaining the methodology and data of the research are included in the next chapter. Sections one and two of the chapter discuss the research approach and offer details on the source of data as well as the size of the research sample. Sources of data, ways they were analyzed, trustworthiness concerns, and the research’s boundaries were all described in this chapter.

## The rationale for the research approach

We have used mixed methods for conducting the assessment. The reason for choosing this approach is that it allows the researcher to gain and analyze both types of data. Since research will be ineffective and inefficient if it uses just one kind of data, the mixed method approach is essential here. The selection of the approach was influenced by an increase in the research’s credibility thanks to different data collection techniques used (George, 2021).

## Research sample and data sources

Samples for this research study have been drawn from medical laboratories, sensor equipment, electronic health data, and various other sources that offer good health data for public health. The sample includes 500 people’s data. Through the community health centres and their online networks, groups of female participants will be invited to join. To make the findings useful, the study involves adults of all age groups and with diverse backgrounds. The participants selected for the study must agree to taking part and giving their informed permission for privacy.

No one’s name or private details will be shown while collecting the data. Instead, the participants will use codes or numbers to show their answers. For this reason, no one can figure out who wrote those messages. Before any further questions are asked, permission will be gotten to ensure the safety of all participants. They will be told the target of the study and the steps they should take before getting started. They are allowed not to answer those questions if they find them uncomfortable. The research will ensure that no problems arise for the volunteers (White, 2020).

All possible means will be used to ensure the rights of each individual are safeguarded. Similarly, their rights will remain intact, since they will not be revealed in any way. So, they must ensure that the participants’ rights are considered and they are dealt with respect. Thus, it will protect people by ensuring their consent and keeping their identity and privacy safe and intact, whether collecting or using the information obtained during or after the event. If a person no longer needs them, the required laws and regulations will promptly ensure that their information is erased or destroyed (2019).

## Data Collection Method

For this study, data was collected using both qualitative and quantitative methods. The biggest benefit of a mixed data-gathering strategy is that you can gather, examine, and analyze qualitative as well as numeric data at the same time. Gathering data and interpretation using multiple means may improve their validity, credibility, and reliability (Kuhn, 2023). With this approach, people can gather details about public health and data through both interviews and surveys.

The survey will be conducted online to ask about their needs and present health problems. There are several questions about heart and cancer problems in the survey used to collect information from patients (Aguinis, Hill and Bailey, 2019). The survey has been designed so that the surveyors can understand details about a person’s health and healthcare in both numbers and in words. Many kinds of sensors will also record and monitor data about individual patients’ health. The information will also come from electronic medical files and recent advancements in medical science. For the survey, information and data about the health of 500 participants will be collected. This model will guide in collection of Heart disease information, thanks to the medication and heart and cancer disease detection features it will include. Data is collected on participants aged 18 and over so that results or information about their health conditions will be meaningful.

## Data analysis method

It is a process that evaluates raw data and ensures its cleanup. The data collection process may include both errors and missing items, so you first clean the data during preprocessing and then use it for additional analysis. Highly accurate, error-free, and clean data ensures good model web application results, however, errors in the data can affect its efficiency and success (Tiwari, 2021).

Collection from different sources has completed the gathering of both types of data, and adjustments were made to the survey questionnaire so that sound data could be used for the study. It’s necessary to clean data, as the information gathered can include numerous errors that would not be useful for the study and could weaken the results and accuracy of the proposed study (Caracelli and Greene, 1993).

Python programming will be used to make the model. Some of them are Jupyter, Django, and the pickle tools. I will use Python for the model, as it offers many useful tools for handling the data that was acquired. There are many functions and methods in the Python library to handle issues with errors, missing values, outliers, and duplicate data.

You can use the fill() and drop() functions in Panda’s Python to work with any lack of data you found in the health-related data obtained. Identifying and dealing with abnormal cases is another strategy for handling data. Applying the RobustScaler class from the Python, scikit-learn library to work with outliers. To eradicate the duplication of values, an option is to use drop-duplicates()(Surya Gunawan et al., 2020).

Feature selection focuses on keeping only relevant inputs and clearing out the noise in the data. Feature selection involves picking out the appropriate features that match the requirements of the learning model. By using feature selection, the research study will identify and choose the features that will create the best model. For developing models and applications in healthcare, supervised feature selection plays a valuable role in the process (Obi Tayo, 2023).

When there are labels for the data, the supervised feature selection method is used to discover the essential features to boost the efficiency of regression and classification models, and the proposed model or web application depends on SVM, L.R., and decision tree (Calzon, 2023).

When using the wrapping method, the search technique is applied to feature selection by making, evaluating, and comparing several combinations. Here, algorithms are taught by using a combination of a few features in a cyclical manner. Some other methods from the wrapper approach that you can use are backward elimination, forward selection, recursive feature elimination, and exhaustive feature selection. In the case of the filter method, the statistic measures provide the support for choosing the best features. When using this technique, feature selection occurs in the beginning as it is not influenced by the learning algorithm used. With the aid of filter methods, the model will get rid of columns or features that are redundant or not necessary. In this method, you might choose to work with Fisher’s score, chi-square test, missing value ratio, or information gain techniques. By using the embedded method, we combine the features of both the wrapper and filter approaches by looking into how different features affect each other. The process involves reviewing and selecting the main feature that can help a specific training run. Techniques included in it are regularisation and the significance of random forest (Gupta, 2020).

## Model Implementation

The support vector machine algorithm will be implemented to assist in developing the model or web application meant to suggest and supply medications for heart and cancer illnesses. It is a supervised ML algorithm that applies a classification algorithm to sort out two or more group problems. It employs the kernel trick to process nonlinear data and is suitable for a model that is meant to detect or categorize values or inputs (Orrù et al., 2020). It is possible to use an SVM algorithm-based model with both categorical and constant variables. An SVM algorithm will form a hyperplane in a multidimensional space to separate several sets of data (Sharma and Yadav, 2020).

One of the main methods for the development of the health model is linear regression. One of the main benefits of using linear regression is that the results are not hard to understand. Quantifying the link between various independent variables and dependent ones is one of the uses of linear regression (Silhavy, Silhavy and Prokopova, 2017). In linear regression, you can use simple linear regression to find out how a constant value affects another constant value. It uses one variable that is under the researcher’s control and tries to connect it to another variable through the best possible linear relationship (Singh and Kumar, 2020).

Many scientists prefer the decision tree algorithm that makes use of a teacher. A decision tree algorithm separates the data into several subpopulations based on the major difference between the input variables (Charbuty and Abdulazeez, 2021). The model included in the research can be carried out using either of these two types of decision tree algorithms: one can find both continuous variable decision trees and categorical variable decision trees. The decision tree algorithms bring many helpful features to the health-based model and the webpage. It is capable of using the numerical data and the categorical data required for the analysis (Maji and Arora, 2018).

## Results

It is considered correct classification when the score from the support vector machine algorithm is 88%. This is a good sign, because it seems the outcomes are being produced with high reliability. Any linear regression model with a success rate of 95% is viewed as reliable and OK to use. Its categorization rate/outcome should be 86.7% for the decision tree to be right and practical. The model can forecast diseases accurately with decision tree algorithm (86.7%), by using linear regression (95%), and by employing a supporting vector machine (88%).

## Exports to web application

Predicting and providing treatment for heart and cancer diseases will be done using the machine learning model packaged with pickle and integrated into the web application (Jain, Debnath and Jain, 2022). All in all, the model will be trained by using the health-related data that has been received. When it is done, the exported information is added to the web application suggested for the model. The trained model will allow the users to get predictions about heart and cancer diseases after entering their health-related data into the program. In conclusion, the model planned for predicting heart and cancer illnesses among members of the public will be delivered through the web application, allowing people to get the correct predictions and required treatment.

## 3.9 Summary

This chapter offers assistance in making an application to predict heart and cancer diseases. In this study, both qualitative and quantitative information are used. This strategy makes sure that the results of research can be used effectively. More than 500 research samples were collected from medical studies, sensors and health records. Participants’ information is no longer seen but is protected through the use of codes instead. Data collection will be done by having persons take part in an online questionnaire and use sensor equipment. It is also part of the process to make the information accurate and reliable through data cleaning. Jupyter and Django can be used in Python to help with building models. Part of disease prediction relies on algorithms such as support vector machine, logistic regression and decision tree. Results indicate high accuracy rates: Linear regression gives 88% and 95%, while decision trees provide %, which is approximately. With the model in a public web application, users get prognosis and advice for treatment based on the health information they enter.

# Chapter 4 – Implementation

## Introduction

A healthcare webpage gives people the chance to assess their cancer risks as well as the possibility of a heart attack. It relies on machine learning to take medical data given by users and use special algorithms to give forecasts.

Accessible forms for filling in important health information are provided on the Website because of its user-friendly design. After that. These inputs are handled with advanced algorithms that have received wide development based on different datasets. As soon as a visitor enters, the site gives an instant estimate of their risks for heart attack, stroke, or cancer.

