TRIBHUVAN UNIVERSITY

INSTITUTE OF SCIENCE AND TECHNOLOGY (IOST)

BHAKTAPUR MULTIPLE CAMPUS

Dudhpati-17, Bhaktapur

A

PROJECT REPORT

ON

"ADVANCED HEALTH PREDICTION SYSTEM"

In partial fulfillment of the requirements for the Bachelor's Degree in Computer Science and Information Technology

(Course Code: CSC-404)

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)

BONAFIDE CERTIFICATE

Certified that, this project report entitled "**ADVANCED HEALTH PREDICTION SYSTEM"** is the Bonafide work of **Mrs. Anusha Wagle (15132/074), Mr. Arjun Shrestha (15133/074), Mrs. Nikita Rjal (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"** has been prepared under my supervision by **Mrs. Anusha Wagle, Mr. Arjun Shrestha, Mrs. 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**" submitted by **Mrs. Anusha Wagle, Mr. Arjun Shrestha, Mrs. 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

Supervisor

Department of Computer

Science & IT

Bhaktapur Multiple Campus

…….

External Examiner

.....

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 website called "Online Book Store with Recommendation System". We are also appreciative of the efforts of BSc. CSIT coordinator **Mr. Sushant Paudel** 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 data mining techniques 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 who spent their days and life treating the patients. 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: data mining, datasets, support vector machines (SVM), machine learning, regression analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI: Artificial Intelligence

ANN: Artificial Neural Networks

ER: Entity Relation

FN: False Negative

FP: False Positive

ID3: Iterator Dichotomiser 3

IDE: Integrated Development Environment

KDD: Knowledge Discovery in Databases

KNN: K-Nearest Neighbors

ML: Machine Learning

PHP: Hypertext Preprocessor

RFNN: Recurrent Fuzzy Neural Network

SVM: Support Vector Machines

TP: True Positive

TN: True Negative

UI: User Interface

UCI: University of California Irvine

1 CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Data mining is a powerful technology that uses the data which are already existed in the various databases and alters it into a new result or product. The use of Artificial Intelligence (AI) and database management are done in data mining for extracting new patterns from comparatively larger data sets along with the knowledge which are associated with those patterns. The concrete work is to extract the data through automatic or semi-automatic means. The different parameters included in data mining includes clustering, forecasting, path analysis and predictive analysis. Data mining involves analyzing certain amount of information to locate certain patterns of occurrence to predict future tendencies, using several processes of effective data collection, warehousing and computer processing. With this kind of functionality, it becomes very vital and makes a purpose in terms of predicting the disease of the patients providing the essential health information. These kinds of findings become quite beneficial in healthcare industry as it makes the work of health professionals far more convenient. Eyeing the advancements that have been made in the field of technologies, a smart health prediction system which could foresee the disease is the need of time considering the gains it would provide to the health professionals and doctors.

Advanced Health Prediction system incorporates the application of data mining techniques 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 who spent their days and life treating the patients. 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. Data mining 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.

1.2 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].

1.3 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 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.

1.4 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 smart 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 system is limited to doctors and health professionals only and not made and bound to the public users.
- The study doesn't encompass the analysis and comparison of all the existing data-mining algorithm.

1.5 REPORT ORGANIZATION

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

Chapter 2: This chapter covers all the history, methods, requirement specification and feasibility analysis and structured system requirements.

Chapter 3: Design of Advanced Health Prediction System project is explained in detail with all the necessary diagrams and brief functionality.

Chapter 4: Process of implementation and testing is described along with all the tools used for the development.

Chapter 5: Conclusion and future scope of the application are explained.

2 CHAPTER 2: REQUIREMENT ANALYSIS AND FESIBLITY STUDY

2.1 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 [3].

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 [4].

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 [5].

2.2 ALGORITHM

2.2.1 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine or 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.

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.

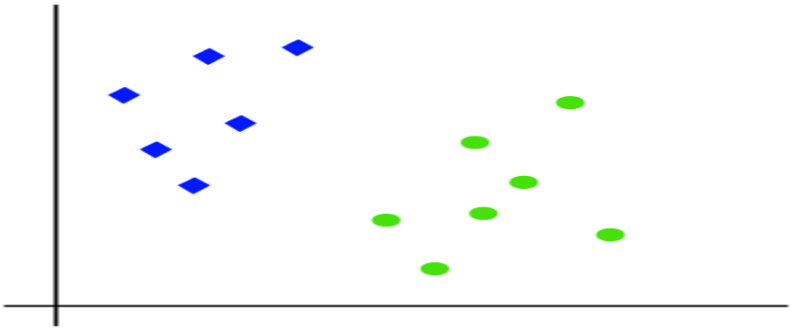
Hyper plane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyper plane of SVM.

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.

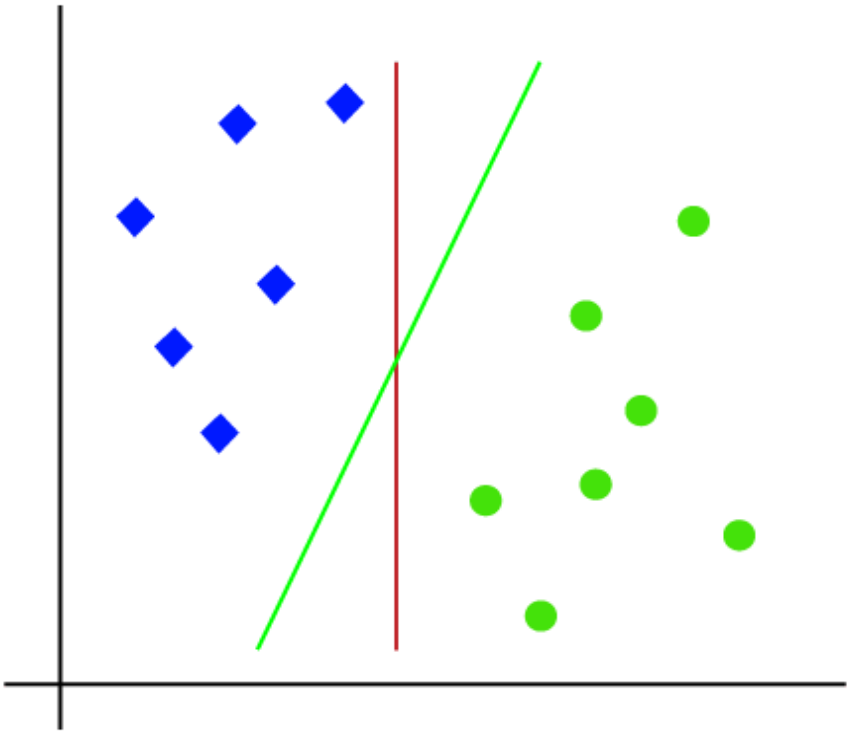
2.2.2 HOW DOES SVM WORKS?

