**Abstract**

The healthcare domain is one of the prominent research fields in the current scenario with the rapid improvement of technology and data. It is difficult to handle the huge amount of data of the patients. It is easier to handle this data through Big Data Analytics. There are a lot of procedures for the treatment of multiple diseases across the world. Machine Learning is an emerging approach that helps in prediction, diagnosis of a disease. This paper depicts the prediction of disease based on symptoms using machine learning. Machine Learning algorithms such as Naive Bayes, Decision Tree and Random Forest are employed on the provided dataset and pre- dict the disease. Its implementation is done through the python pro- gramming language. The research demonstrates the best algorithm based on their accuracy. The accuracy of an algorithm is determined by the performance on the given dataset**.**

**INDEX**

1. [**Introduction**](#_bookmark0) **1**

[1.1 Overview](#_bookmark1) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[1.1.1 Motivation](#_bookmark2) . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[1.1.2 Objective](#_bookmark3) . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1. [**Literature Survey**](#_bookmark4) **4**

[2.1 Study Of Project](#_bookmark5) . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

1. [**Problem Statement**](#_bookmark6) **15**
2. [**Project Requirement**](#_bookmark7) **17**
   1. [**EXTERNAL INTERFACE REQUIREMENT**](#_bookmark8)18
      1. [**User Interface**](#_bookmark9)18
      2. [**Hardware Interfaces:**](#_bookmark10)18
      3. [**Software Interfaces**](#_bookmark11)18
   2. [**NON FUNCTIONAL REQUIREMENT**](#_bookmark12)19
      1. [**PerformanceRequirements**](#_bookmark13)19
      2. [**Safety Requirement**](#_bookmark14)19
      3. [**Software Quality Attributes**](#_bookmark15)19
3. [**System Analysis**](#_bookmark16) **21**
   1. [**System Architecture**](#_bookmark17)22
      1. [**Module**](#_bookmark19)22
      2. [**Data Flow Diagram**](#_bookmark20)22
   2. [**UML DIAGRAMS**](#_bookmark24)25

**6 References…………………………………………………………………….30**

**List of Figures**

* 1. [**System Architecture-l**](#_bookmark18)22
  2. [**Data Flow diagram**](#_bookmark21)24
  3. [**Data Flow diagram**](#_bookmark22)24
  4. [**Data Flow diagram**](#_bookmark23)25
  5. [**Use case Diagram**](#_bookmark25)26
  6. [**Activity Diagram**](#_bookmark26)27
  7. [**Sequence Diagram**](#_bookmark27)28
  8. [**Class Diagram**](#_bookmark28)29

# Chapter 1

# INTRODUCTION

**1.1 OVERVIEW**

At present, when one suffers from particular disease, then the person has to visit to doctor which is time consuming and costly too. Also if the user is out of reach of doctor and hospitals it may be difficult for the user as the disease can not be identified. So, if the above process can be completed using a automated program which can save time as well as money, it could be easier to the patient which can make the process easier. There are other Heart related Disease Pre- diction System using data mining techniques that analyzes the risk level of the patient. heart Disease Predictor is a web based application that predicts the heart disease of the user with respect to the symptoms given by the user. heart Disease Prediction system has data sets collected from different health related sites. With the help of heart Disease Predictor the user will be able to know the probability of the disease with the given symptoms. As the use of internet is growing every day, people are always curious to know different new things. People always try to refer to the internet if any problem arises. People have access to internet than hospitals and doctors. People do not have immediate option when they suffer with particular disease. So, this system can be helpful to the people as they have access to internet 24 hours.

* + 1. **Motivation**

The current systems working on heart disease prediction works on a small dataset. The aim of our system is to work on a larger dataset to increase the efficiency of the overall system.our system is easy to heart predict disease give output quickly the heart Diesase prediction functioning depends on Natural lan- guage processing that helps users to submit their problem about the health.

* + 1. **Objective**
       - To implement Na¨ıve Bayes Classifier that classifies the disease as per the input of the user.
       - To develop web interface platform for the heart prediction of the disease.
       - to give immediaitily Output to user(heart predict the disease Quickly

# 

# CHAPTER 2 LITERATURE SURVEY

2.1 STUDY OF PROJECT

1. **Paper Name: Design And Implementing Heart Disease Prediction Using Naives Bayesian**

Author:Anjan Nikhil Repaka,Sai Deepak Ravikanti.