The primary purpose of this program is to help consumers simply check whether any health risks might be there for them. As a result, those using the Website’s forecasting tool can take part in enhancing healthcare by allowing early diagnosis.

## Methodology

It meant gathering medical data, using decision trees to handle heart attack cases, using support vector machines in cancer prediction models, and building all this into a Django-powered website that anyone can use. So, the predictions became accurate, resulting in users having a satisfying experience.

## Machine Learning Model training steps

**1. Gathering and Preparing Data**

**Data Gathering:** Machine learning models use datasets that are collected in large enough amounts. At this step, we gathered cancer diagnosis and heart stroke risk data from research institutes and public health care services. They contained a variety of details about a patient such as age group, previous illnesses, lifestyles, and what tests were done.

**Preprocessing of the Data:** The data set was greatly prepared before any further steps. For this, the missing values had to be filled, all the data was to be kept in the same format, and steps were needed to correct the outliers to complete the cleaning. So each step would be on the same level and one step wouldn’t be more important due to having a different number, normalization and scaling were done. Taking this step helps ensure that the data is complete and proper for any future work with modelling (Maharana, Mondal and Nemade, 2022).

**2. Model Selection and Training**

**Support Vector Machines (SVM) for Cancer Detection:** One of the key things about SVMs is their ability to manage data with challenging boundaries. SVM is designed to classify data by finding the ideal line (hyperplane) between the points and splitting the data into separate groups. In cancer diagnostics, SVM was used because of its key benefits. (Mayers 2020).

**Heart Attack Decision Trees:** This is a type of supervised learning that relies on decision trees as their model. With the use of a decision tree, a predictive model for heart stroke prediction. Specifically, this model examined different risk markers and determined how likely an individual is to have a cardiac arrest (Ozcan & Peker, 2023).

This aimed at maximizing the training of the models through data partitioning into the training and test set. Through this, testing the quality of the model was conceivable. This prompted the utilization of hyperparameter improvement techniques like framework looking and irregular scanning to fine-tune the model boundaries for ideal performance. The significant target of this strategy was to upgrade precision, decrease mix-ups, and ensure that there could be no under-fitting or over-fitting.

**3. Validation and Evaluation of the Model**

**Performance Metrics:** Different assessments were utilized to evaluate the models’ performances. For this reason, conventional measures like accuracy, precision, recall, and F1-score were analyzed after proper calculation for each model to evaluate its ability to stay away from false /negatives (Eker et al., 2019).

## Development and Integration of Websites

**Django Framework:** To this end, Django, which has scaling capabilities and can be delegated a safe web framework with heaps of local treats, was utilized to develop the healthcare website. Its MVT architecture has contributed greatly to simplifying the development process while ensuring effective backend management (Bairagi et al., 2021).

**User-friendly Interface:** A natural and straightforward user interface was created using HTML, CSS, and Bootstrap. The responsive design of Bootstrap made the Website capable of adapting to different screen sizes, improving the user experience on smartphones, tablets, and desktop computers.

**Model Integration:** Using Python’s serialization method (e.g., pickle), the trained machine learning models were quickly incorporated into the Website’s backend. Through this combination, the Website was able to receive input from users, process it using trained models, and display the predictive findings right on the user interface (Ndzomga, 2023).



Figure Process diagram of machine learning model training

## Key Features

**Cancer Detection (SVM)**

The Cancer Prediction Model with Support Vector Machines exhibits exceptional accuracy, with a 97.2% precision rate in predicting probable cancer cases. Its ability to thoroughly analyze a wide range of datasets and make accurate predictions is what makes it so strong. This approach is excellent at early detection and offers priceless insights that make preventative healthcare treatments possible. Its dependability in evaluating risks provides users with tailored insights to help them make educated decisions and possibly take preventative action. This method ensures consumers rapid access, boosts health-awareness campaigns, and fosters interaction. It fits smoothly into a web-based platform (Cervantes et al., 2020).

**Heart Stroke Prediction (Decision Tree)**

The Heart Stroke prediction model constructed with Decision Trees is remarkable in that it has an accuracy rate of 94.63%. This well-known interpretable model provides customers with easily understandable insights into heart-related difficulties, therefore assisting them in understanding their risk factors. It does a great job of estimating the risk of myocardial infarctions and making suggestions for preventive care tactics (Dev et al., 2022). Its smooth integration into the web interface ensures easy access and usage and allows for rapid and simple user interaction. It assists by offering precise projections, empowering consumers to choose healthcare providers wisely, and promoting a proactive approach to better health management.

## Pages of Website (Site map)

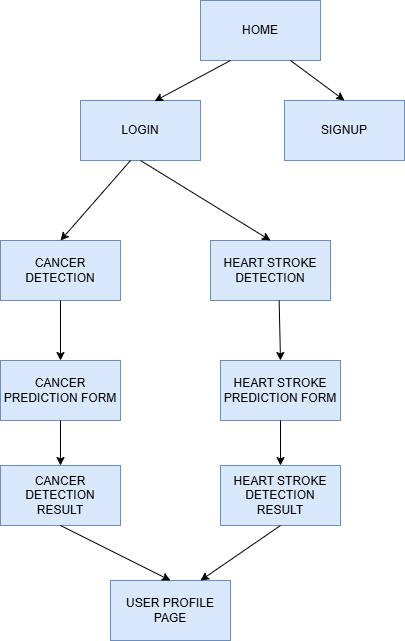


Figure Site map of the Website

**DFD:**

**A diagram of a flowchart

AI-generated content may be incorrect.**

Figure DFD Level 0 & Level 1

**Level 0 DFD (Context Diagram)**

This is the first level of DFD known as the Context Level.

We can see that the healthcare prediction system forms an entire process. It looks at the communication between users and the data sources in the system.

**Main Components:**

User:

* Creates an account or enters their login credentials.
* They add their details into the system for a prediction (for health disorders such as cancer or stroke).
* Takes a look at the result once it has been processed by the system.

Healthcare System (Main Process):

* Deals with authentication and predictions.
* Puts data into the Prediction Datasets and takes data out of them.

Prediction Datasets:

* Developer must add the data or models that are used to make predictions.
* Provides the needed information using the information typed in by the patient.

Purpose: To explain the main functions of the system without going into detail about how they are performed inside the corporation.

**Level 1 DFD – Detailed View**

It shows how the main process at Level 0 is divided into specific and more detailed activities. It highlights the way data is exchanged in different stages of the system.

Detailed Components:

User:

* You should begin by signing up or logging in to Elligence.
* Provide their medical information (for example, age, weight, blood pressure, and others).
* Views prediction results after processing.

Healthcare System:

* Verifies user credentials.
* Removes users’ login information from the Login & Register Database.
* Brings the user to the Home Page once they have successfully logged in.

Home Page:

Let users browse through the different options and pick the right prediction.

* Cancer Prediction
* Heart Stroke Prediction

Identifying the Risk of Cancer / Predicting Heart Stroke:

* People provide their personal health details within the required fields.
* The system takes the input, and then sends it to the model.
* All outcomes are put into storage in the correct data stores.
* Cancer Prediction Result Stored
* Heart Stroke Prediction Result Stored

Model Prediction Result:

* Gathers the predictions that were made by the selected model.
* Shows the final answer to the user.

Profile Page:

* Gives users the ability to see their past diagrams of possible outcomes.
* Return back to the home page with a click.

Data Stores Used:

* Information About New Users Was Saved in Database. Keeps user credentials.
* Cancer Prediction Result Stored: Saves cancer-related prediction outcomes.
* Heart Stroke Prediction Result Stored: Saves heart stroke-related prediction outcomes.

**Index/ home page: -**

When someone visits a website, they initially view the index page, sometimes referred to as the home page. It gives a summary of the goal of the Website, might contain some details about the products and services provided, and frequently has links to other parts of the site.