2.2.2.1 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:



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:



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 [6].

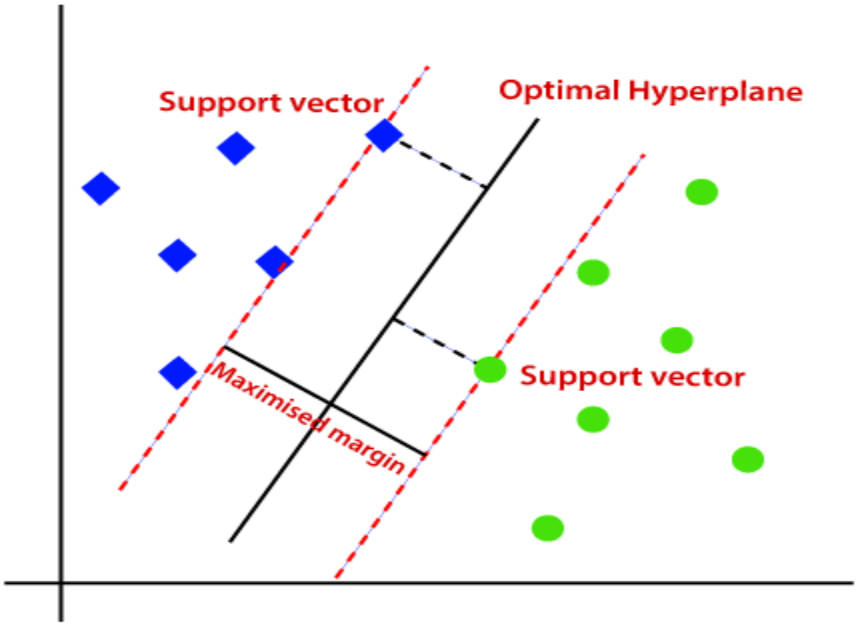
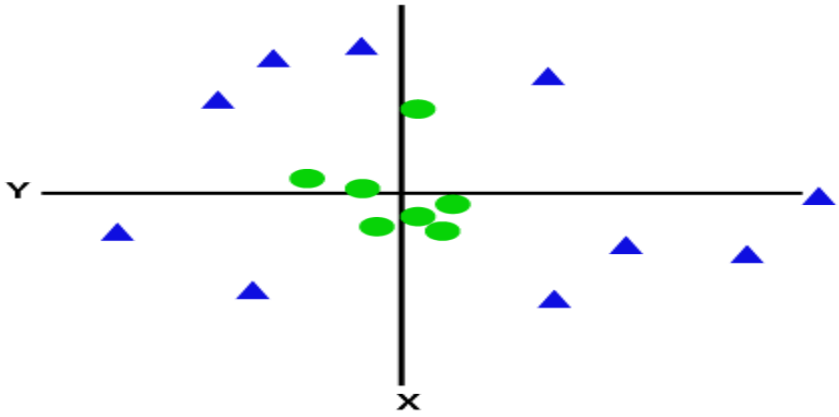


Figure 1 : Linear SVM

2.2.2.2 Non-Linear SVM:

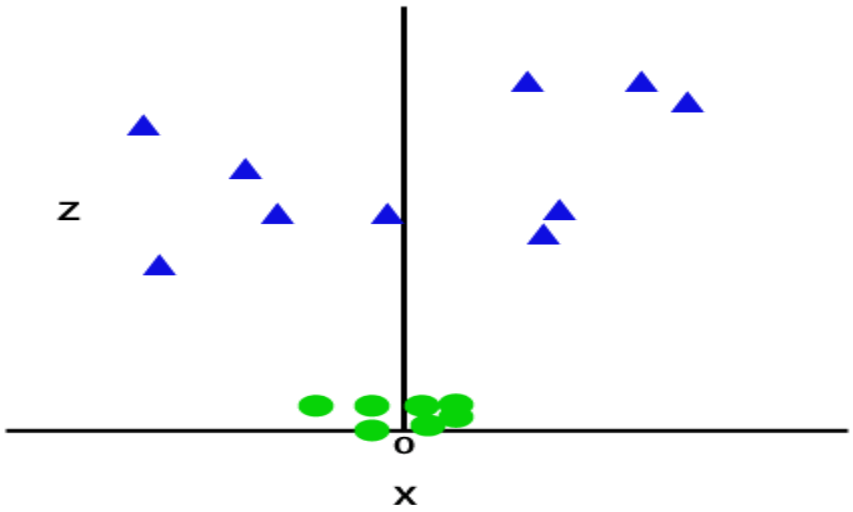
If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:



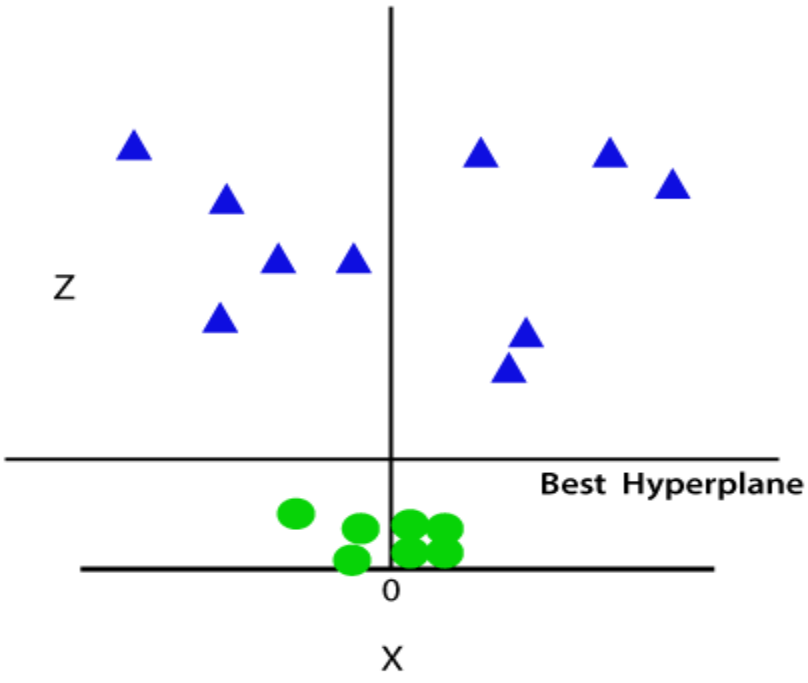
So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

