TRIBHUVAN UNIVERSITY

INSTITUTE OF SCIENCE AND TECHNOLOGY (IOST)

BHAKTAPUR MULTIPLE CAMPUS

Dudhpati-1, Bhaktapur



A

PROJECT REPORT

ON

“ADVANCED HEALTH PREDICTION SYSTEM USING SUPPORT VECTOR MACHINE (SVM)”

In partial fulfillment of the requirements for the Bachelor’s Degree in Computer

Science and Information Technology

(Course Code: CSC412)

SUBMITTED TO

Department of Computer Science and Information Technology

Bhaktapur Multiple Campus

SUBMITTED BY

Anusha Wagle (15132/074)

Arjun Shrestha (15133/074)

Nikita Rijal (15148/074)

Date of Submission : 2022/02/10

# BONAFIDE CERTIFICATE

Certified that, this project report entitled “ADVANCED HEALTH PREDICTION SYSTEM USING SUPPORT VECTOR MACHINE (SVM)” is the Bonafide work of Ms. Anusha Wagle (15132/074), Mr. Arjun Shrestha (15133/074), Ms. Nikita Rijal (15148/074) who carried out the research under my supervision. Certified further, that to the best of my knowledge the work reported here in does form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

…………………..

Mr. Pratik Timalsena

Project Supervisor

Date:…………………

# SUPERVISOR’S RECOMMENDATION

It is my pleasure to recommend that a report on “ADVANCED HEALTH PREDICTION SYSTEM USING SUPPORT VECTOR MACHINE (SVM)” has been prepared under my supervision by Ms. Anusha Wagle, Mr. Arjun Shrestha, Ms. Nikita Rijal in partial fulfillment of the requirement of the degree of Bachelor of Science in Computer Science and Information Technology (BSc.CSIT). This report is satisfactory and is an original work done by them to process for the future evaluation.

……………………

Mr. Pratik Timalsena

Supervisor

Department of Science and IT

Bhaktapur Multiple Campus

Date:………………..

# CERTIFICATE OF APPROVAL

The undersigned certify that he has read and recommended to the Department of Computer and Information Technology for acceptance, a project report entitled “ADVANCED HEALTH PREDICTION SYSTEM USING SUPPORT VECTOR MACHINE (SVM) ” submitted by Ms. Anusha Wagle, Mr. Arjun Shrestha, Ms. Nikita Rijal in partial fulfillment for the degree of Bachelor of Science in Computer Science and Information Technology (BSc.CSIT), Institute of Science and Technology, Tribhuvan University.

…………………….. ………………….

Mr. Pratik Timalsena External Examiner

Supervisor

Department of Computer

Science & IT

Bhaktapur Multiple Campus

………………………………

Sushant Poudel

Program Coordinator

# ACKNOWLEDGEMENT

We would like to express our appreciation to all those who provided us the possibility to complete the report. A special gratitude we want to give to our project supervisor, Mr. Pratik Timalsena, whose contribution in providing suggestions and encouragement helped us to coordinate our project especially in writing this report.

Furthermore, we are highly indebted to Bhaktapur Multiple Campus for constant guidance and supervision, as well as for providing all the necessary ICT infrastructure and friendly environment for the successful completion of the project of developing a web-based application called “Advanced Health Prediction System Using Support Vector Machine (SVM)”. We are also appreciative of the efforts of BSc. CSIT coordinator Mr. Sushant Poudel without his supporting role, the project would have been nowhere near completion. Special thanks go to our team mate, who were of great help to assemble the parts and gave their time and effort to accomplish the project. We would like to express our gratitude to our project supervisor Mr. Pratik Timalsena who took keen interest on our project and guided us throughout the project by providing all the necessary ideas, information and knowledge for developing website. He has supervised us well especially in our project presentation that has improved our presentation skills.

We are thankful and fortunate enough to get constant support from every teaching staff of BSc. CSIT department which helped us to successfully complete our project. We would also like to extend our regards to all the non-teaching staff of BSc. CSIT department for their timely support.

With respect,

Anusha Wagle (15132/074)

Arjun Shrestha (15133/074)

Nikita Rijal (15148/074)

# ABSTRACT

Advanced Health Prediction system incorporates the application of supervised machine learning algorithm for the prediction of disease from the symptoms which are given by the users to the system. This system enables the doctor/patient to know about the disease according to the symptoms. The study aims to provide relief to many of the busy doctors as well as patients who spent their days and life visiting the hospitals for treatment. The datasets of the diseases produced by the medical industries contain a wealth of hidden information which could be interpreted into useful ones through data mining techniques. The proposed model utilizes the Support Vector Machines (SVM) as the data mining or machine learning algorithm due to its high accuracy than any other algorithms in the field of the general disease prediction system.

Keywords: prediction, datasets, support vector machines (SVM), machine learning, data mining.

# TABLE OF CONTENTS

[BONAFIDE CERTIFICATE i](#_Toc95405616)

[SUPERVISOR’S RECOMMENDATION ii](#_Toc95405617)

[CERTIFICATE OF APPROVAL iii](#_Toc95405618)

[ACKNOWLEDGEMENT iv](#_Toc95405619)

[ABSTRACT v](#_Toc95405620)

[TABLE OF CONTENTS vi](#_Toc95405621)

[LIST OF FIGURES viii](#_Toc95405622)

[LIST OF ABBREVIATIONS ix](#_Toc95405623)

[Chapter 1: Introduction 1](#_Toc95405624)

[1.1. Introduction 1](#_Toc95405625)

[1.2. Problem Statement 1](#_Toc95405626)

[1.4. Scope and Limitations 2](#_Toc95405627)

[1.5. Development Methodology 3](#_Toc95405628)

[1.6. Report Organization 3](#_Toc95405629)

[Chapter 2: Background Study and Literature Review 4](#_Toc95405630)

[2.1. Background Study 4](#_Toc95405631)

[2.2. Literature Review 5](#_Toc95405632)

[Chapter 3: System Analysis 7](#_Toc95405633)

[3.1. System Analysis 7](#_Toc95405634)

[3.1.1. Requirement Analysis 7](#_Toc95405635)

[i. Functional Requirements 7](#_Toc95405636)

[ii. Non Functional Requirements 9](#_Toc95405637)

[3.1.2. Feasibility Study 9](#_Toc95405638)

[i. Technical Feasibility 9](#_Toc95405639)

[ii. Operational Feasibility 10](#_Toc95405640)