Abstract :—Data mining, a great developing technique that revolves around exploring and digging out significant information from massive collection of data which can be further beneficial in examining and drawing out patterns for making business related decisions. Talking about the Medical domain, imple- mentation of data mining in this field can yield in discovering and withdrawing valuable patterns and information which can prove beneficial in performing clinical diagnosis. The research focuses on heart disease diagnosis by consider- ing previous data and information. To achieve this SHDP (Smart Heart Disease Prediction) is built via Navies Bayesian in order to predict risk factors concern- ing heart disease. The speedy advancement of technology has led to remarkable rise in mobile health technology that being one of the web application. The re- quired data is assembled in a standardized form. For predicting the chances of heart disease in a patient, the following attributes are being fetched from the medical profiles, these include: age, BP, cholesterol, sex, blood sugar etc... The collected attributes acts as input for the Navies Bayesian classification for predicting heart disease. The dataset utilized is split into two sections, 80% dataset is utilized for training and rest 20% is utilized for testing. The proposed approach includes following stages: dataset collection, user registration and login (Application based), classification via Navies Bayesian, prediction and se- cure data transfer by employing AES (Advanced Encryption Standard). There- after result is produced. The research elaborates and presents multiple knowl- edge abstraction techniques by making use of data mining methods which are adopted for heart disease prediction. The output reveals that the established diagnostic system effectively assists in predicting risk factors concerning heart diseases.

1. **Paper Name:Application of Machine Learning in Disease Prediction Author: Pahulpreet Singh Kohli,Shriya Arora**

Abstract :The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal dis- eases and therefore, increase the survival rate of patients significantly. In this paper, we apply different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

1. **Paper Name:Disease phenotype similarity improves the prediction of novel disease-associated microRNAs**

Author:Duc-Hau Le

Description : —Many studies have shown roles of miRNAs (microRNAs) on human disease and a number of computational methods have been proposed to predict such associations by ranking candidate microRNAs according to their relevance to a disease. Among them, network-based methods are becoming dominant since they well exploit the “disease module” principle in miRNA func- tional similarity networks. Of which, Random Walk with Restart (RWR) algorithm- based method on a miRNA functional similarity network, namely RWRMDA,

is state-of-theart one. The use of this algorithm was inspired from its success in predicting disease gene because “disease module” principle also exists in pro- tein interaction networks. Besides, many other algorithms were also designed for prediction of disease genes. However, they have not yet been utilized for disease microRNA prediction. In this study, we proposed a method, namely RWRHMDA, for prediction of disease-associated miRNAs. This method was based on RWRH algorithm, which was successfully proposed for disease gene prediction on a heterogeneous network of genes and disease phenotypes. In par- ticular, we used this algorithm to rank disease candidate miRNAs on a heteroge- neous network of phenotypes and miRNAs, which was constructed by integrat- ing a shared target gene-based microRNA functional similarity network and a disease phenotype similarity network. Comparing the prediction performance of RWRHMDA with that of RWRMDA on a set of 35 disease phenotypes, we found that RWRHMDA significantly outperformed RWRMDA irrespective of parameter settings since it better exploited “disease module” principle. In addi- tion, using RWRHMDA method, we identified eight novel Alzheimer’s disease- associated miRNAs.

1. **Paper Name:Efficient Heart Disease Prediction System using Decision Tree Author: Purushottam,Prof. (Dr.) Kanak Saxena,Richa Sharma**

Description :—Cardiovascular disease (CVD) is a big reason of morbidity and mortality in the current living style. Identification of Cardiovascular disease is an important but a complex task that needs to be performed very minutely, efficiently and the correct automation would be very desirable. Every human being can not be equally skillful and so as doctors. All doctors cannot be equally skilled in every sub specialty and at many places we don’t have skilled and specialist doctors available easily. An automated system in medical diagnosis would enhance medical care and it can also reduce costs. In this study, we have designed a system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the user’s requirement. The performance of the system is evaluated in terms of classification accuracy and the results shows that the sys- tem has great potential in predicting the heart disease risk level more accurately

1. **Paper Name:Diabetes Disease Prediction Using Data Mining Author:Deeraj Shetty,Kishor Rit,Sohail Shaikh.**

Abstract:Data mining is a subfield in the subject of software engineering. It is the methodical procedure of finding examples in huge data sets including tech- niques at the crossing point of manufactured intelligence, machine learning, insights, and database systems. The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Our examination concentrates on this part of Medical conclusion learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination is to assemble Intelligent Diabetes Disease Predic- tion System that gives analysis of diabetes malady utilizing diabetes patient’s database. In this system, we propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient’s database and analyze them by taking various attributes of diabetes for prediction of diabetes disease.