$z=x^2 +y^2$

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



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



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:

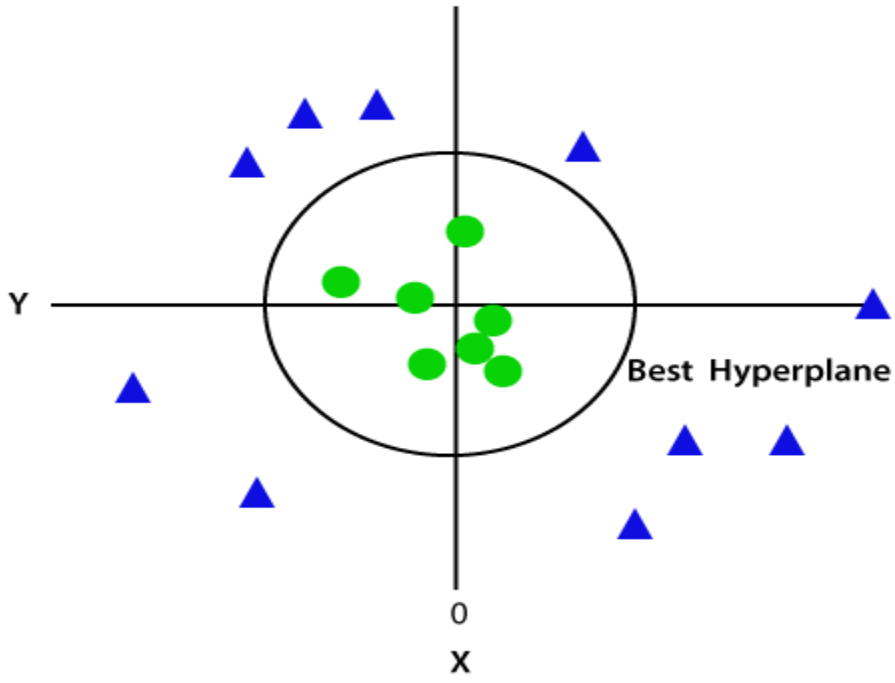


Figure 2: Non-Linear SVM

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

2.3 PSEUDOCODE OF SVM

```
Define number of features+1 as F and SVs+1 as SV
FOR each SV
        FOR each feature of the SV
          Read streamed data
          Convert it to float
          Store into array_SVs [SV][F]
        END FOR
END FOR
Read streamed data
Convert it to float
Store into array_αy [0] (b value)
FOR each SV
        Read streamed data
        Convert it to float
        Store into array_αy [SV]
END FOR
FOR each feature
        Read streamed data
        Convert it to float
        Store into array_test [F]
END FOR
FOR each feature
        Clear array_AC [F]
END FOR
FOR each SV
        FOR each feature of the SV
          array_AC [F] += array_αy [SV] * array_SVs [SV][F]
        END FOR
END FOR
FOR each feature
        Distance_value += array_AC [F] * array_test [F]
END FOR
Distance_value -= b
IF (Distance_value ≥ th) THEN
        RETURN 1
ELSE
   RETURN -1
END IF
```

Figure 3:Pseudocode of SVM [7]

2.4 REQUIREMENT SPECIFICATION

2.4.1 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.

2.4.1.1 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.