[iii. Economic Feasibility 10](#_Toc95405641)

[iv. Schedule Feasibility 10](#_Toc95405642)

[3.1.3. Analysis 11](#_Toc95405643)

[ Data Modelling using ER Diagram 11](#_Toc95405644)

[ Process Modelling using level-0 and level-l DFD. 12](#_Toc95405645)

[Chapter 4: System Design 14](#_Toc95405646)

[4.1. Design 14](#_Toc95405647)

[4.2. Algorithm Details 16](#_Toc95405648)

[Chapter 5: Implementation and Testing 21](#_Toc95405649)

[5.1. Implementation 21](#_Toc95405650)

[5.1.1. Tools Used 21](#_Toc95405651)

[5.1.2. Implementation Details of Modules 22](#_Toc95405652)

[5.2. Testing 22](#_Toc95405653)

[5.2.1. Test Cases for Unit Testing 23](#_Toc95405654)

[5.2.2. Test Cases for System Testing 23](#_Toc95405655)

[5.3. Result Analysis 23](#_Toc95405656)

[Chapter 6: Conclusion and Future Recommendations 24](#_Toc95405657)

[6.1. Conclusion 24](#_Toc95405658)

[6.2. Future Recommendations 24](#_Toc95405659)

[References 26](#_Toc95405660)

# LIST OF FIGURES

[Figure 1: Use Case Diagram of Advanced Health Prediction System 8](#_Toc95322572)

[Figure 2: Time schedule for Advanced Health Prediction System. 10](#_Toc95322573)

[Figure 3: ER-Diagram of Advanced Health Prediction System. 11](#_Toc95322574)

[Figure 4: level 0 DFD of Advanced Health Prediction System 12](#_Toc95322575)

[Figure 5: level-1 DFD of Advanced Health Prediction System. 13](#_Toc95322576)

[Figure 6: System Architecture 14](#_Toc95322577)

[Figure 7: Database Schema 15](#_Toc95322578)

[Figure 8: A dataset having two tags (green and blue) 17](#_Toc95322579)

[Figure 9: Creating lines between the classes 17](file:///C:\Users\admin\Desktop\project%20report1.docx#_Toc95322580)

[Figure 10 : Linear SVM 18](#_Toc95322581)

[Figure 11: Non-Linear SVM 18](#_Toc95322582)

[Figure 12: Adding Z Dimension 19](#_Toc95322583)

[Figure 13: Best Hyperplane in 3-D 19](#_Toc95322584)

[Figure 14: Converting into 2-D with z=1 19](#_Toc95322585)

[Figure 15: Pseudocode of SVM [7] 20](#_Toc95322586)

# LIST OF ABBREVIATIONS

AI: Artificial Intelligence

AUC: Area Under the Curve

ANN: Artificial Neural Networks

CSS: Cascading Style Sheets

DFD: Data Flow Diagram

ER: Entity Relation

HTML: Hypertext Markup Language

IDE: Integrated Development Environment

JS: Javascript

KNN: K-Nearest Neighbors

MAE: Mean Absolute Error

ML: Machine Learning

RMSE: Root Mean Squared Error

ROC: Receiver Operating characteristic

SMO: Sequential minimal optimization

SQL: Structured Query Language

SVM: Support Vector Machines

UI: User Interface

WEKA: Waikato Environment for Knowledge Analysis

# Introduction

## Introduction

Advanced Health Prediction system incorporates the application of supervised machine learning algorithm for the prediction of disease from the symptoms which are given by the users to the system. This system enables the doctor/patient to know about the disease according to the symptoms. The datasets of the diseases produced by the medical industries contain a wealth of hidden information which could be interpreted into useful ones through data mining techniques. The proposed model utilizes the Support Vector Machines (SVM) as the data mining or machine learning algorithm due to its high accuracy than any other algorithms in the field of the general disease prediction system. Machine Learning has become a present trend for attaining the diagnostic results. The Support Vector Machines (SVM) are chosen for predicting various diseases faster. The aim of SVM is to find the best classification function to distinguish between members of the two classes in the training data.

## Problem Statement

The advanced health prediction system is especially designed for the doctors to help them in being more accurate and reliable. As all the people in the world keep their big faith in doctors for treatment, they are humans as well. The doctors can commit mistakes as well which indeed the system aids them in knowing about the disease of the patients.

The efficiency and accuracy of decisions will decrease when humans are put into stress and immense work. Imagine a doctor who has to examine five patient records; he or she will go through them with ease. However, if the number of records increases with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only five records to be analyzed [1]. From the medical point of view, different factors such as valuable chronic diseases, the results of treatment, the current improvements, or conflicting deterioration, subjective and objective evaluation of the patient, and also other valuable data can highly effect on the diagnosis and decision-making in patients’ treatment [2].

* 1. **Objectives**

The main aim of the project is to build a health prediction system through the application of data mining.

The objectives of the project are:

* To develop a web based interface for predicting the disease according to the symptoms of the patients.
* To implement the Support Vector Machines (SVM) as the tool or algorithm for prediction of the disease relevant to the input of the doctors.
* To help the doctors in reducing their efforts and making them more reliable in decision making.
* To provide the platform for two way communication between patient and doctors for consultation

## Scope and Limitations

The project aims in providing a web platform for predicting the occurrences of disease with regards to various symptoms. The scope of advanced health prediction system is to have a clinical decision making in diagnosis of the patients. The system could reduce the errors that the human or the doctors would possibly make in any phase of their medical error. The doctor would get full assist from the system and help them enhance the safety of the patients. The system would be promising as the data modeling and analysis tools or algorithm, for instance, Support Vector Machines has the potential of generating a knowledge-rich environment which can aid significantly in improving the quality of making quantifiable decision.

Though the project offers benefits to the doctors and health professionals, there are some limitations which are discussed below:

* The system doesn’t recommend the medication of the disease.
* The study doesn’t encompass the analysis and comparison of all the existing data-mining and machine algorithm.

## Development Methodology

The projects or application are considered as successful after well management. There are numerous software development methodologies to choose from but the project can be managed efficiently only upon the best selection of methodology which would fit the project at hand. Every methodology has its own set of advantages and disadvantages, and it exists for a variety of reasons. The project has been implemented following the agile software development methodology. Considering the wider benefits and the features offered by agile, it has been chosen for application development. While developing the application through agile methodology, core focus are given upon the people, prototypes, collaboration and iteration.

## Report Organization

Chapter 1: This chapter explains the introduction, problem statement, objectives, scope and limitation and development methodology of the system.