1. Paper Name:Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheuma- toid Arthritis

Author:Shang-Ming Zhou1 \*, Fabiola Fernandez-Gutierrez1 , Jonathan Kennedy1

Abstract:1) To use data-driven method to examine clinical codes (risk factors) of a medical condition in primary care electronic health records (EHRs) that can accurately predict a diagnosis of the condition in secondary care EHRs. 2) To develop and validate a disease phenotyping algorithm for rheumatoid arthri- tis using primary care EHRs.This study linked routine primary and secondary care EHRs in Wales, UK. A machine learning based scheme was used to iden- tify patients with rheumatoid arthritis from primary care EHRs via the fol- lowing steps: i) selection of variables by comparing relative frequencies of Read codes in the primary care dataset associated with disease case compared to non- disease control (disease/non-disease based on the secondary care diagnosis); ii) reduction of predictors/associated variables using a Random Forest method, iii) induction of decision rules from decision tree model. The proposed method was then extensively validated on an independent dataset, and compared for per- formance with two existing deterministic algorithms for RA which had been developed using expert clinical knowledge.

1. **Paper Name:Prediction of Disease Infection of Welsh Onions by Rust Fungus Based on Temperature and Wetness Duration**

Author:Hiroyuki Takanashi, Hiromitsu Furuya and Seiji Chonan Abstract:The style of agriculture practiced in Japan and other countries in Asia is small scale compared to that in North America. While systematic produc- tion and management systems have been maintained in Europe and America, Japanese agricultural style tends to depend on past experiences, and applica- tion of agricultural chemicals is guided by the calendar and past experiences. Japan is also advanced in the field of plant disease prediction. This paper fo- cuses on a prediction model of disease infection for a foliar parasite on Welsh onions. Rust fungus disease is the most typical disease on Welsh onions, and the Weibull probability density function is appropriate for approximating the infection rate of the disease. The model utilizes temperature and wetness dura- tion to predict the infection of Welsh onions by rust fungus. Producers, then, can use the model to determine the day on which the infection rate will rapidly increase, then carry out appropriate countermeasures to the disease. The pro- posed prediction method is applicable to several infections found throughout Asia.

1. **Paper Name:Neurodegenerative disease prediction based on gait analysis sig- nals acquired with force-sensitive**

Author:ger Selzler 1 , James R. Green 1 , Rafik Goubran. Abstract:Neurodegenerative diseases such as Parkinson’s Disease (PD), Hunt- ington Disease (HD), and Amyotrophic Lateral Sclerosis (ALS) affect the lives of thousands of people around the world. One of the consequences of such dis- eases occurs in the motor neurons of the patients, resulting in problems in move- ment, causing a change in gait pattern. Force sensitive resistors can be used to measure the force/pressure between the shoe and the patient’s foot, provid- ing information about the gait dynamics when the patient walks. This project uses signals from the Gait Dynamics in Neuro-Degenerative Disease database to extract features for classification of neurodegenerative diseases (NDD). Manu- ally labelled features from the database are used for comparison with previous studies. Time series signals is also used, where algorithms for signal reliability, feature extraction and feature selection are implemented, allowing real-time signal processing and classification. Multiple feature sets are used for classifi- cation with algorithms such as K-nearest neighbor, Support Vector Machines, and Decision Trees, and the performance of these algorithms are then reported. This study presents a realtime system with accuracy exceeding 82% for the aforementioned diseases. Finally, a discussion about possible improvements for future studies are presented.