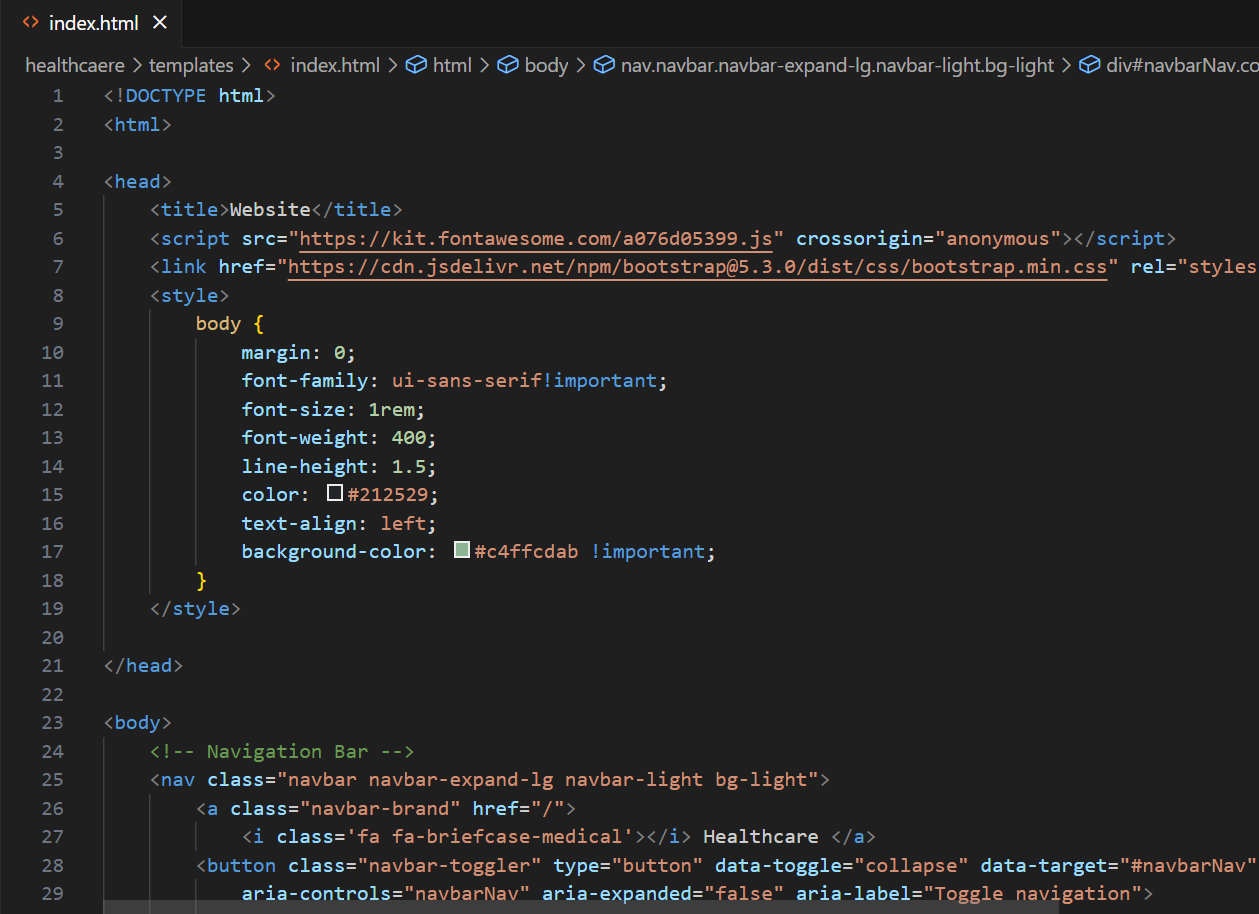


Figure Index Page1

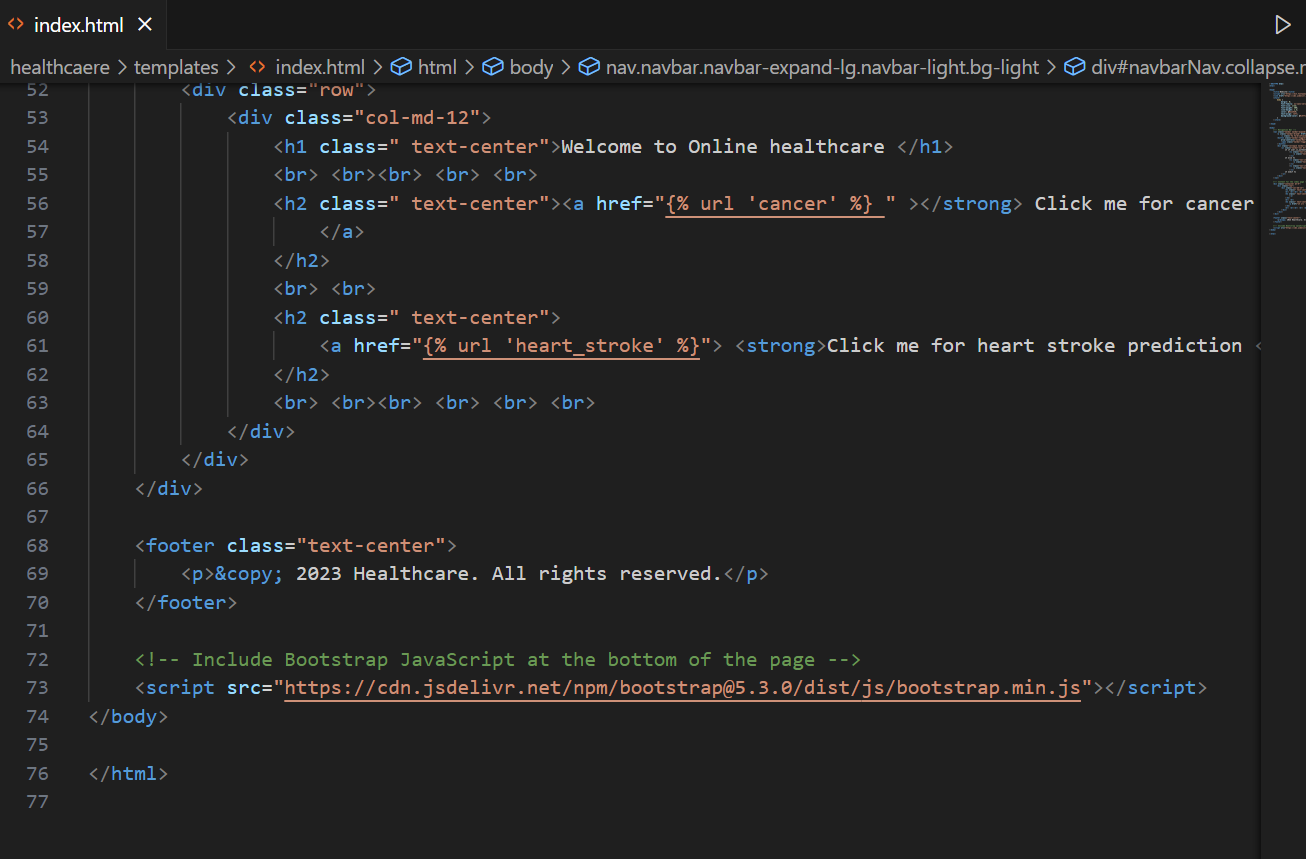


Figure 5 Code of index.html



Figure 6 Home page of the application1

A screenshot of a computer

AI-generated content may be incorrect.

Figure Home page of the application2

**Login page**

The login page is where users enter their login information to gain access to their accounts. This page lets users enter their username and password if they’re not logged in.