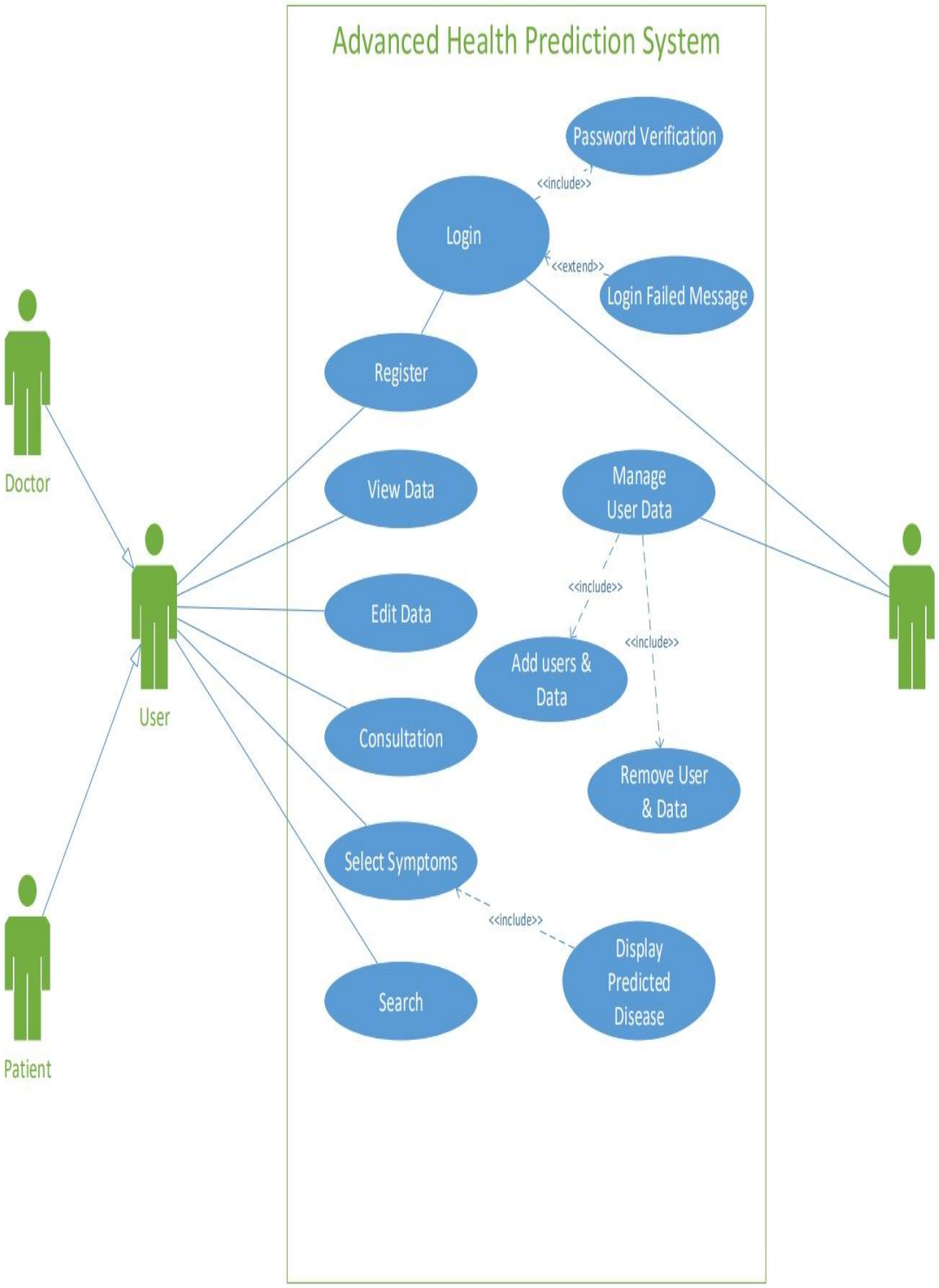


Figure 4:Use Case Diagram

2.4.2 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.

2.5 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 will be performed prior to working on the project:

2.5.1 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 is done with the current equipment and existing software technology. It needs an IDE.

2.5.2 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.

2.5.3 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.