Chapter 2: This chapter covers all description of fundamental theories, general concepts terminologies and literature review related to the project.

Chapter 3: System analysis of Advanced Health Prediction System project is explained in detail with all the requirements analysis including functional and non-functional requirements, feasibility analysis and structural approach diagrams.

Chapter 4: Structural design approach and details of algorithm used in advanced health prediction system are described in this chapter.

Chapter 5: Process of implementation and testing are described along with all the tools used for the development.

Chapter 6: Conclusion and future recommendation of the system are explained.

# Background Study and Literature Review

## Background Study

A supervised machine learning method, the support vector machine (SVM) algorithm [3], has demonstrated high performance in solving classification problems in many biomedical fields, especially in bioinformatics [4]. In contrast to logistic regression, which depends on a pre-determined model to predict the occurrence or not of a binary event by fitting data to a logistic curve, SVM discriminates between two classes by generating a hyperplane that optimally separates classes after the input data have been transformed mathematically into a high-dimensional space. Because the SVM approach is data-driven and model-free, it may have important discriminative power for classification, especially in cases where sample sizes are small and a large number of variables are involved (high-dimensionality space). This technique has recently been used to develop automated classification of diseases and to improve methods for detecting disease in the clinical setting.

For analyzing or working with the dataset, Support Vector Machines (SVM) is chosen for predicting various diseases and helping the health professionals in diagnosing faster. The SVM have gained very much admiration in the field of data mining and pattern classification. The aim of SVM is to find the best classification function to distinguish between members of the two classes in the training data. The SVM approach is specifically selected due to its various benefits and the essential advantage being that it can overawe ‘the high dimensionality problem’ with a comparative ease. The data-driven nature, quality of generalization and knowing its ability to work with the datasets having many attributes make it an approach far beyond the capacities of other traditional methods. A final goal was to demonstrate the applicability of the SVM approach by creating a demonstration web-based classification tool.

## Literature Review

The literature review is mainly written for understanding about the project and to know the importance of the topic. It is generally an account of which has been published by the researcher which help in developing own ideas. It provides idea or summary about the previous research conducted on that field.

According to Sujatha, the paper suggests that data mining has become one of the most motivating areas which have grown and become widely popular in health-related organization. The real task or work of data mining involves the extraction of data by automatic and semi-automatic means. There are various areas of mining which mainly include clustering, forecasting, and path analysis [5].

According to Gomathi & Priyaa, described that two different data mining classification techniques were used for the prediction of various diseases and their performance was compared to evaluate the best classifier. An imperative challenge in data mining and machine learning areas is to develop specific and computationally proficient classifiers for medical applications [6].

According to Thiyagaraj & Suseendran, the paper discusses about the data mining techniques for kidney-related disease. Kidney disease is a major issue in low-income countries. 60% of deaths worldwide are because of kidney-related issues with the help of data mining in healthcare frauds and abuses can be detected. It helps physicians to identify the best treatment for a particular disease. It can produce fast analysis reports, operational efficiency and reduce operational cost. There are also some of the disadvantages such as data ownership problems, privacy and security related issues for human data administration etc. Various algorithms are used at different stages of analysis and prediction of disease [7].

The authors of this project, Narander Kumar and Sabita Khatri [8], have researched and made comparisons of different algorithms such as k-NN, Naïve Bayes, Random Forest, J48, using performance measures like ROC, kappa statistics, RMSE and MAE in WEKA tools, and also compared the classifiers on various accuracy measures. The conclusion reached of this research was that Random Forest has better accuracy for chronic kidney dataset that was used.

In this project the authors, Monika Gandhi and Dr. Shailendra Singh [9], have analyzed different data mining algorithms like Naïve Bayes, Neural network and decision tree algorithms for their accuracy on prediction of Heart Disease.

The authors Marija Sultana, Afrin Haider and Md.Shorif Uddin [10], have analyzed algorithms such as K-star, J48, SMO, Bayes Net and Multilayer Perceptron Network using WEKA tools for heart disease prediction dataset. The performance of these datamining techniques in acquired by combination of results of measures such as predictive accuracy, ROC curve and AUC value. The result obtained is the SMO and Bayes network show more optimum result than their other mentioned counterparts.

In this project, the authors Girija D.K, Dr. M.S. Shashidhara and M.Giri [11], make use of Neural networks to make predictions regarding presence of uterine fibroid disease. The experimental results show an accuracy of 98% using the Multilayer perceptron neural network and data mining.

# System Analysis

## System Analysis

### Requirement Analysis

### Functional Requirements

The functional requirements are the requirements that describe what the system does i.e., the functioning or behavior of the system. The functional requirements of the Advanced Health Prediction System are:

* The admin should be able to manage all the users including doctors and patients.
* The doctors should be able to make changes to their personal information if needed.
* The doctors should be able to predict the disease from the symptoms which would be chosen.
* The doctors should have additional dialog box to provide feedbacks.
* The patients should be able to predict the disease from the symptoms.
* The system should suggest the patient to whom he/she needs to consult based on predicted disease.
* The system should suggest the best hospitals to the patients based on the predicted disease.

**Use Case Diagram**

The use case diagram of advanced health prediction system using data mining comprises of all the aspects that a normal use case requires. This use case diagram illustrates how the model flows from one phase to another. The actors are the doctors and use cases are the functions operated by them. The use case diagrams represents the system requirements from user’s perspective.