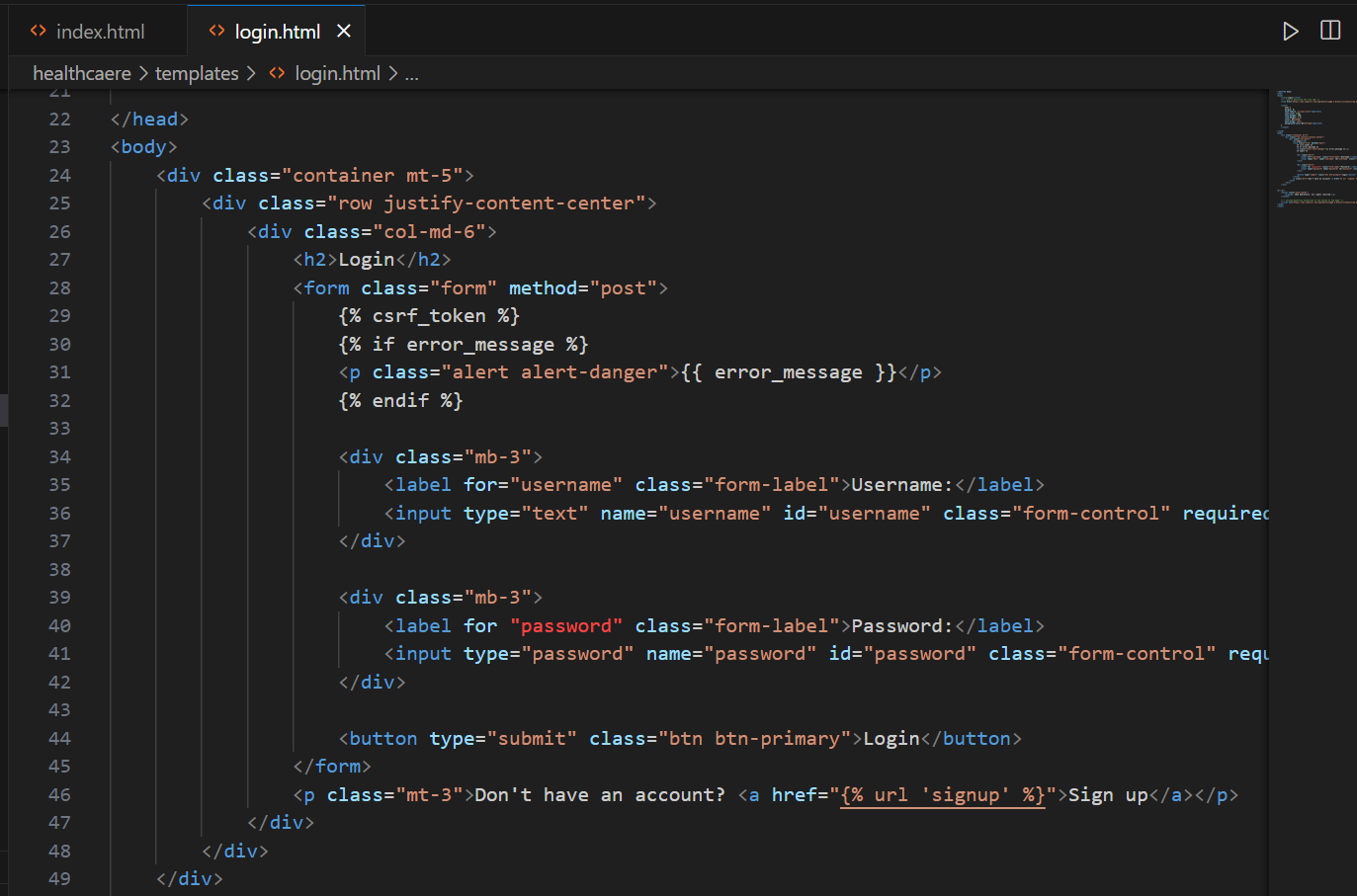


Figure 8 code of login.html

A screenshot of a login screen

AI-generated content may be incorrect.

Figure 9 Login page

**Signup page**

This page is where new users register and create accounts on the Website. It gathers essential data, including email addresses, passwords, and usernames.

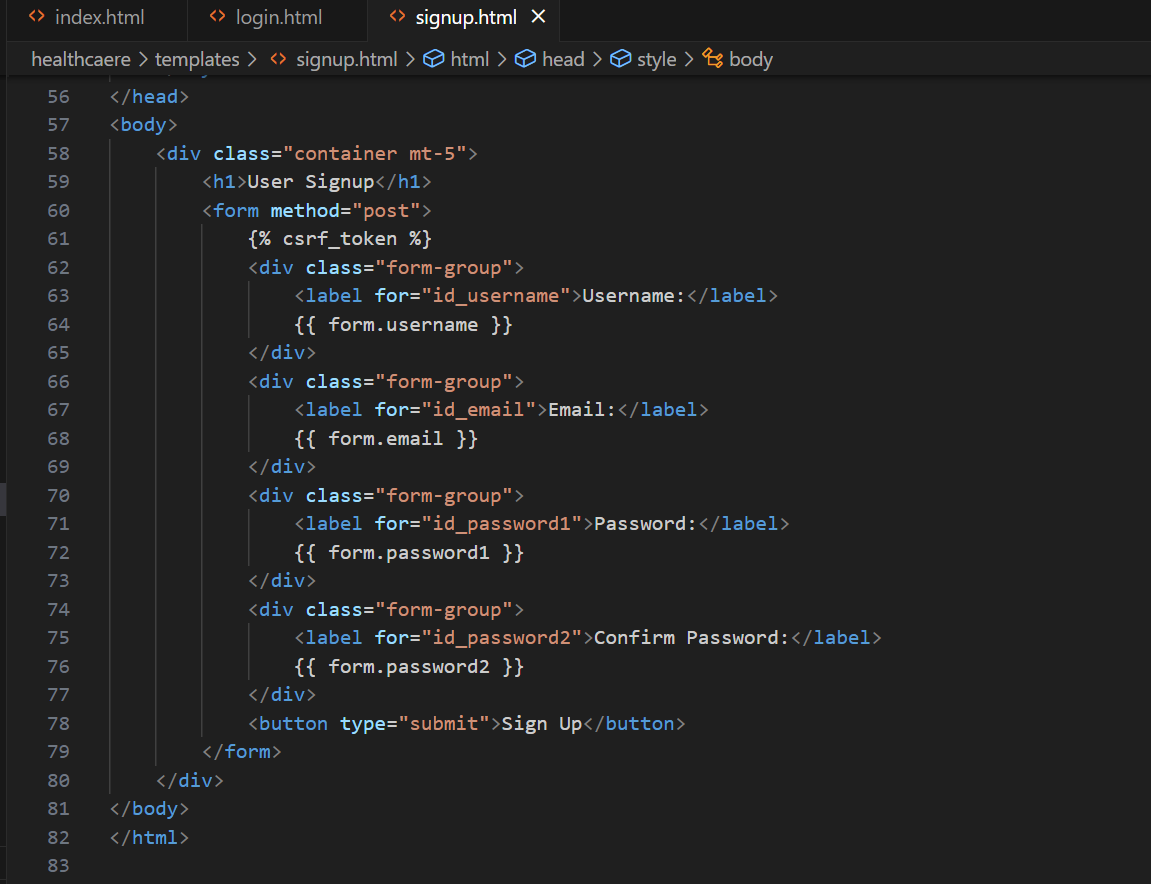


Figure 10 Signup.html code

A screenshot of a login form

AI-generated content may be incorrect.

Figure 11 Signup page

**Prediction Forms:** Distinct pages for predictions of heart attacks and cancer. Individuals provide data about their health, which is sent to machine learning models so they can make predictions. The prediction form pages are designed according to the independent features of both machine learning models.



Figure 12 Cancer prediction form

A screenshot of a cell phone

AI-generated content may be incorrect.

Figure 13 Cancer prediction form page

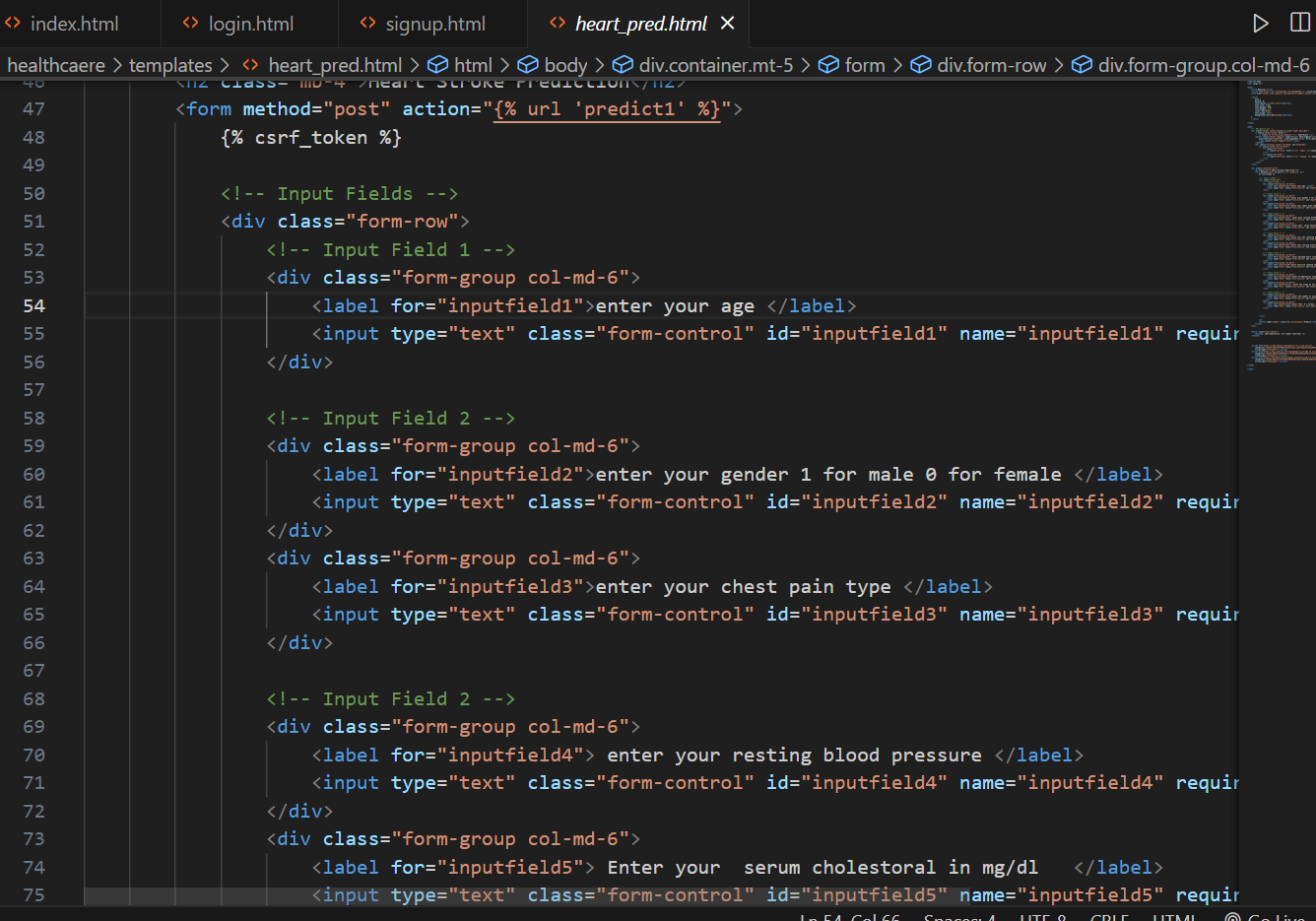


Figure 14 Heart stroke prediction form.html

A screenshot of a computer

AI-generated content may be incorrect.