1. **Paper name:Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning**

Author:Rohit Binu Mathew, Sandra Varghese, Sera Elsa Joy, Swanthana Susan Alex

Abstract:Hospitals are the most widely used means by which a sick person gets medical check-ups, disease diagnosis and treatment recommendation. This has been a practice by almost all the people over the world. People consider it as the most reliable means to check their health status. The proposed system is to cre- ate an alternative to this conventional method of visiting a hospital and making an appointment with a doctor to get diagnosis. This research intends to apply the concepts of natural language processing and machine learning to create a chatbot application. People can interact with the chatbot just like they do with another human and through a series of queries, chatbot will identify the symp- toms of the user and thereby, predicts the disease and recommends treatment. This system can be of great use to people in conducting daily check-ups, makes people aware of their health status and encourages people to make proper mea- sures to remain healthy. According to this research, such a system is not widely used and people are less aware of it. Executing this proposed framework can help people avoid the time-consuming method of visiting hospitals by using this free of cost application, wherever they are.

1. **Paper Name:Chatbot for Disease Prediction and Treatment Recommenda- tion using Machine Learning**

Author:Rohit Binu Mathew, Sandra Varghese, Sera Elsa Joy, Swanthana Susan Alex

Abstract: Hospitals are the most widely used means by which a sick person gets medical check-ups, disease diagnosis and treatment recommendation. This has been a practice by almost all the people over the world. People consider it as the most reliable means to check their health status. The proposed system is to cre- ate an alternative to this conventional method of visiting a hospital and making an appointment with a doctor to get diagnosis. This research intends to apply the concepts of natural language processing and machine learning to create a chatbot application. People can interact with the chatbot just like they do with another human and through a series of queries, chatbot will identify the symp- toms of the user and thereby, predicts the disease and recommends treatment. This system can be of great use to people in conducting daily check-ups, makes people aware of their health status and encourages people to make proper mea- sures to remain healthy. According to this research, such a system is not widely used and people are less aware of it. Executing this proposed framework can help people avoid the time-consuming method of visiting hospitals by using this free of cost application, wherever they are. .

# CHAPTER 3

# PROBLEM STATEMENT

There are many tools related to disease prediction. But particularly heart re-lated diseases have been analyzed and risk level is generated. But generally there are no such tools that are used for prediction of general diseases. So Dis- ease Predictor helps for the prediction of the general diseases.

# CHAPTER 4

# PROJECT REQUIREMENT

* 1. **EXTERNAL INTERFACE REQUIREMENT**
     1. **User Interface**

Application Based Heart Disease Predication.

* + 1. **Hardware Interfaces:**

RAM : 8 GB

As we are using Machine Learning Algorithm and Various High Level Libraries Laptop

RAM minimum required is 8 GB.

Hard Disk : 40 GB

Data Set of CT Scan images is to be used hence minimum 40 GB Hard Disk memory is required.

Processor : Intel i5 Processor

Pycharm IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required

IDE : Pycharm

Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that makes typing feasible and fast.

Coding Language : Python Version 3.5

Highly specified Programming Language for Machine Learning because of avail- ability of High Performance Libraries.

Operating System : Windows 10

Latest Operating System that supports all type of installation and development Environment

* + 1. **Software Interfaces Operating System:** Windows 10

IDE:Spyder

Programming Language : Python

* 1. **NON FUNCTIONAL REQUIREMENT**
     1. **PerformanceRequirements**

The performance of the functions and every module must be well. The over- allperformance of the software will enable the users to work eciently. Perfor- mance of encryption of data should be fast. Performance of the providingvirtual environment should be fastSafety Requirement.

•The application is designed in modules where errors can be detected and xedeasily. This makes it easier to install and update new functionality if required.

* + 1. **Safety Requirement**

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

* + 1. **Software Quality Attributes**

Our software has many quality attribute that are given below:-

Adaptability: This software is adaptable by all users.

Availability: This software is freely available to all users. The availability of the software is easy for everyone.

Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.

Reliability: The performance of the software is better which will increase the reliabilityof the Software.

User Friendliness: Since, the software is a GUI application; the output gener- ated is much user friendly in its behavior.

Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled.

Security: Users are authenticated using many security phases so reliable secu- rity is provided.

Testability: The software will be tested considering all the aspects.