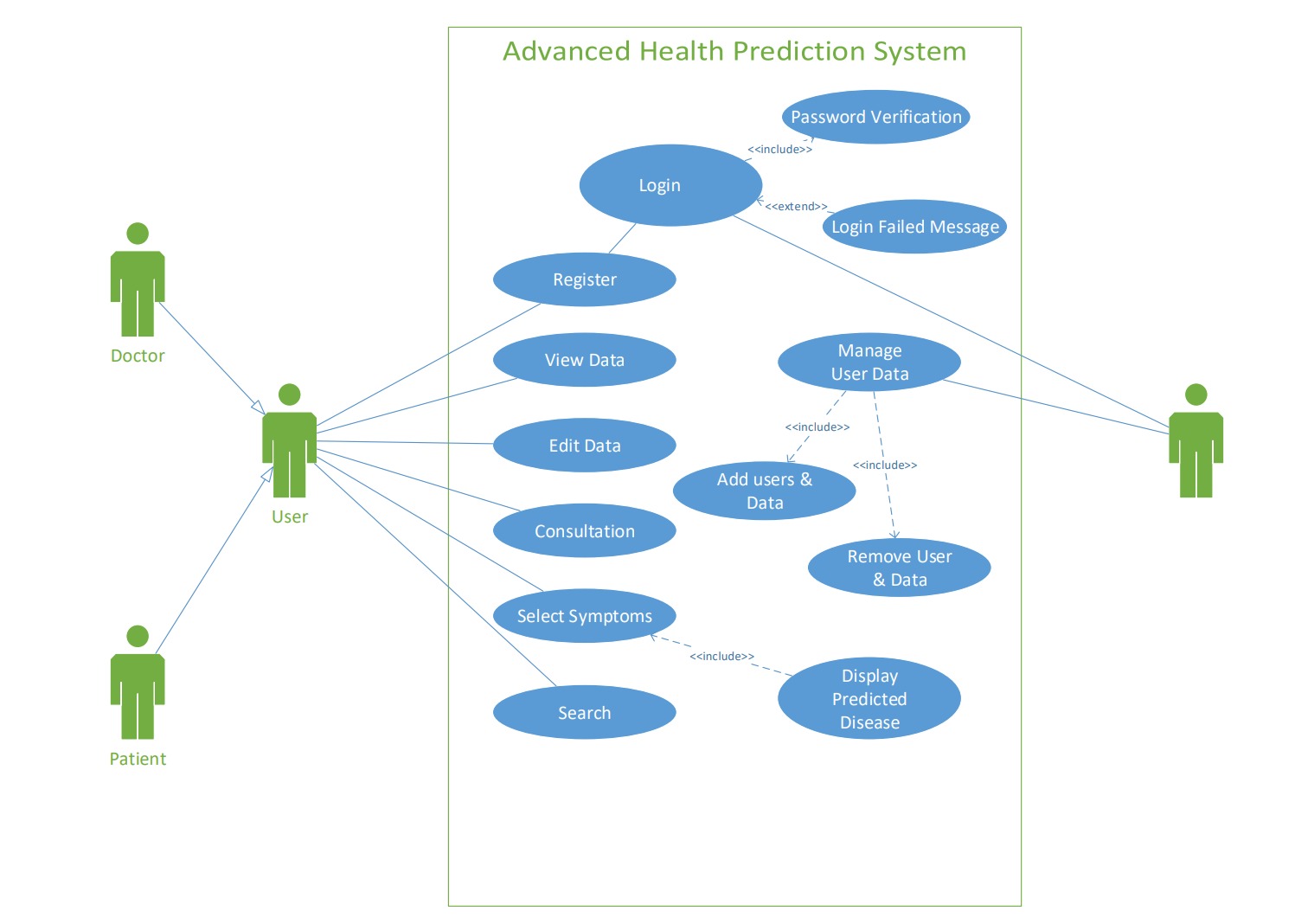


Figure 1: Use Case Diagram of Advanced Health Prediction System

### Non Functional Requirements

It defines the criteria according to which the system must work. The requirements that are not covered by functional requirements are covered by non-functional requirements. It includes following requirements:

Security: Users will only be able to access the system using authorized username and Password. Only the authorized user can change their profile content or password.

Usability: This system is very easy to use and is easy to understand. It is user friendly.

Performance: The developed system is able to provide correct and accurate functioning. Every feature presented by the system has quick response time.

Maintainability: The system needs to be maintained. Updating database, monitoring and error handling should be carried out on a daily basis.

Reliability: Users will be able to get the information about the activities of the organization. Each Registration for the course by the user will be notified to the admin.

### Feasibility Study

Feasibility study means to analyze and evaluate the proposed project that indicates if a project is viable or not. It estimates the existing business environment, problems and opportunities, and resources required which will ultimately lead to success of the project. A feasibility study is an analysis of how successfully a project can be completed, accounting for factors that affect such as economic, scheduling and technical factors. In parallel, the study also determines if the system can be built correctly and precisely on time with available resources meeting all the constraints.

Following feasibility analysis have been performed prior to working on this project:

### Technical Feasibility

Technical Feasibility study defines how feasible a system is from a technical point of view. The developed system is technically feasible. It provides the technical guarantee of accuracy, reliability and security. The work for the project “advanced health prediction system” is done with the current equipment and existing software technology. It needs an IDE.

### Operational Feasibility

This project is operationally feasible as mentioned in the log report which is included in the appendix. Operational feasibility is related to the operational capabilities of the system. The user requirements have been taken into consideration. So, there is no question of resistance from the users that can undermine the possible application benefits.

### Economic Feasibility

Economic Feasibility study is the study that determines whether a system is economically feasible or not. The system is economically feasible. It does not require any additional hardware or software as the interface for this system is developed using the existing resources and technologies available and software used are freely available with free of cost.

### Schedule Feasibility

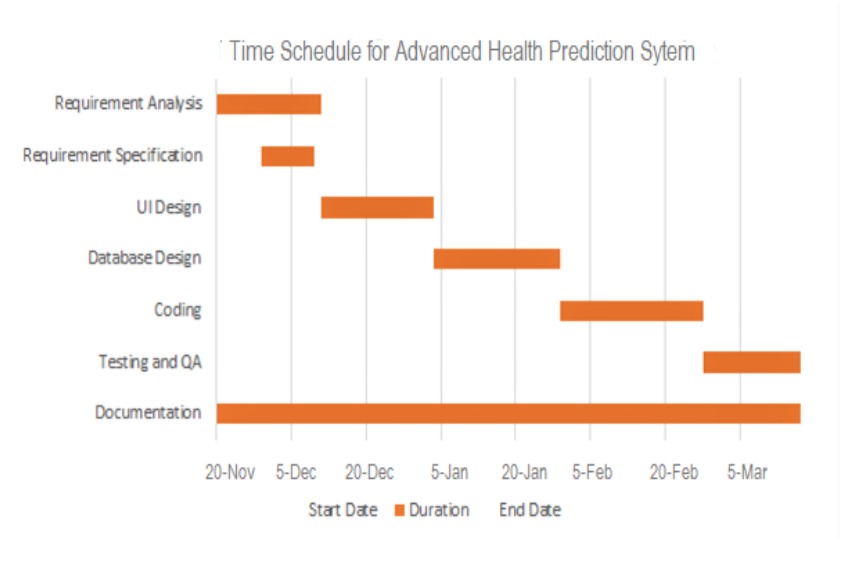


Figure 2: Time schedule for Advanced Health Prediction System.