Figure 15 Heart stroke prediction form page

**Prediction Results:** Users can view their prediction results, including whether they are more likely to develop cancer or have a heart attack, after providing their data. This page comes into role after the user input of the related model input.



Figure 16 heart prediction result



Figure 17 Cancer prediction result.html

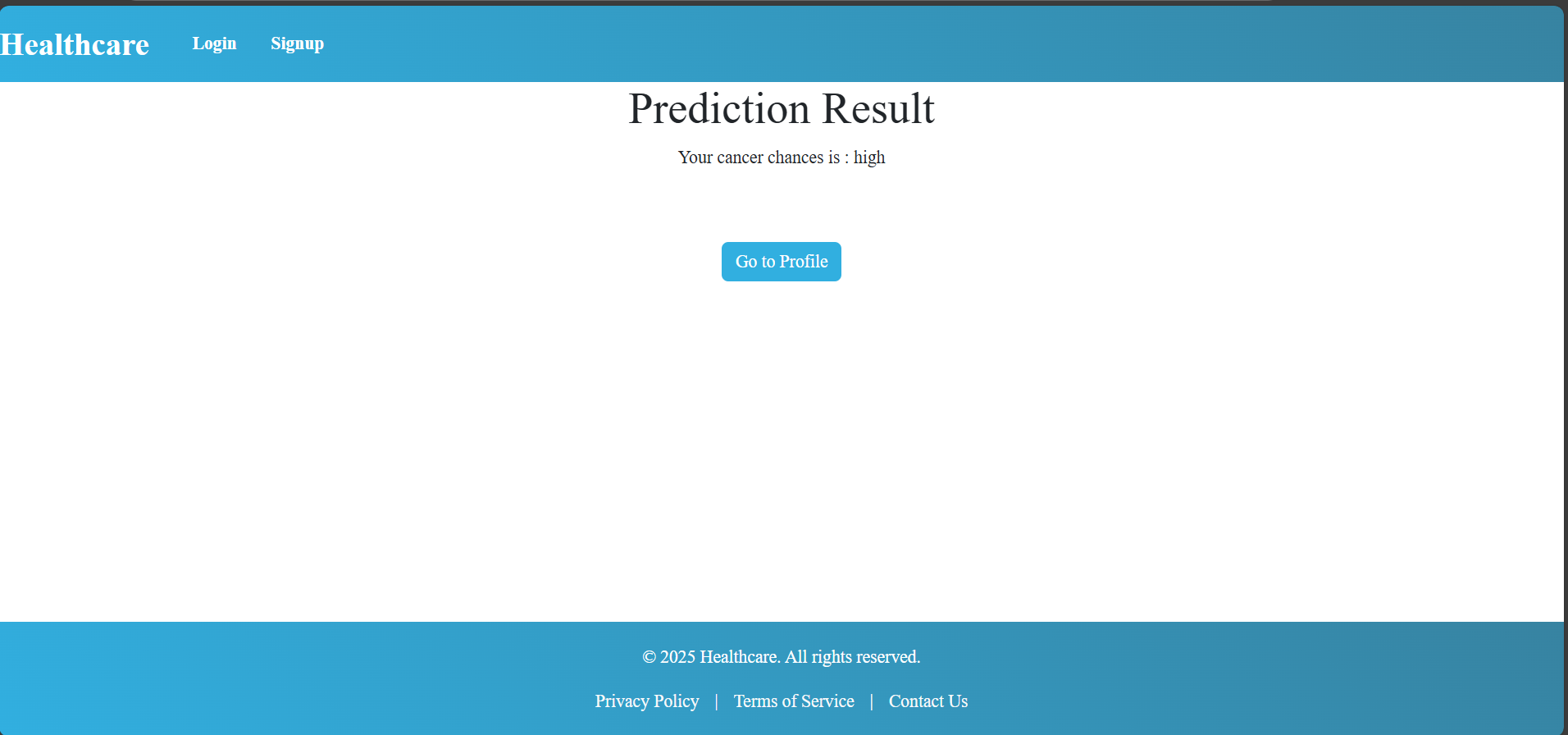


Figure 18 cancer prediction result page

**User Profile Page:**

**A white card with blue text

AI-generated content may be incorrect.**

Figure User Profile Page

**Admin Panel:** An area that is only available to admins that gives them the ability to view website statistics, manage user accounts, and carry out other tasks related to administration (Ambrozowicz, 2022).

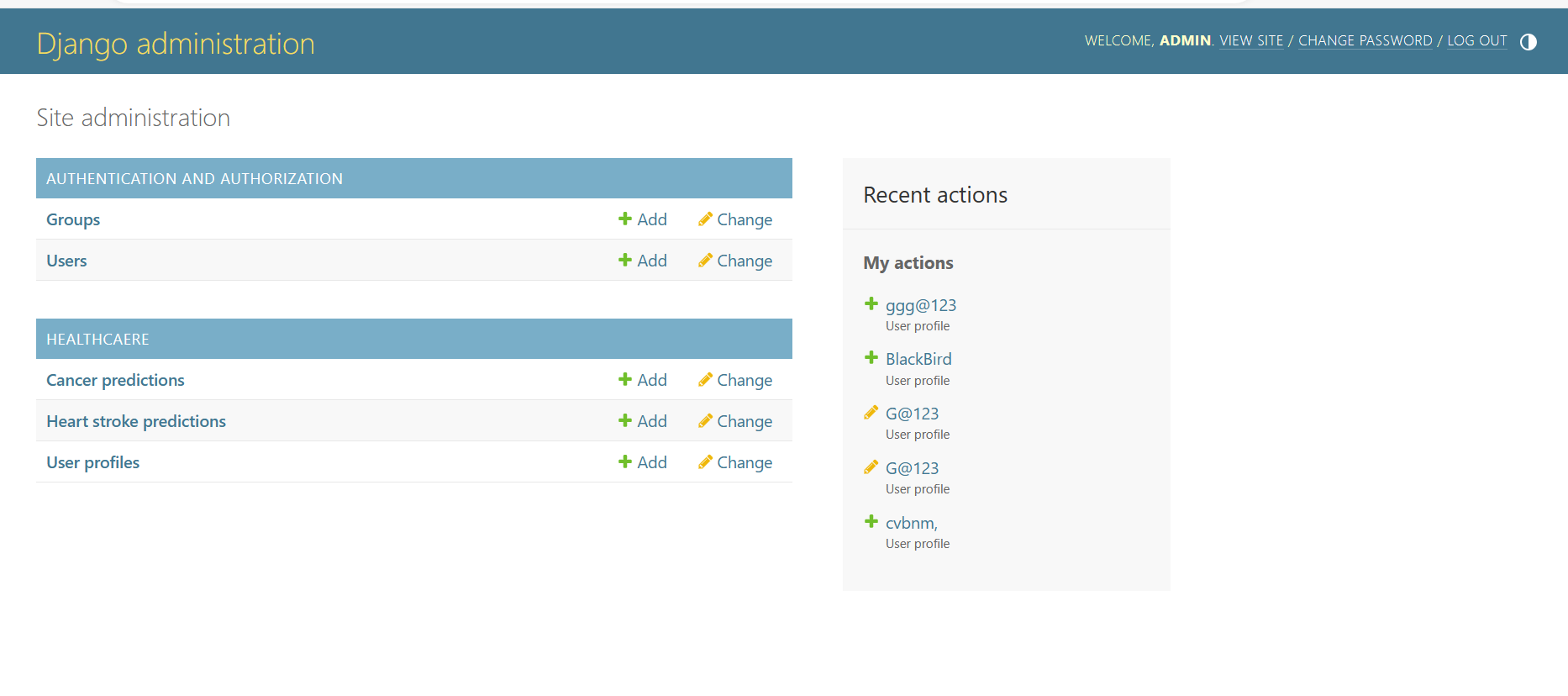


Figure 20 Admin panel of the Website

## Results and Outcomes

The models provide dependable insights, as evidenced by the achieved accuracy levels of 94.63% for decision tree modelling and 97.2% for SVM for cancer forecasting and heart stroke prediction, respectively. With such high accuracy levels, the healthcare platform’s dependability is increased as the models demonstrate their ability to generate precise predictions from the provided input data. By providing users with helpful information about their health status, the outcome illustrates the potential for precise early identifications and efficient risk assessments. With the help of such accurate forecasts, consumers can be proactive in reducing risks and enhancing the quality of healthcare.