# CHAPTER 5

# SYSTEM ANALYSIS

* 1. **SYSTEM ARCHITECTURE**

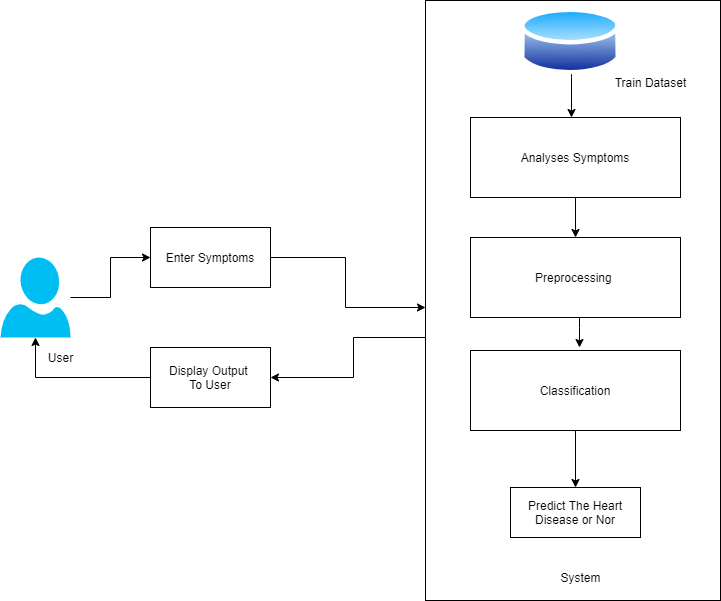


Figure 5.1: System Architecture-l

* + 1. **Module**
       - Preprocessing
       - Analyses Symptoms
       - Classification
    2. **Data Flow Diagram**

In Data Flow Diagram,we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and cir- cle show our system,In DFD1 we show actual input and actual output of system input of our system is text or image and output is rumor detected like wise in

DFD 2 we present operation of user as well as admin.

Figure 5.2: Data Flow diagram

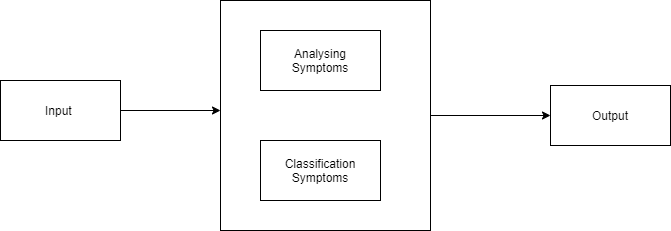


Figure 5.3: Data Flow diagram

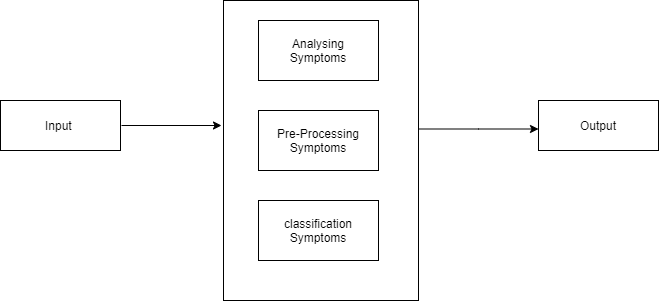


Figure 5.4: Data Flow diagram

* 1. **UML DIAGRAMS**

Unified Modeling Language is a standard language for writing software blueprints.The UML may be used to visualize,specify,construct and document the artifacts of

a software intensive system. UML is process independent, although optimally it should be used in process that is use case driven, architecture-centric, iterative, and incremental. The Number of UML Diagram is available.

* Use case Diagram.
* Activity Diagram.
* Sequence Diagram.
* Class Diagram.