### Analysis

### Data Modelling using ER Diagram

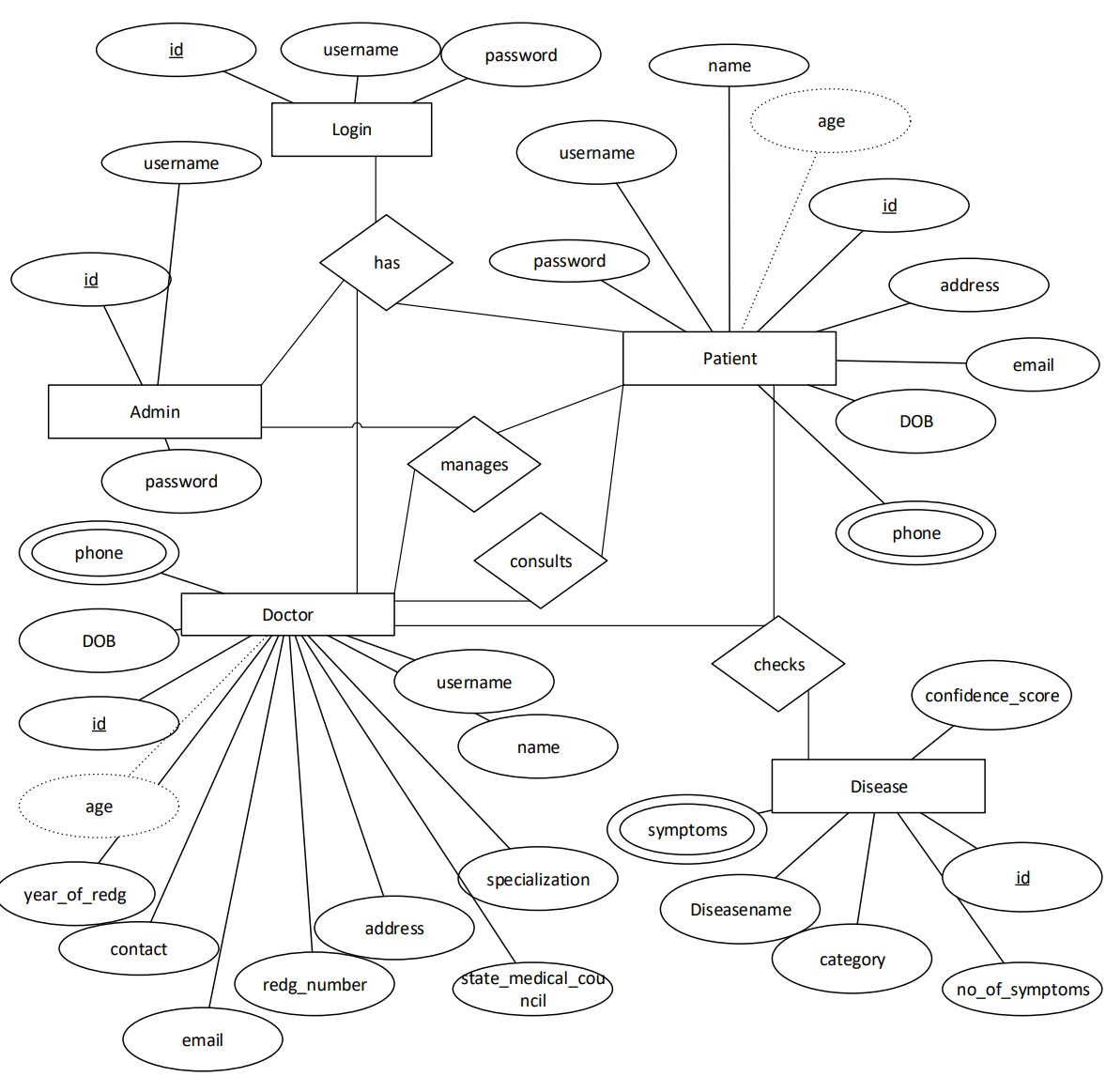


Figure 3: ER-Diagram of Advanced Health Prediction System.