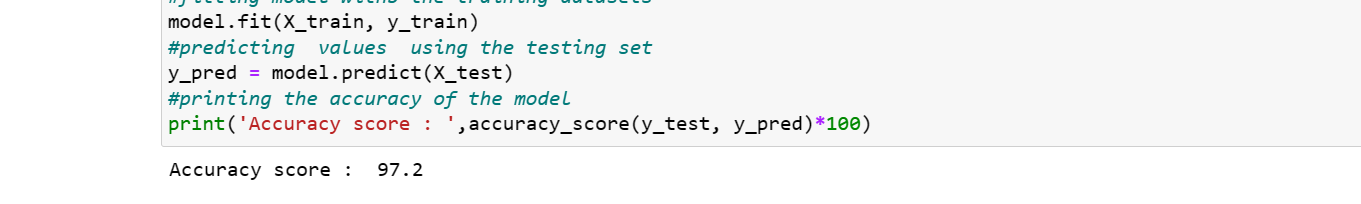


Figure 21 Cancer prediction model accuracy

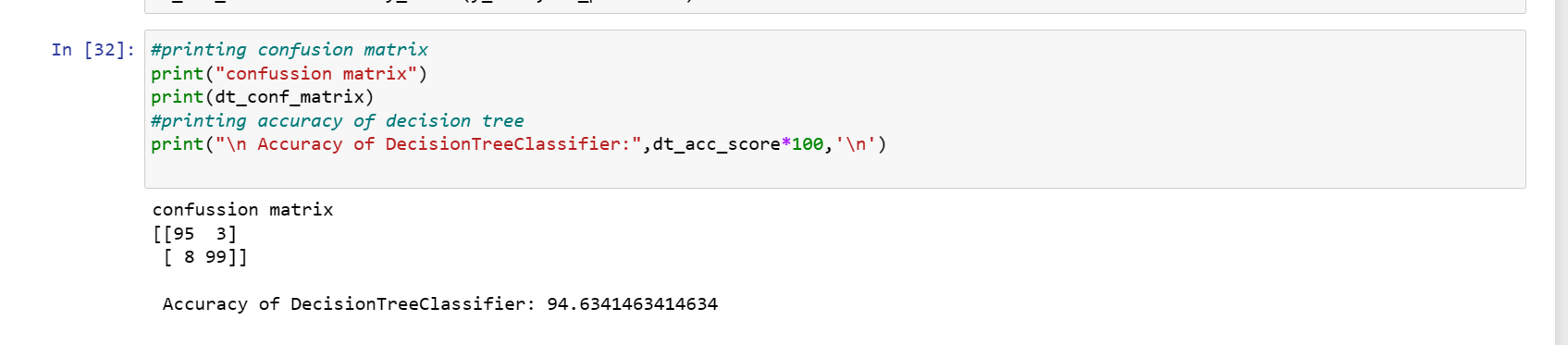


Figure 22Heart stroke model accuracy

## Conclusion

The usage of intelligent medical technology has advanced the way we handle our healthcare. By applying machine learning, we are able to predict cancer with nearly 97.2% accuracy and heart attacks with around 94.63% accuracy. Such programs help us understand health issues by acting as trusted assistants. Creating better healthcare decisions is hassle-free for users of the Website because they can easily check if they may be facing health challenges.

The goal of this project is to highlight the use of technology to find issues earlier and boost our health services. The service will help many people stay healthy and maintain their well-being.

## Summary

With this website, medical experts can assess the risk for heart attacks and cancers using Django-made web design and AI decision-making models. This involved constructing a Django website and creating a model. For cancer, the SVM can reach an accuracy of 97.2%, and the decision tree is able to detect heart attacks with an accuracy of 94.63%. The main features include prediction forms, machine learning for real-time results, and authentication for users. Several pages can be found on the Website, for example, the Admin Panel, the forms for predictions, a way to Signup, and another for Login. Machine learning and an easy-to-use interface have been successfully combined in the report for health risk assessment.

# Chapter 5 – Discussion and Analysis

**Technical Approach and Methodology:** Relying on reliable healthcare data from major institutions and databases plays a big role in making sure machine learning models are accurate. The team relied on data recording patient age, gender, medical background, habits, and test reports to achieve effective modelling.

It is established that Decision Trees are effective for predicting heart attacks and Support Vector Machines (SVM) for diagnosing cancer. Although Decision Trees provide the degree of openness and comprehension necessary for healthcare decision-making, Support Vector Machines (SVMs) are known for their capacity to deal with high-dimensional data (Parashar, 2023). Improving the stability of models, as well as their ability to be applied to new cases, is possible by making changes to how hyper parameters are found via grid search and random search. After rigorous testing, standard evaluation metrics are used to validate the method.

**Merging Machine Learning Models Smoothly into a Web Platform**: It demonstrates how to integrate trained machine learning algorithms into a Python-based Django web page by using serialization to save information and recall it. It makes the interface both accessible and easy to use by allowing predictions to be created immediately based on user input. The development of user-friendly prediction forms created specifically for heart attack and cancer evaluations guarantees the collection of specific disease characteristics, enabling precise predictions that meet specified model specifications (Françoise, Baptiste Caramiaux and Sanchez, 2021). Furthermore, the utilization of HTML, CSS, and Bootstrap to make a responsive and versatile U.I. underlines the devotion to giving the ideal survey insight on various gadgets. With its responsive plan, individuals of different gadget types may effectively get to and use the application, empowering comprehensiveness and expanding client take-up (Sanghvi, 2023).

**Important Features and Pages that Improve User Experience:** The proposal states clearly the crucial parts and structures that need to be improved to enhance the user experience within the medical application. For both coronary failure and malignant growth, the changes made ensure that reviews consider the vital infectious elements that aid in accurate estimations. This clearly shows how central stage upkeep, information protection, and effective organization are. It enables smooth workflow, ensures the organization’s client information is secure, and manages investigations on the site. Taking good care of the main areas of the application lets it run dependably, safely, and with customer-oriented highlights, also contributing to increased user satisfaction.

**Model Performance and Outcomes:** As a result of meticulous model creation, the application showed impressive accuracy for both respiratory failure prediction (94.63%) and cancer diagnosis (97.2%). Having a closer look at exactness, accuracy, review, and F1-score reveals that these measures tell us how successfully the model can deliver the correct answers. The choice trees used in the respiratory failure detection and backing vector machines utilized by the disease prediction models were completely prepared on a range of datasets derived from electronic health records (Nabipour et al., 2020). Therefore, this advantage comes with added reliability and excellent accuracy, which are very useful for confirming the credibility. They assist in providing brief counseling and monitoring health programs, as well as letting people pick their own solutions. Using improved simulations, medical staff can decide the best way to treat patients and arrange their resources.

**Ethical Considerations in Healthcare Predictive Models:** There are ethical matters in using A.I. models in the healthcare industry concerning transparency, neutrality, and information confidentiality. As a result, when using models, large datasets must be used while upholding persistent privacy and data anonymization. Such systems include the GDPR and HIPAA (Morley et al., 2020).

It is also necessary to tackle the issues of algorithmic and racial bias. Healthcare procedures should be evaluated to determine whether or not they have discriminatory features that could negatively impact specific populations. Particularly in sensitive medicinal cases, the bias-reducing steps, as well as the assumption of equal forecast, are theses that should be highlighted. Understanding a model is important if it is to gain the confidence, trust, and acceptance of patients and healthcare professionals (Farhud & Zokaei, 2021).

**Future Implications and Scalability:** While creating the application, healthcare experts need to think about its future potential and its use to scale it for a larger audience. One possible method is to include forecasts about other diseases in the platform in addition to cancer and heart attacks. To serve an expanding user base, an application needs to handle a lot of data and users without performance issues.  
In addition, discussing potential collaborations with health facilities or incorporating live data streams in the dynamic forecasts may showcase the application’s malleable nature and versatility (Pugliese, Regondi and Marini, 2021). By highlighting a plan for future improvement, possible partnerships, or study expansion in healthcare technology, the progressive strategy of the project demonstrates its contribution to the betterment of healthcare facilities.

**Limitations and Recommendations:** Still, even with the innovation, there are certain limitations in the health application that need to be addressed before it can be legally applied. Furthermore, they cannot easily be re-trained due to their use of static older data. Therefore, it is important to regularly update the system while integrating live input data streams. Complete privacy must be maintained for patients and biases should be reduced, despite using data protection methods. Being strict about complying with HIPAA and GDPR, and regularly testing for bias, may reduce ethical problems. It is still challenging to understand how machine learning models make decisions. Applying the method of model-agnostic interpretability can show why predictions were made, which should make users believe the results more.

Scalability challenges should be handled, especially for situations with rising powers on the prediction side or as customers need more. Better system performance and convenience for users could be provided by enhancing the architecture and increasing the resources. It is important for future research to assess models using guidelines suited for real-time clinical use. Teamwork with healthcare companies increases the generalizability of our model by exposing it to different types of data.