2.5.4 Schedule Feasibility

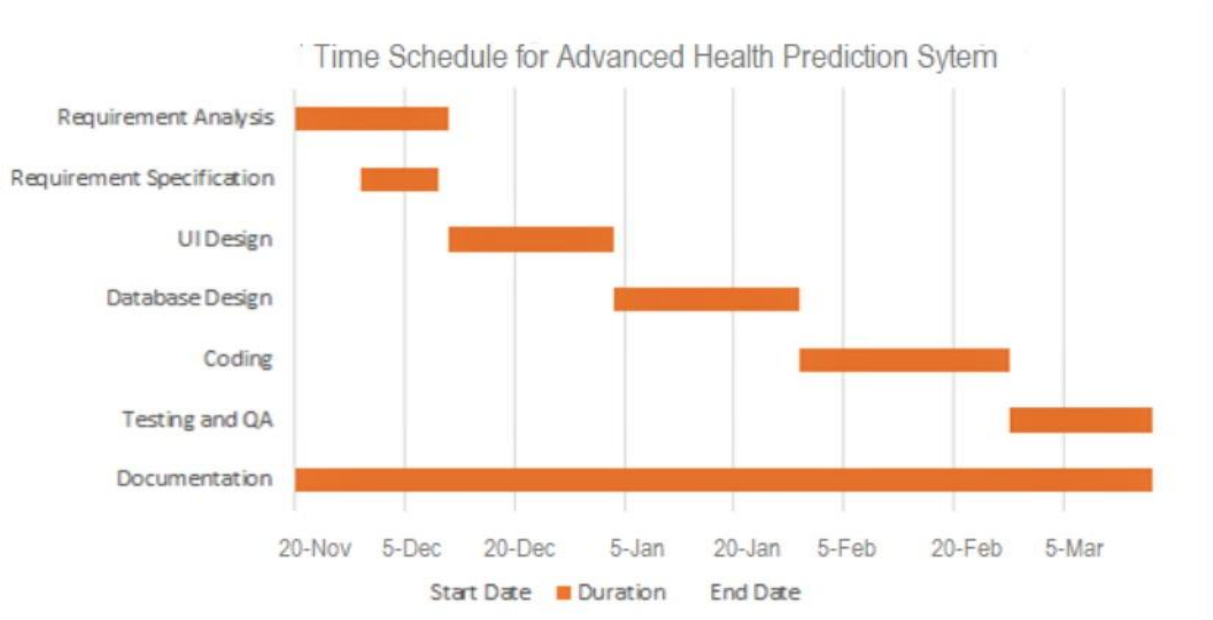


Figure 5: Time schedule for Advanced Health Prediction System

2.6 Structuring System Requirements

2.6.1 Data Modeling (ER Diagram)

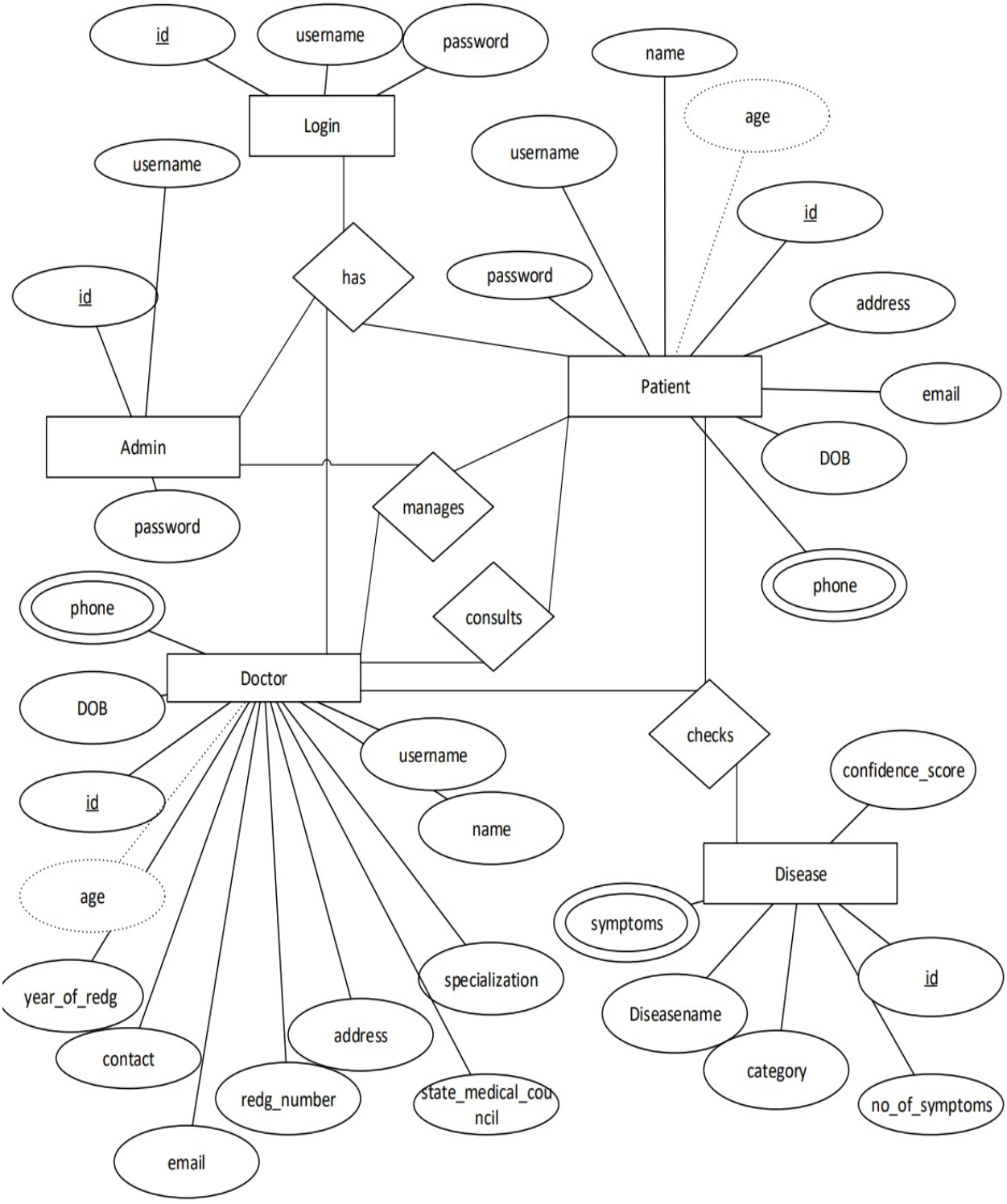


Figure 6: ER-Diagram