### Process Modelling using level-0 and level-l DFD.

1. Level 0 DFD

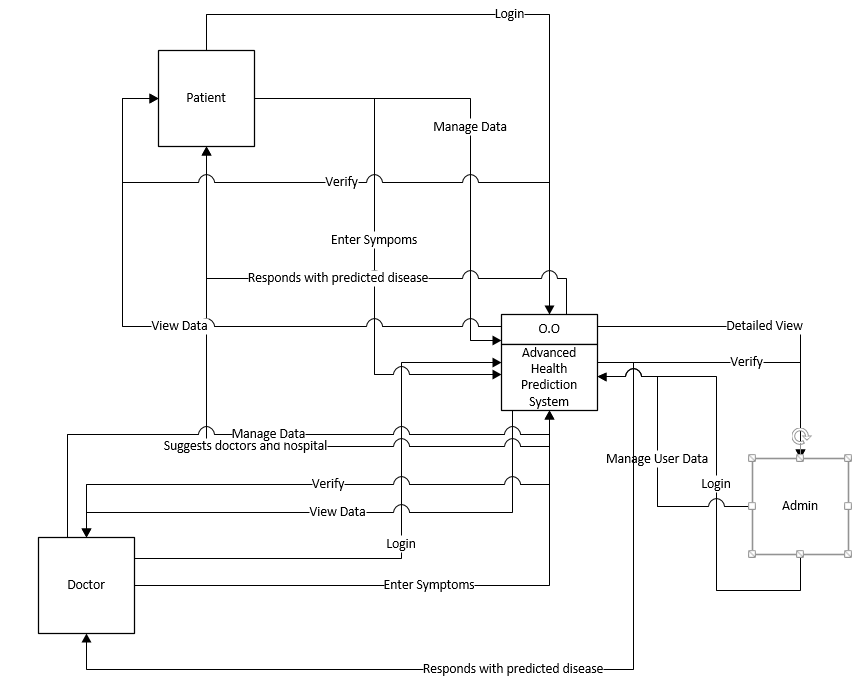


Figure 4: level 0 DFD of Advanced Health Prediction System

1. **Level-1 DFD**

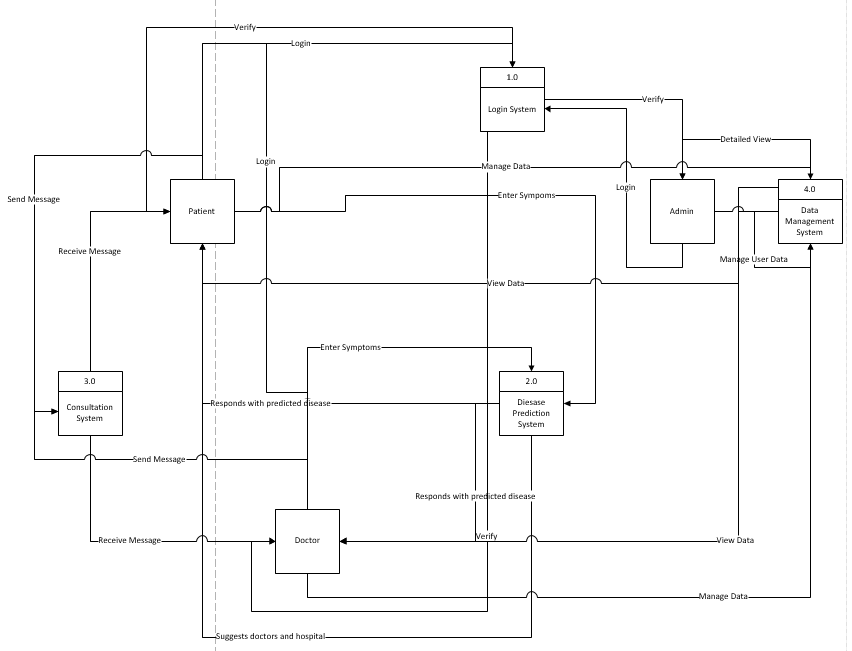


Figure 5: level-1 DFD of Advanced Health Prediction System.

# System Design

## Design

System design is the process of defining the elements of a system such as the architecture, modules and components, the different interfaces of those components and the data that goes through that system. It is meant to satisfy specific needs and requirements of a business or organization through the engineering of a coherent and a well-running system. The advanced health prediction system predicts the disease of the patients for the doctors based on the symptoms or the general information doctors provide to the system through the symptoms. The architecture of the health prediction system encompass various datasets, the datasets are changed into smaller units and they are classified based on data mining algorithms later on the classified data in then processed into SVM and goes into the prediction model via the inputs of the user.

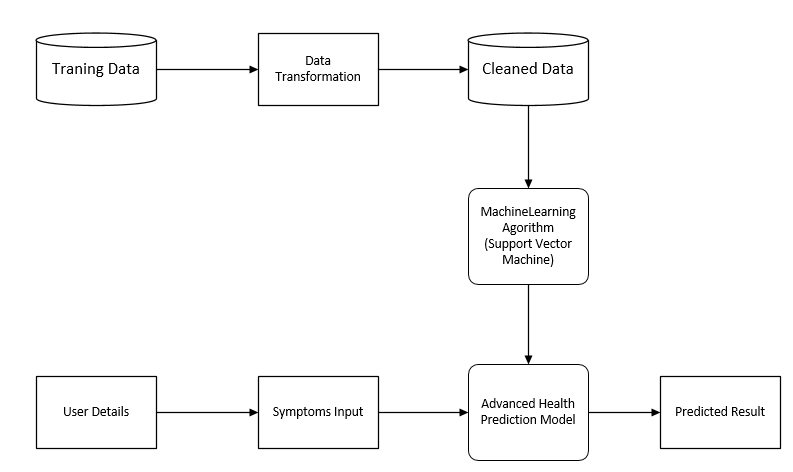


Figure 6: System Architecture

* **Database Design**

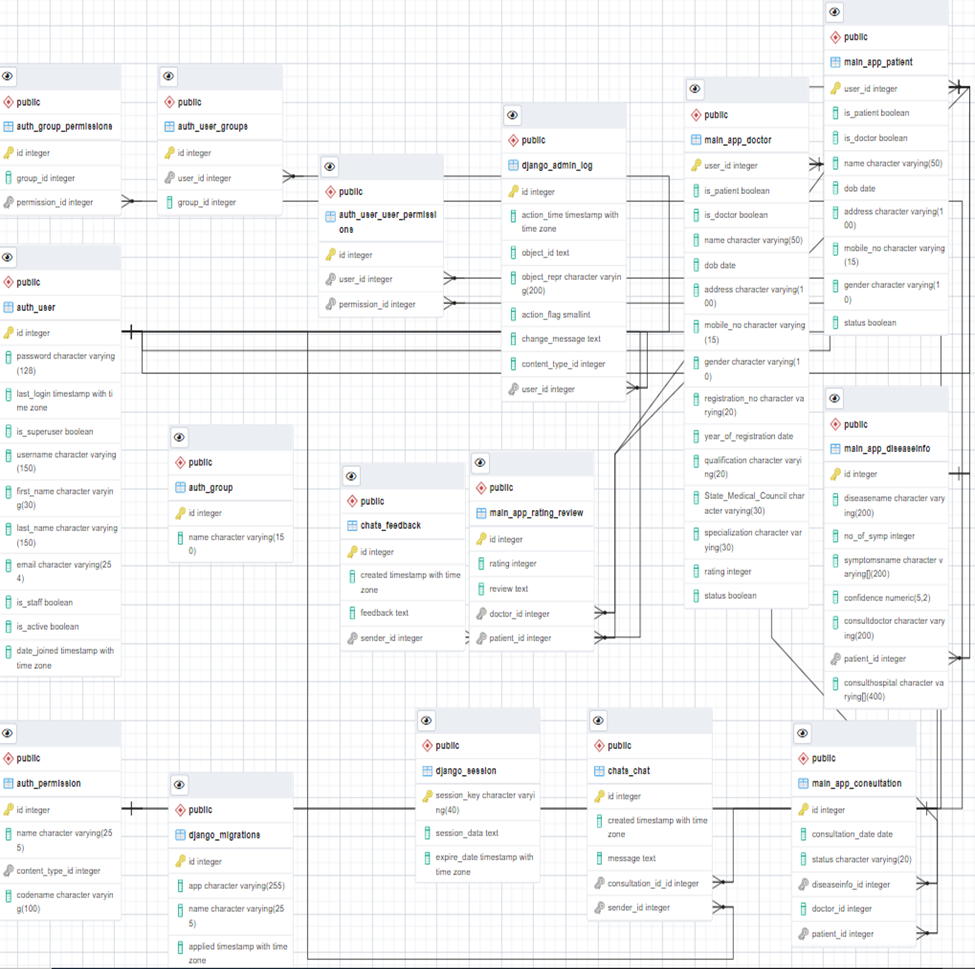
****

Figure 7: Database Schema

* **Forms and Report Design**
* **Interface and Dialogue Design**

## Algorithm Details

The Advanced Health Prediction System is developed using supervised machine learning algorithm, Support Vector Machine. SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Two important terminologies of Support Vector Machine are Hyper-plane and support vector.