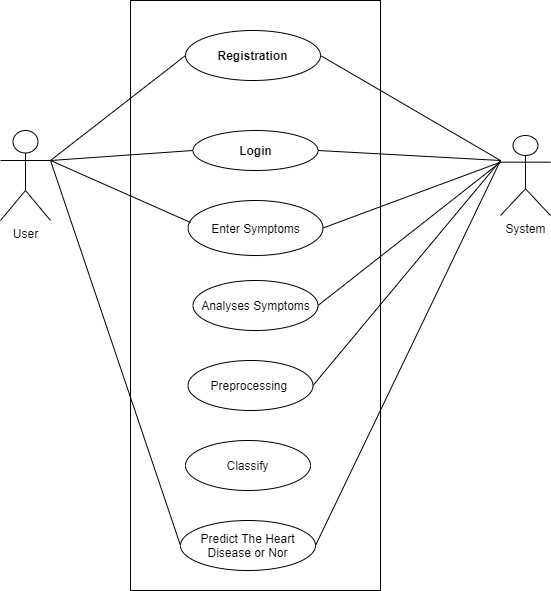


Figure 5.5: Use case Diagram

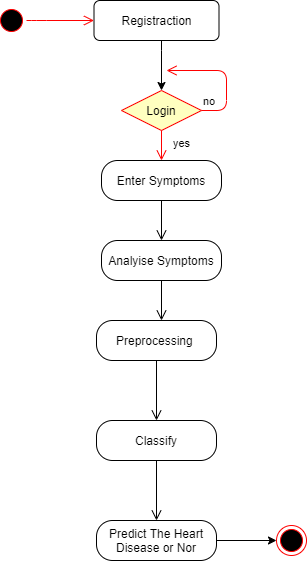


Figure 5.6: Activity Diagram

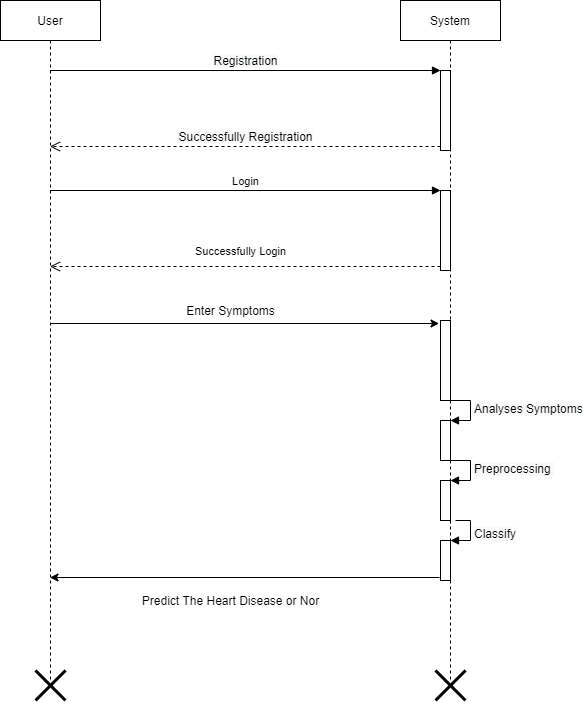


Figure 5.7: Sequence Diagram

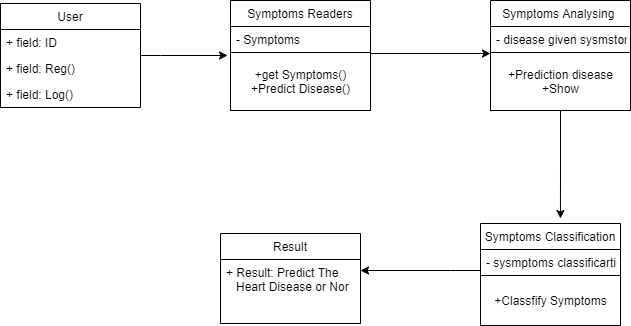


Figure 5.8: Class Diagram

# CHAPTER 6 REFERENCES

* + 1. D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, “A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics,” IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3033–3049, 2015.
    2. M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, “Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System,” IEEE Communications, Vol. 55, No. 1, pp. 54–61, Jan. 2017.
    3. M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, ”Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring,” ACM/Springer Mobile Networks and Applications, Vol. 21, No. 5, pp. 825C845, 2016
    4. M. Chen, P. Zhou, G. Fortino, “Emotion Communication System,” IEEE Access, DOI: 10.1109/ACCESS.2016.2641480, 2016.
    5. J. Wang, M. Qiu, and B. Guo, “Enabling real-time information service on telehealth system over cloud-based big data platform,” Journal of Sys- tems Architecture, vol. 72, pp. 69–79, 2017.
    6. Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, “Healthcps: Healthcare cyber-physical system assisted by cloud and big data,” IEEE Systems Journal, 2015.
    7. K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, “Localization based on social big data analysis in the vehicular networks,” IEEE Trans- actions on Industrial Informatics, 2016.