2.6.2 Process Modeling (0-level and 1 level DFD)

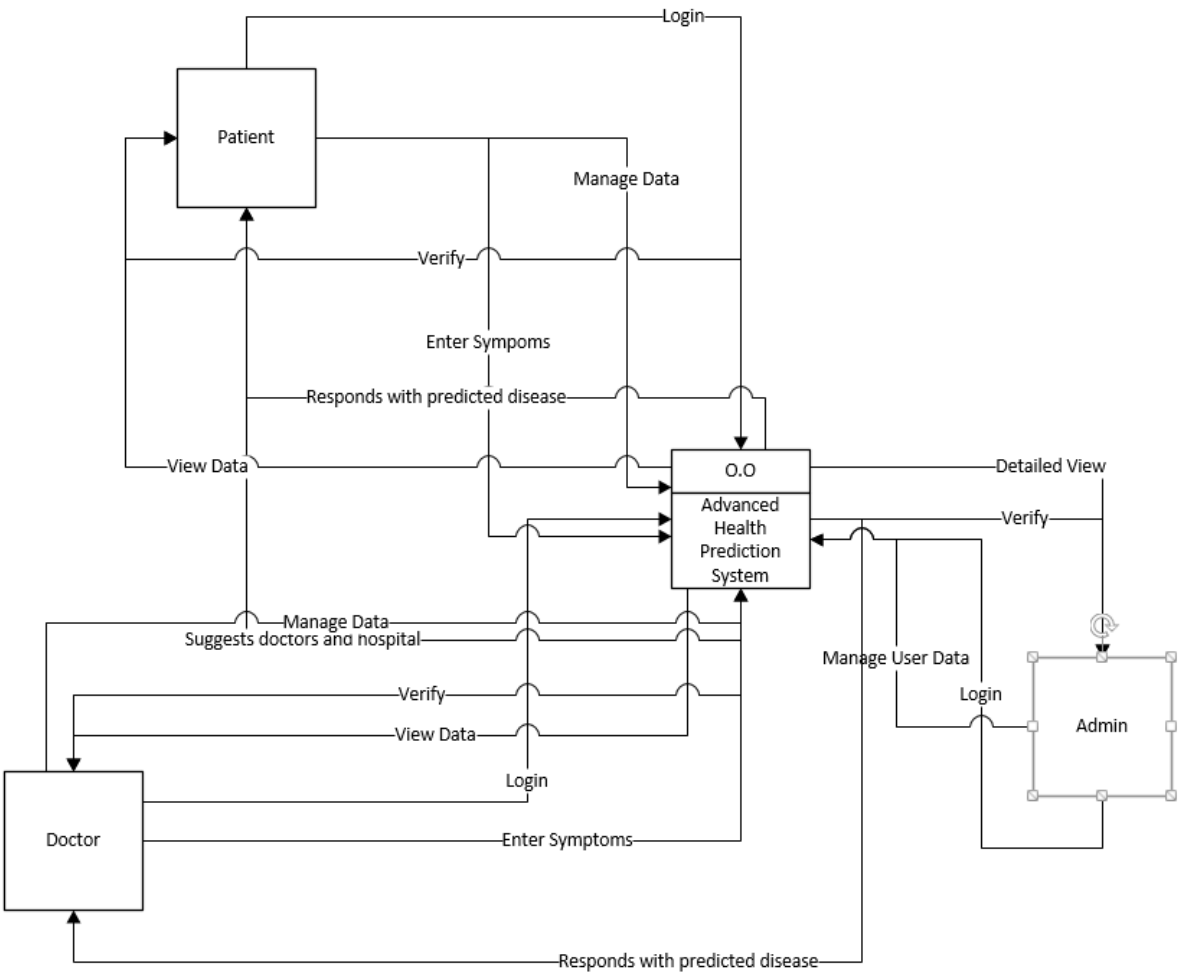


Figure 7: 0-level DFD

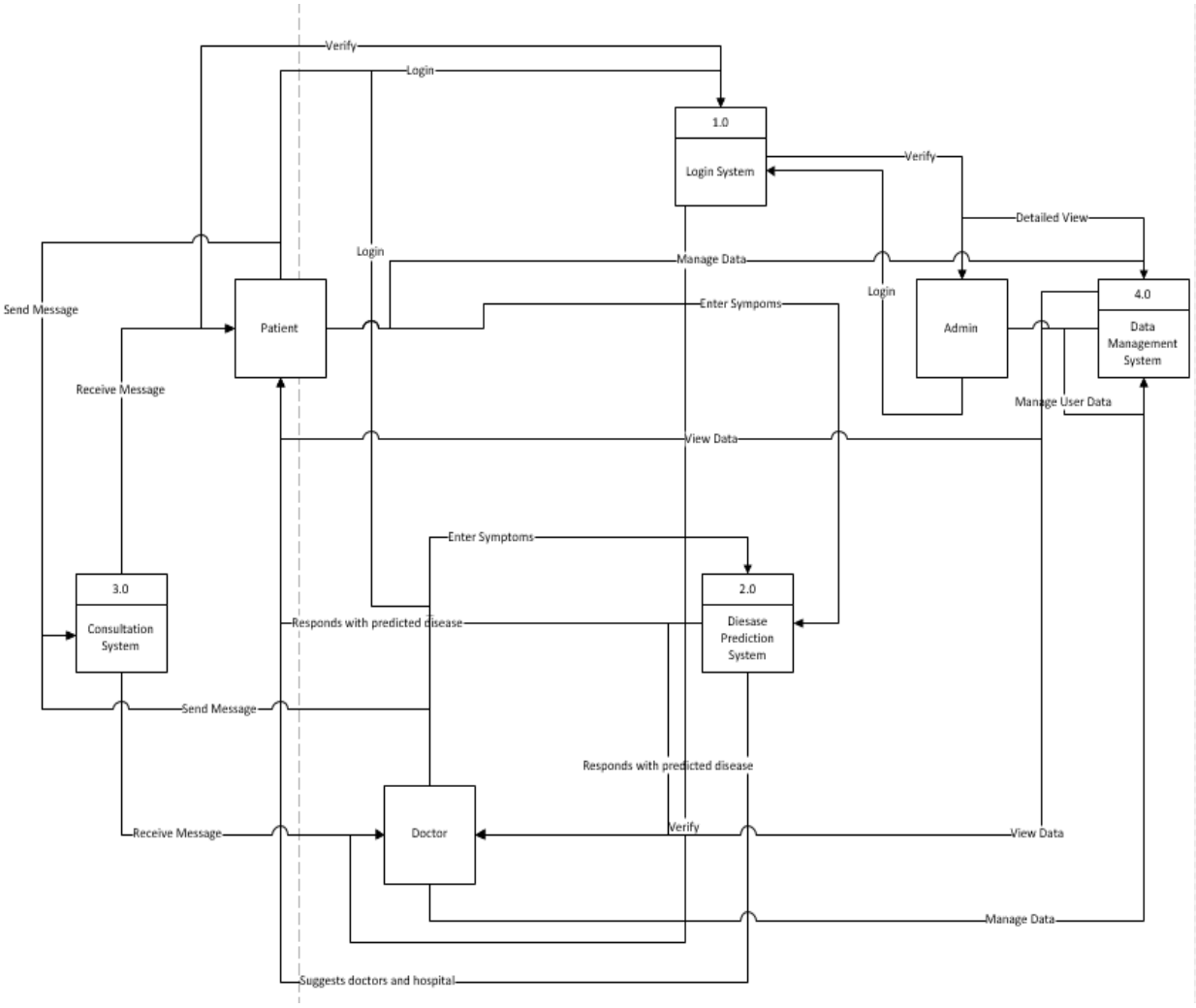


Figure 8:1-level DFD

3 CHAPTER 3: SYSTEM DESIGN

3.1 System architecture and overview

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.