Hyper plane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyper plane of SVM.

Support Vectors: The data points or vectors that are the closest to the hyper plane and which affect the position of the hyper plane are termed as Support Vector. Since these vectors support the hyper plane, hence called a Support vector.

SVM can be of two types:

* Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
* Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

**HOW DOES SVM WORKS?**

**Linear SVM:**

The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2. We want a classifier that can classify the pair(x1, x2) of coordinates in either green or blue. Consider the below image:



*Figure 8: A dataset having two tags (green and blue)*

So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:

Figure 9: Creating lines between the classes

Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a hyper plane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyper plane is called as margin. And the goal of SVM is to maximize this margin. The hyper plane with maximum margin is called the optimal hyper plane [12].



*Figure 10 : Linear SVM*

Non-Linear SVM

If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:

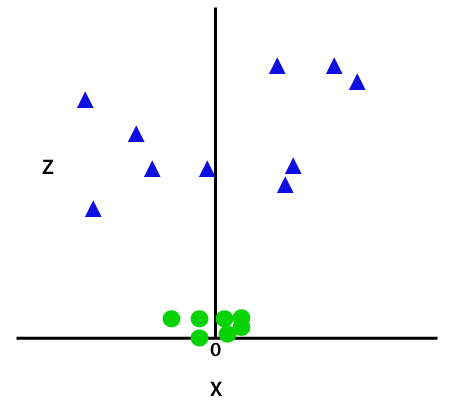


*Figure 11: Non-Linear SVM*

So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

z=x2 +y2

By adding the third dimension, the sample space will become as below image:



*Figure 12: Adding Z Dimension*

So now, SVM will divide the datasets into classes in the following way. Consider the below image:



*Figure 13: Best Hyperplane in 3-D*

Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with z=1, then it will become as:

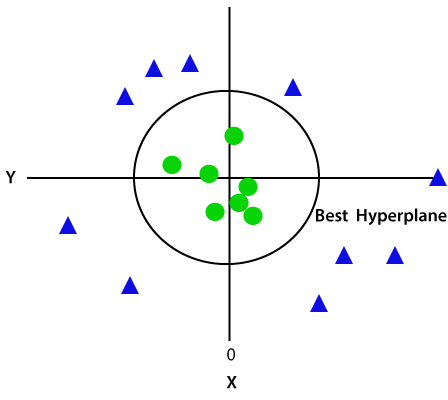
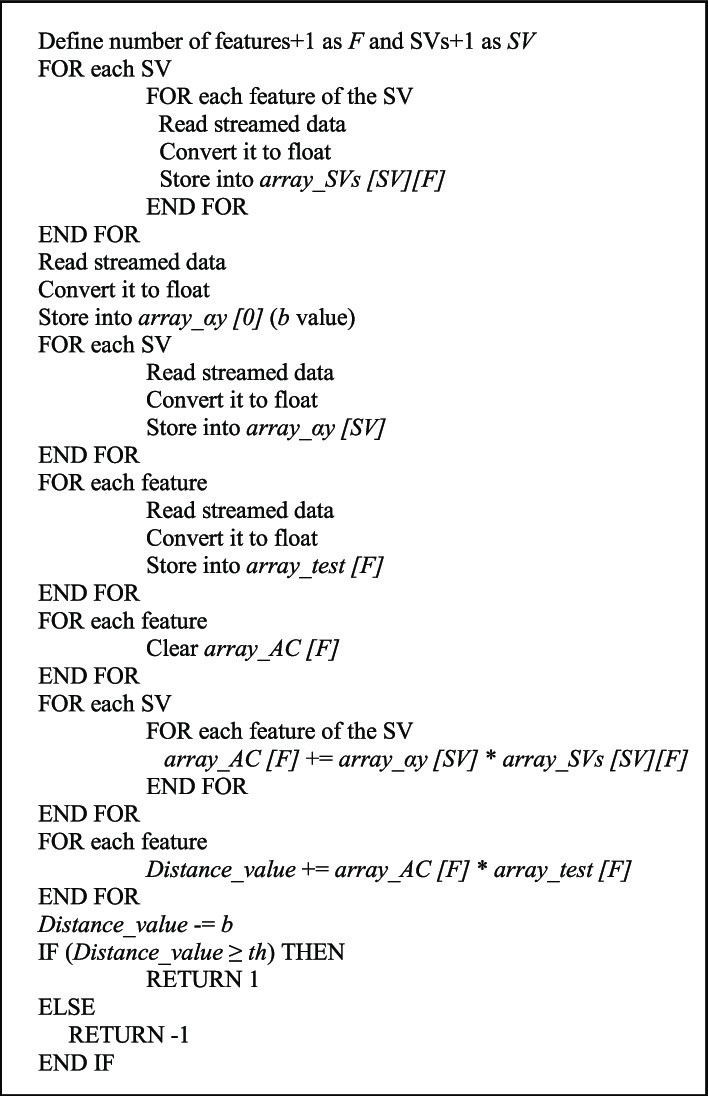


Figure 14: Converting into 2-D with z=1

Hence we get a circumference of radius 1 in case of non-linear data.

**Pseudocode of SVM**



*Figure 15: Pseudocode of SVM* [13]

# Implementation and Testing

## Implementation

The implementation of the logic has been done in an interactive python notebook. Initially, the ipynb file was generated from Jupyter Notebook but further supervision with proper maintenance was conducted through visual studio code. The entire python code have been written in visual studio code which is an amazingly powerful multiplatform text editor allowing the developer to use any of the new features. The goal was to create a prediction model in notebook which would be used in the application as a library for predicting the disease from symptoms.

### Tools Used

**Front End Tool**

The Front-End part of the system would be handled through HTML5, CSS3 and JavaScript making the design look very much attractive and user friendly.