## Summary

It clarifies how to develop an app that predicts heart and cancer diseases. Trustworthy data sets, as well as support vector machines and decision trees, are important for medical diagnoses. Bringing machine learning into the web interface makes it more accessible to users. Significant improvements in terms of user experience and high accuracy levels for heart attack prediction (94.63%) and cancer disease prediction (97.20%) are noteworthy. It includes future considerations, scaling restrictions, mentions privacy, biased results, and commits to dedicating continuous watch for anything legal and ethical.

# Chapter 6 - Conclusion

Because of the advances in machine learning, healthcare is moving towards better management, especially when it comes to predicting heart and cancer problems. More people are struggling to predict illnesses due to complications from their lifestyle choices and the environment. Even so, the paper offers a solution that applies SVM, linear regression, and decision tree methods, aiming to boost the accuracy of illness prediction. The main aim of the project was to focus on developing an app for identifying likelihood of heart and cancer diseases. These A.I. techniques provided excellent outcomes, with the choice tree getting 86.7% precision, SVM reaching an 88% accuracy rate, and straight relapse achieving 95% accuracy. This further increases the accuracy, allowing people with health problems to get guidance and treatment online instead of needing in-person visits.

The report aims to demonstrate the main virtues of the app, encourage medical plans, and detail A.I.’s role in healthcare. Additionally, it explores the potential of the software in managing illnesses and how that affects improving health results for patients, answering several analysis questions that cover these areas. By utilizing new clinical advancements, it is now possible to foresee illnesses, such as cancer and heart attacks, with almost 100% exactness. It is important because it is precise and open-minded. By implementing these modern applications on the site, people can efficiently choose their best medical options. It brings a proactive approach to handling medical care organizations.

## Summary

The research indicates that the app performs well by predicting illness based on a variety of factors like life patterns and environmental conditions. Such technological integration, on top of being highly accurate, also supports improvements in the field of healthcare for convenient medical decision-making.

# References

Ahsan, M.M. and Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, p.102289. doi: <https://doi.org/10.1016/j.artmed.2022.102289>.

Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, 30, p.100. doi: <https://doi.org/10.1016/j.imu.2022.100924>.

Alanazi, R 2022, ‘Identification and Prediction of Chronic Diseases Using Machine Learning Approach’, *Journal of Healthcare Engineering*, vol. 2022, p. e2826127.

Alharthi, H. (2018). Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *Journal of Infection and Public Health*, 11(6), pp.749–756. doi <https://doi.org/10.1016/j.jiph.2018.02.005>.

Ambrozowicz, M 2022, *How to build a Django admin panel?* Forest Admin Blog, viewed 14 November 2023, <https://www.forestadmin.com/blog/how-to-build-a-django-admin-panel/>.

Bansal, A., Ahirwar, M.K. and Shukla, P.K. (2018). A Survey on Classification Algorithms Used in Healthcare Environment of the Internet of Things. *International Journal of Computer Sciences and Engineering*, 6(7), pp.883–887. Doi: <https://doi.org/10.26438/ijcse/v6i7.883887>.

Battineni, G., Sagaro, G.G., Chinatalapudi, N. and Amenta, F. (2020). Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. *Journal of Personalized Medicine*, [online] 10(2). doi: <https://doi.org/10.3390/jpm10020021>.

Calzon, B. (2023). *Learn Here Different Ways of Data Analysis Methods and Techniques*. [online] BI Blog | Data Visualization & Analytics Blog | Datapine. Available at: <https://www.datapine.com/blog/data-analysis-methods-and-techniques/> [Accessed 6 Nov. 2023].

Center for Drug Evaluation and Research (2019). *IRBs and Protection of Human Subjects*. [online] U.S. Food and Drug Administration. Available at: <https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/institutional-review-boards-irbs-and-protection-human-subjects-clinical-trials> [Accessed 6 Nov. 2023].

Dahiwade, D, Patel, G & Meshram, E 2019, *Designing Disease Prediction Model Using Machine Learning Approach*, IEEE Xplore, pp. 1211–1215.

Farhud, D. D., & Zokaei, S. (2021). Ethical Issues of Artificial Intelligence in Medicine and Healthcare. *Iranian Journal of Public Health*, *50*(11). <https://doi.org/10.18502/ijph.v50i11.7600>

Fatima, N., Liu, L., Hong, S., and Ahmed, H. (2020). Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis. *IEEE Access*, 8, pp.150360–150376. doi: <https://doi.org/10.1109/access.2020.3016715>.

George, T. (2021). *An introduction to mixed methods research*. [online] Scribbr. Available at: <https://www.scribbr.com/methodology/mixed-methods-research/#:~:text=Mixed%20methods%20research%20combines%20elements> [Accessed 6 Nov. 2023].

Gupta, A. (2020). *Feature Selection Techniques in Machine Learning*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/> [Accessed 6 Nov. 2023].

Islam, M., Hasan, M., Wang, X., Germack, H. and Noor-E-Alam, M. (2018). A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. *Healthcare*, [online] 6(2), p.54. doi <https://doi.org/10.3390/healthcare6020054>.

Javeed, A., Khan, S.U., Ali, L., Ali, S., Imrana, Y. and Rahman, A. (2022). Machine Learning-Based Automated Diagnostic Systems Developed for Heart Failure Prediction Using Different Types of Data Modalities: A Systematic Review and Future Directions. *Computational and Mathematical Methods in Medicine*, [online] 2022, p.9288452. doi: <https://doi.org/10.1155/2022/9288452>.

Khan, Y., Qamar, U., Wesaf, N., and Khan, A. (2019). Machine Learning Techniques for Heart Disease Datasets. *Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC ’19*. doi: <https://doi.org/10.1145/3318299.3318343>.

Kuhn, G. (2023). *What is Mixed-Method Research? [+ Examples & Benefits]*. [online] www.driveresearch.com. Available at: <https://www.driveresearch.com/market-research-company-blog/what-is-mixed-mode-data-collection-marketing-research-firm-syracuse-ny/> [Accessed 6 Nov. 2023].

Kumar, T. (2022). A Review of Role of Machine Learning Models in Coronary Heart Disease Detection Accuracy. *Inflibnet.ac.in*. [online] doi <http://gnanaganga.inflibnet.ac.in:8080/jspui/handle/123456789/1101>.

Li, J., Zhou, Z., Dong, J., Fu, Y., Li, Y., Luan, Z. and Peng, X. (2021). I am predicting breast cancer 5-year survival using machine learning: A systematic review. *PLOS ONE*, 16(4), p.e0250370. doi: <https://doi.org/10.1371/journal.pone.0250370>.

M., S. and Chattu, V.K. (2021). A Review of Artificial Intelligence, Big Data, and Blockchain Technology Applications in Medicine and Global Health. *Big Data and Cognitive Computing*, 5(3), p.41. doi: <https://doi.org/10.3390/bdcc5030041>.

Maji, S. and Arora, S. (2018). Decision Tree Algorithms for Prediction of Heart Disease. *Information and Communication Technology for Competitive Strategies*, [online] pp.447–454. doi <https://doi.org/10.1007/978-981-13-0586-3_45>.

Mayers, G 2020, *Classifying Malignant or Benignant Breast Cancer using SVM*, Analytics Vidhya, viewed 14 November 2023, <https://medium.com/analytics-vidhya/classifying-malignant-or-benignant-breast-cancer-using-svm-fe36f139dd21>.

Mienye, I.D., Sun, Y. and Wang, Z. (2019). Prediction performance of improved decision tree-based algorithms: a review. *Procedia Manufacturing*, 35, pp.698–703. doi <https://doi.org/10.1016/j.promfg.2019.06.011>.

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of A.I. in health care: A mapping review. *Social Science & Medicine*, *260*, 113172. <https://www.sciencedirect.com/science/article/pii/S0277953620303919>

Ndzomga, FS 2023, *How To Integrate Machine Learning Models To Wer Web Applications*, MLearning.ai, viewed 14 November 2023, <https://medium.com/mlearning-ai/how-to-integrate-machine-learning-models-to-wer-web-applications-778ee9c62149>.

Obi Tayo, y B. (2023). *Exploring Data Cleaning Techniques With Python*. [online] KDnuggets. Available at: <https://www.kdnuggets.com/2023/04/exploring-data-cleaning-techniques-python.html> [Accessed 6 Nov. 2023].

Ozcan, M & Peker, S 2023, ‘A classification and regression tree algorithm for heart disease modeling and prediction’, *Healthcare Analytics*, vol. 3, p. 100130.

Parashar, N. (2023, August 1). *The Role of Artificial Intelligence in Cancer Detection*. Medium. <https://medium.com/@niitwork0921/the-role-of-artificial-intelligence-in-cancer-detection-8c19606fdee8>

Rafique, R., Islam, S.M.R. and Kazi, J.U. (2021). Machine learning in the prediction of cancer therapy. *Computational and Structural Biotechnology Journal*, 19, pp.4003–4017. doi: <https://doi.org/10.1016/j.csbj.2021.07.003>.