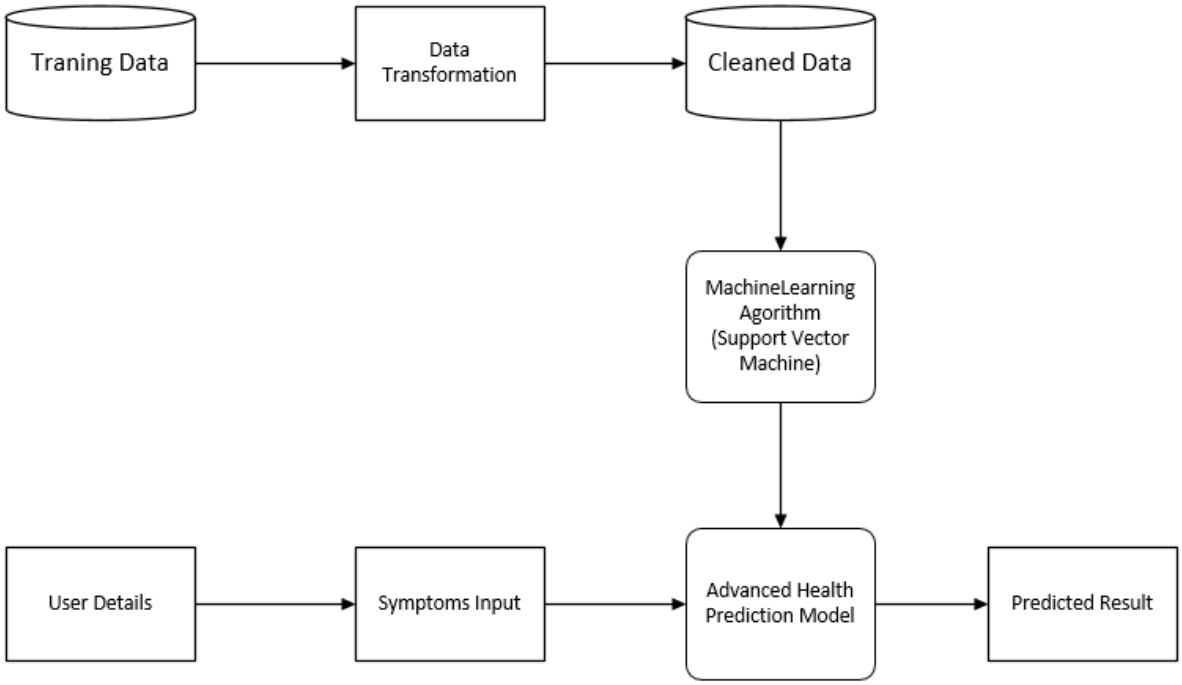


Figure 9: System Architecture

3.2 System Design

3.2.1 Database Scheme

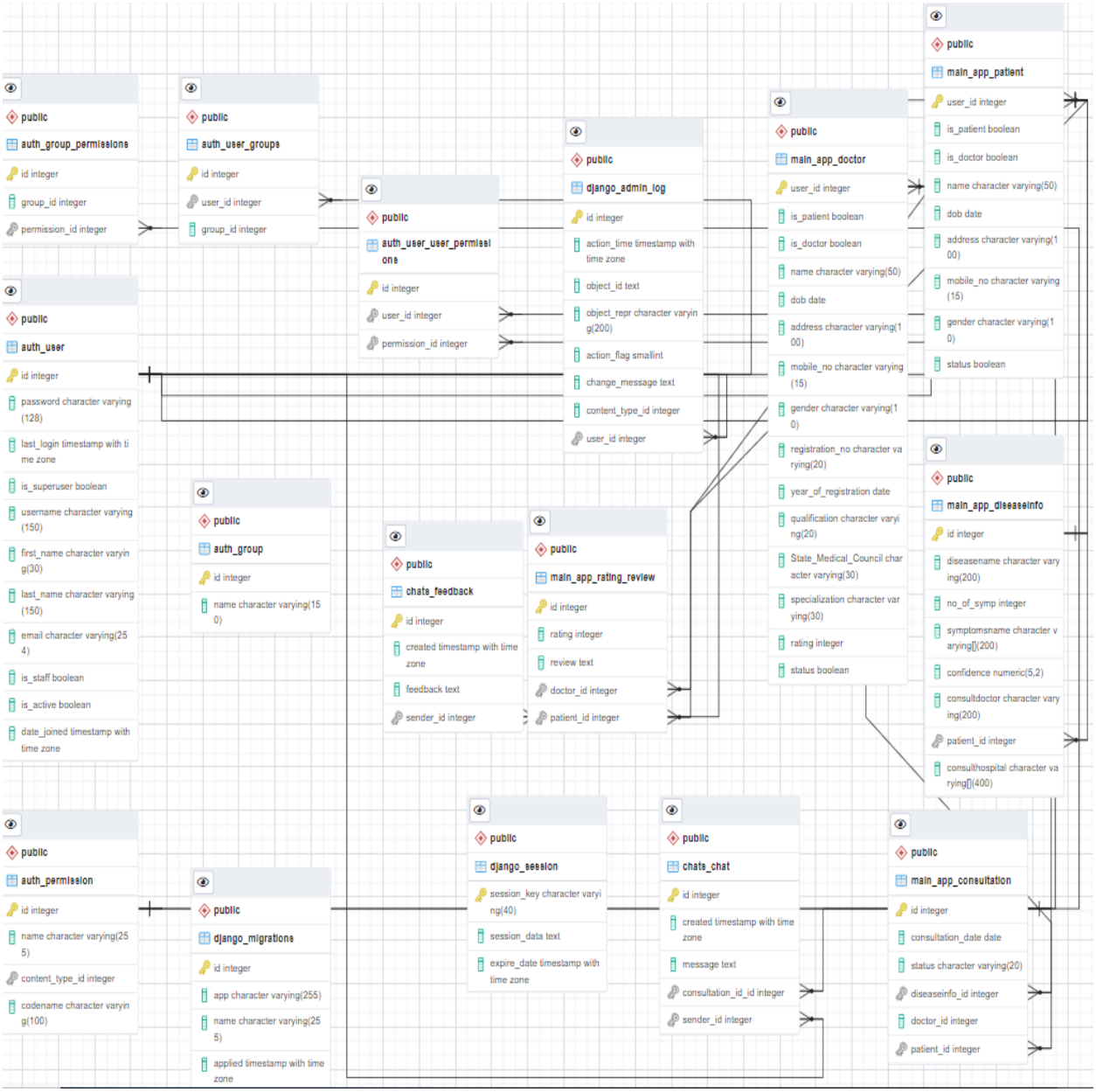


Figure 10: Database Scheme

3.2.2 Class Diagram

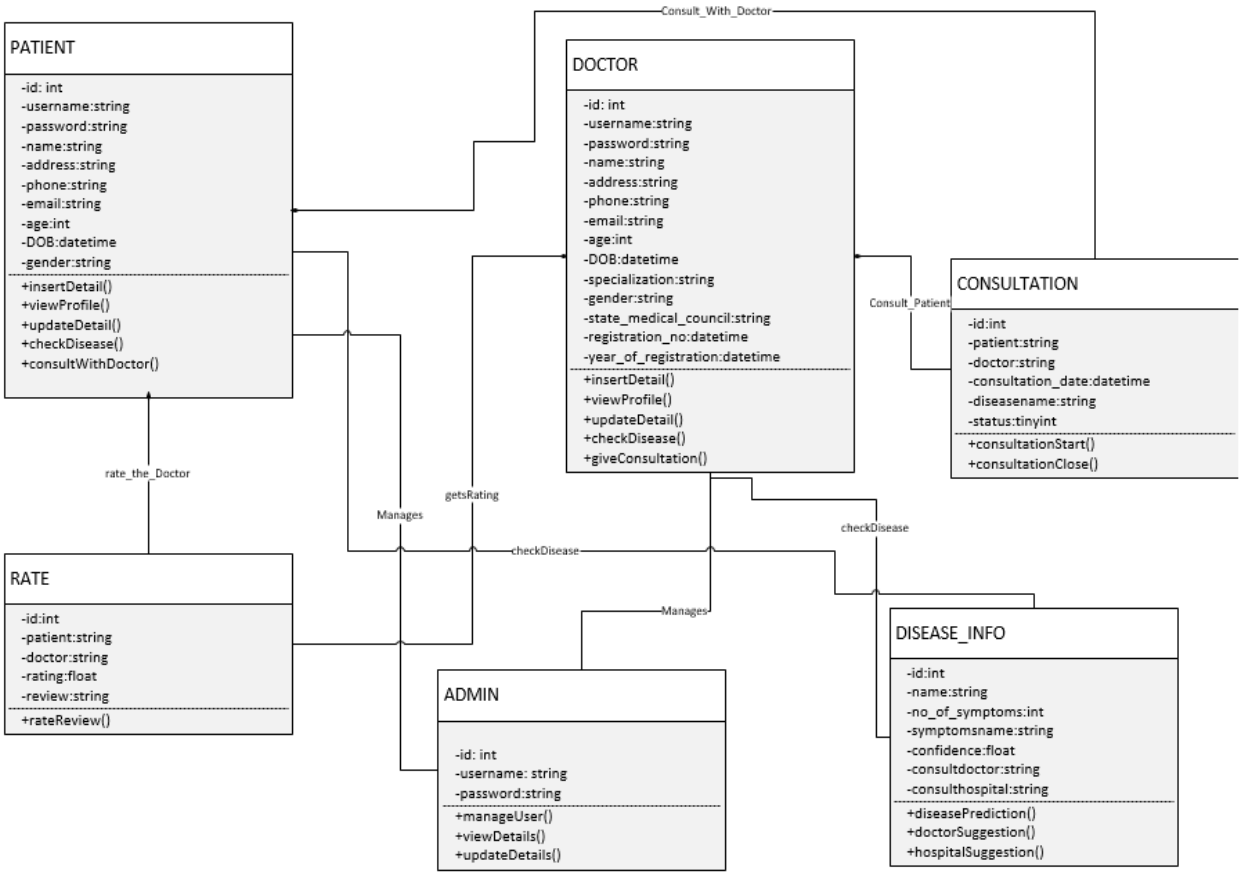


Figure 11: Class Diagram

3.2.3 Activity Diagram

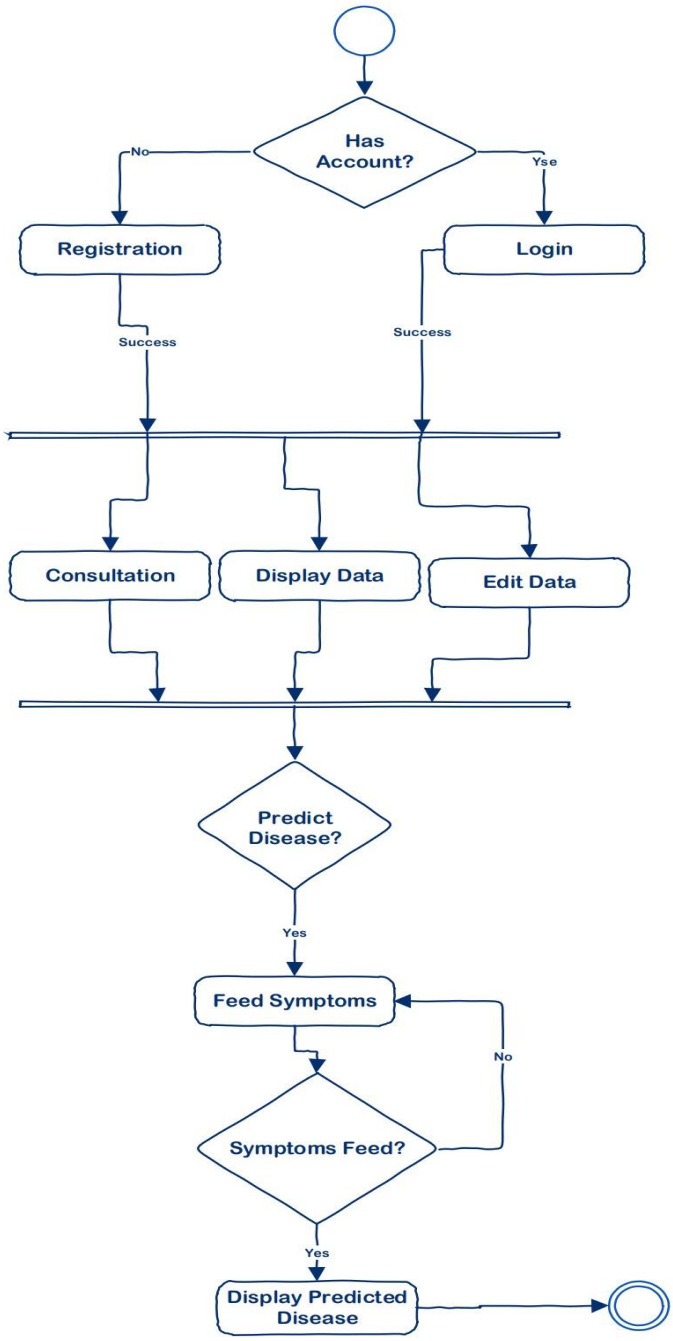


Figure 12: Activity Diagram

4 Chapter 4: SYSTEM IMPLEMENTATION AND TESTING

4.1 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.

4.2 Tools Used

4.2.1 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.

i. 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.

ii. 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.

iii. iii. JavaScript: The JavaScript 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.

4.2.2 Back-End Tool

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

i. 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.

4.3 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.

- **Integration Testing:**

Integration testing is testing in which a group of components is combined to produce output. The Integration Testing helps to verify the functional, performance, and reliability between the modules that are integrated. Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. This testing will

be done after the unit testing and will be used for exposing the problems that might ascend through the combination of components. The objective is to take unit tested components and build a program structure that has been dictated by design.

- **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.

5 CHAPTER 5: CONCLUSION

5.1 Conclusion

Data mining 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.

6 REFERENCES

[1] Ambiga, R., Hariraman, S. & Anantharaj, B., DEEP LEARNING IN SMART, 2018.

[2] S. G. H. &. S. R. Afifi, "A system on chip for melanoma detection," *ELSEVIER,* pp. 57-68, 2018.

[3] S. K. &. S. H. Dehkordi, "Prediction of disease based on prescription using data," *Tehran: Springer,* 2018.

[4] K. Gomathi and P. S., "Multi Disease Prediction using Data Mining Techniques," 2017.

[5] S. &. N. A. Sujatha, "A Survey of Health Care Prediction Using Data," 2018.

[6] M. &. S. G. Thiyagaraj, "Research of Chronic Kidney Disease based on," *IJRTE,* vol. 8(2S11), pp. 115-120, 2019.

[7] "Javapoint," [Online]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm. [Accessed 04 Feb 2022].