1. HTML5: Hyper Text Markup Language (HTML) is known as the standard markup language which is used for the development and creation of web pages and web applications. HTML code provides an overall framework of how the site will look.
2. ii. CSS3: Cascading Style Sheets (CSS) is a language that is used to describe the style of an HTML document. It represents how HTML elements must be displayed. The application system would get proper decorations only through CSS and would specify how it would documented to the user. Furthermore, the framework of CSS implemented would be Bootstrap version 4.3.1 along with some external CSS for it.
3. iii. JavaScript: The JS can be used to manipulate the system in response to events. The html pages gets enhanced due to the use of JavaScript and the system gets rendered in an interactive and dynamic fashion. The JavaScript gets implemented with the help of its library jquery.

**Back-End Tool**

The Back-End part of the system would be controlled through Python with Django as the framework for it.

1. Python and Django: Django is considered as a high-level python web framework which predominantly helps in developing the web application. Django aids in the elimination of repeated processes, making the development process simple and time-saving. The use of Python with Django will be the core for implementation of application. The python gets implemented with the Django version 3.0.3.

Database Platforms

Database portion of this system is handled through the implementation of PostgreSQL 13.5 and pgAdmin 4. PostgreSQL, also known as Postgres, is a free and open-source relational database management system emphasizing extensibility and SQL compliance. pgAdmin 4 is a complete rewrite of pgAdmin, built using Python and Javascript/jQuery. A desktop runtime written in NWjs allows it to run standalone for individual users, or the web application code may be deployed directly on a web server for use by one or more users through their web browser. The software has the look and feel of a desktop application whatever the runtime environment is, and vastly improves on pgAdmin III with updated user interface elements, multi-user/web deployment options, dashboards, and a more modern design.

### Implementation Details of Modules

## Testing

Testing is the process of determining whether the system works effectively and efficiently. Testing does not only include debugging. It also checks for quality assurance, validation and verification, reliability and availability estimation. Testing tracks the progress of the project and build a confidence in the developer. Different kinds of testing that will be carried out for the proper development of application are illustrated below:

* Unit Testing:

The Unit Testing deals with the functional correctness of the system and each unit of the systems will be separated for identifying, analyzing and fixing the flaws. It focuses on the smallest unit of software design. In this, we test an individual unit or group of interrelated units. The unit testing includes those features which are vital to the performance of unit which is under test. This encourages the developers to amend the source code without immediate concerns about how such modifications might disturb the functioning of other units. This testing will be done after the completion of a distinct unit before integration. It is often done by the programmer by using sample input and observing its corresponding outputs.

* System Testing:

It is the kind of testing which is done to a complete integrated system. The whole interaction of the components of the system will get tested validating that the system works as per the requirements. There will be a functional testing performed under this testing process where the overall functioning of the system will be verified to perform as per the requirements of the user. This software is tested such that it works fine for the different operating systems. It is covered under the black box testing technique. In this, we just focus on the required input and output without focusing on internal working. In this, we have security testing, recovery testing, stress testing, and performance testing.

### Test Cases for Unit Testing

### Test Cases for System Testing

## Result Analysis

# Conclusion and Future Recommendations

## Conclusion

Machine learning can have a huge impact in the field of medical and technical science. The wonders that data mining techniques and algorithms could bring in the field of medical would be quite fascinating to watch. The work for the doctors would far or less get reduced. The advanced health prediction system would be a kind of achievement in the health care sector as the doctors would be using it for the prediction of general diseases through the symptoms of the patients. The implementation of SVM as the algorithm is supposed to bring more accuracy in prediction than any other technique. In addition, the SVM was preferred due to its efficiency in the medical field and previous research work done in the technique. The accuracy obtained from using the SVM as the algorithm was round to 93% of the whole accuracy. Various secondary data collection and primary data collection techniques were taken as a source for supporting the research project. Many research papers and previous works were checked. Furthermore, questionnaires were prepared and distributed to various health authorities and doctors for understanding their views regarding the system. Majority of the results indicated that the system might be a cause for improving the medical system and making the doctors or human gods more reliable. In conclusion, advanced health prediction system has potential in fulfilling the requirements of the doctors and might be a good source in prediction of diseases through symptoms.

## Future Recommendations

The tech world is ever changing and evolving. In the field of Technology, it can never be said that the system is a mere perfect and no enhancements can be made. The field of technology has so much depth in it that there would be advancements every single day. Therefore considering the improvements, the data mining would get progressed as well mining various knowledge concealed in the medical data. The improvements that can be thought out in future are:

* The system can be developed for the public user as well by improving the security system and addition of essential features.
* History about the disease for a user can be kept as a log and recommendation can be implemented for medications.
* There is no recommendation of medication of the disease which can be implemented in the project.
* Medical report analysis of patients can be included for better prediction result and medication.
* The project is a web application which can be made a mobile application as well.

# References

|  |  |
| --- | --- |
| [1] | S. G. H. &. S. R. Afifi, "A system on chip for melanoma detection," *ELSEVIER,* pp. 57-68, 2018. |
| [2] | Ambiga, R., Hariraman, S. & Anantharaj, B., DEEP LEARNING IN SMART, 2018. |
| [3] | C. C and V. V, "Machine Learning," *Support Vector Machine,* vol. 20, pp. 273-297, 1995. |
| [4] | K. Ng and M. SK, "Bioinformatics," *De novo classifaication of precursor microRNAs from genomic pseudo harpins using global and intrinsic folding measures,* vol. 23, 2007. |
| [5] | S. K. &. S. H. Dehkordi, "Prediction of disease based on prescription using data," *Tehran: Springer,* 2018. |
| [6] | K. Gomathi and P. S., "Multi Disease Prediction using Data Mining Techniques," 2017. |
| [7] | M. &. S. G. Thiyagaraj, "Research of Chronic Kidney Disease based on," *IJRTE,* vol. 8(2S11), pp. 115-120, 2019. |
| [8] | N. Kumar and S. Khatri, "Implementing WEKA for medical data classification and early disease prediction," *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT),* p. 5, 2017. |
| [9] | Gandhi, M. a. Singh and S. Narayan, "{Predictions in heart disease using techniques of data mining," in *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, 2015, pp. 520-525. |
| [10] | A. Jagtap, P. Malewadkar, O. Baswat and H. Rambade, "Heart Disease Prediction using," vol. 2, no. 2, p. 4, 2019. |
| [11] | D. Girija, M. Shashidhara and M. Giri, "Data mining approach for prediction of fibroid disease using neural networks," p. 5, 2013. |
| [12] | "Javapoint," [Online]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm. [Accessed 04 Feb 2022]. |
| [13] | S. &. N. A. Sujatha, "A Survey of Health Care Prediction Using Data," 2018. |