Sanghvi, K. (2023, April 21). *Artificial Intelligence in Healthcare*. Medium. <https://medium.com/@kavishsanghvi/artificial-intelligence-in-healthcare-79d35439accd>

Sharma, V. and Yadav, S. (2020). Heart Disease Prediction using Machine Learning Techniques. *International Journal of Innovative Technology and Exploring Engineering*, [online] 9(5), pp.1456–1460. doi: <https://doi.org/10.35940/ijitee.e2862.039520>.

Singh, A. and Kumar, R. (2020). Heart Disease Prediction Using Machine Learning Algorithms. *IEEE Xplore*, [online] pp.452–457. doi <https://doi.org/10.1109/ICE348803.2020.9122958>.

Straw, I. and Wu, H. (2022). Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. *BMJ Health & Care Informatics*, [online] 29(1), p.e100457. doi <https://doi.org/10.1136/bmjhci-2021-100457>.

T., M, Mukherji, D, Padalia, N & Naidu, A 2013, ‘A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)’, *International Journal of Computer Applications*, vol. 68, no. 16, pp. 11–15, viewed 23 January 2020, <https://pdfs.semanticscholar.org/668c/99225fea30b6b7f9305b3eab19f71bba9de4.pdf>

Tiwari, N. (2021). *Data Cleaning Using Pandas | Data Cleaning for Beginners*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2021/06/data-cleaning-using-pandas/> [Accessed 6 Nov. 2023].

Tumpa, E.S. and Dey, K. (2022). A Review on Applications of Machine Learning in Healthcare. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. doi <https://doi.org/10.1109/icoei53556.2022.9776844>.

White, M.G. (2020). Why Human Subjects Research Protection Is Important. *The Ochsner Journal*, [online] 20(1), pp.16–33. doi: <https://doi.org/10.31486/toj.20.5012>.

Yahaya, L., David Oye, N. and Joshua Garba, E. (2020). A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques. *American Journal of Artificial Intelligence*, 4(1), p.20. doi: <https://doi.org/10.11648/j.ajai.20200401.12>.

Zhang, B., Shi, H. and Wang, H. (2023). Machine Learning and A.I. in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach. *Machine Learning and A.I. in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach*, Volume 16, pp.1779–1791. doi: <https://doi.org/10.2147/jmdh.s410301>.

Aguinis, H., Hill, N.S. and Bailey, J.R. (2019). Best Practices in Data Collection and Preparation: Recommendations for Reviewers, Editors, and Authors. *Organizational Research Methods*, 24(4), p.109442811983648. doi:https://doi.org/10.1177/1094428119836485.

Ahmed, N., Ahammed, R., Islam, Md.M., Uddin, Md.A., Akhter, A., Talukder, Md.A.-A. and Paul, B.K. (2021). Machine learning based diabetes prediction and development of smart web application. *International Journal of Cognitive Computing in Engineering*, [online] 2, pp.229–241. doi:https://doi.org/10.1016/j.ijcce.2021.12.001.

Bae, J.-M. (2014). Clinical Decision Analysis using Decision Tree. *Epidemiology and Health*, [online] 36, p.e2014025. doi:https://doi.org/10.4178/epih/e2014025.

Bairagi, I.A., Sharma, A., Rana, B.K. and Singh, A. (2021). *UNO: A Web Application using Django*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICAC3N53548.2021.9725577.

Benkessirat, A. and Benblidia, N. (2019). Fundamentals of Feature Selection: An Overview and Comparison. *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*. [online] doi:https://doi.org/10.1109/aiccsa47632.2019.9035281.

Caracelli, V.J. and Greene, J.C. (1993). Data Analysis Strategies for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 15(2), pp.195–207. doi:https://doi.org/10.3102/01623737015002195.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, [online] 408(1), pp.189–215. doi:https://doi.org/10.1016/j.neucom.2019.10.118.

Charbuty, B. and Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, [online] 2(01), pp.20–28. doi:https://doi.org/10.38094/jastt20165.

Dev, S., Wang, H., Nwosu, C.S., Jain, N., Veeravalli, B. and John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, 2, p.100032. doi:https://doi.org/10.1016/j.health.2022.100032.

Eker, S., Rovenskaya, E., Langan, S. and Obersteiner, M. (2019). Model validation: A bibliometric analysis of the literature. *Environmental Modelling & Software*, 117, pp.43–54. doi:https://doi.org/10.1016/j.envsoft.2019.03.009.

Françoise, J., Baptiste Caramiaux and Sanchez, T. (2021). Marcelle: Composing Interactive Machine Learning Workflows and Interfaces. *Marcelle: Composing Interactive Machine Learning Workflows and Interfaces*. doi:https://doi.org/10.1145/3472749.3474734.

G, A., Ganesh, B., Ganesh, A., Srinivas, C., Dhanraj and Mensinkal, K. (2022). Logistic regression technique for prediction of cardiovascular disease. *Global Transitions Proceedings*, [online] 3(1), pp.127–130. doi:https://doi.org/10.1016/j.gltp.2022.04.008.

Jain, A.K., Debnath, N. and Jain, A.K. (2022). APuML: An Efficient Approach to Detect Mobile Phishing Webpages using Machine Learning. *Wireless Personal Communications*, [online] 125. doi:https://doi.org/10.1007/s11277-022-09707-w.

Javaid, M., Haleem, A., Pratap Singh, R., Suman, R. and Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, [online] 3, pp.58–73. doi:https://doi.org/10.1016/j.ijin.2022.05.002.

Kaul, D., Raju, H. and Tripathy, B.K. (2021). Deep Learning in Healthcare. *Studies in big data*, [online] pp.97–115. doi:https://doi.org/10.1007/978-3-030-75855-4\_6.

Maharana, K., Mondal, S. and Nemade, B. (2022). A Review: Data Preprocessing and Data Augmentation Techniques. *Global Transitions Proceedings*, 3(1), pp.91–99. doi:https://doi.org/10.1016/j.gltp.2022.04.020.

Mo, H., Tao, T., Gao, W., Fan, C., Wang, L., Zhao, X. and Wang, H. (2023). Application of Decision Tree Algorithm in Medical Nursing and Health Assessment System. *Application of Decision Tree Algorithm in Medical Nursing and Health Assessment System*. doi:https://doi.org/10.1109/acctcs58815.2023.00066.

Nabipour, M., Nayyeri, P., Jabani, H., S., S. and Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, [online] 8, pp.150199–150212. doi:https://doi.org/10.1109/access.2020.3015966.

Orrù, P.F., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R. and Arena, S. (2020). Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry. *Sustainability*, [online] 12(11), p.4776. doi:https://doi.org/10.3390/su12114776.

Pugliese, R., Regondi, S. and Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Science and Management*, [online] 4. doi:https://doi.org/10.1016/j.dsm.2021.12.002.

Rong, G., Mendez, A., Bou Assi, E., Zhao, B. and Sawan, M. (2020). Artificial Intelligence in Healthcare: Review and Prediction Case Studies. *Engineering*, [online] 6(3). doi:https://doi.org/10.1016/j.eng.2019.08.015.

Shaikh, F.J. and Rao, D.S. (2021). Prediction of Cancer Disease using Machine learning Approach. *Materials Today: Proceedings*. [online] doi:https://doi.org/10.1016/j.matpr.2021.03.625.

Silhavy, R., Silhavy, P. and Prokopova, Z. (2017). Analysis and selection of a regression model for the Use Case Points method using a stepwise approach. *Journal of Systems and Software*, [online] 125, pp.1–14. doi:https://doi.org/10.1016/j.jss.2016.11.029.

Sodhi, P., Awasthi, N. and Sharma, V. (2019). Introduction to Machine Learning and Its Basic Application in Python. *SSRN Electronic Journal*. [online] doi:https://doi.org/10.2139/ssrn.3323796.

Surya Gunawan, T., Aleah Jehan Abdullah, N., Kartiwi, M. and Ihsanto, E. (2020). *Social Network Analysis using Python Data Mining*. [online] IEEE Xplore. doi:https://doi.org/10.1109/CITSM50537.2020.9268866.

Zabor, E.C., Reddy, C.A., Tendulkar, R.D. and Patil, S. (2022). Logistic Regression in Clinical Studies. *International Journal of Radiation Oncology\*Biology\*Physics*, [online] 112(2), pp.271–277. doi:https://doi.org/10.1016/j.ijrobp.2021.08.007.

# Appendix

## Survey questions

1. What is your age?
2. Do you use any tobacco-based products or smoke cigarettes regularly?

* Yes
* No

1. What is your average heart rate or BPM?
2. Do you frequently consume alcoholic drinks?

* Yes
* No

1. What is your blood pressure?
2. Do you have high cholesterol?

* Yes
* No

1. Have you experienced any symptoms of heart or cancer disease?

* Yes
* No

1. Do you experience the severity of shortness of breath?

* Yes
* No

1. Are you on medication/treatment for any heart or cancer disease?

* Yes
* No

Q10. Do you have a family history of heart or cancer disease?

* Yes
* No