**HEART ATTACK RISK PREDICTION USING RETINAL IMAGES**

**ABSTRACT**

The structure and function of the microvascular are significantly influenced by the key cardiovascular disease risk factors of hypertension and heart attacks. Images taken with a fundus camera can be used to spot irregularities in the blood vessels of retina that indicate the extent injury on blood vessels by hypertension and heart attacks. Using machine learning and AI techniques, detecting the preclinical signs that fall below the threshold of an observer. The proposed methodology aimed to investigate the effects of hypertension and heart attacks on morphological characteristics of retinal blood vessels. With a diagnosis of hypertension and heart attack, data scientists collect retinal images. Interference data is removed— information about structures other than that retinal vasculature using the vessel segmentation method, leaving only morphological details about the blood vessel of retina. The method aims to create a system for visual image-based heart disease detection, especially in young people, to identify heart disease. In the study, a dataset of retinal imaging is used, and retinal vessel segmentation is used to separate the vessels in the images. In a number of specialties, such as laryngology, neurosurgery, and ophthalmology, the analysis of blood vessels is crucial for diagnosis, therapy planning and execution, and assessment of clinical outcomes. Therefore, vessel segmentation is a crucial method for using the retinal image to detect heart disease. Changes in the eyes may be a sign of many conditions.

**CHAPTER 1 INTRODUCTION**

The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body’s cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood vessels, for example, veins, arteries, and capillaries. These blood vessels deliver blood all over the body. Abnormalities in normal blood flow from the heart cause several types of heart diseases which are commonly known as cardiovascular diseases (CVD). Heart diseases are the main reasons for death worldwide. According to the survey of the World Health Organization (WHO), 17.5 million total global deaths occur because of heart attacks and strokes. More than 75% of deaths from cardiovascular diseases occur mostly in middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack . Therefore, prediction of cardiac abnormalities at the early stage and tools for the prediction of heart diseases can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases.

Due to the development of advance healthcare systems, lots of patient data are nowadays available (i.e. Big Data in Electronic Health Record System) which can be used for designing predictive models for Cardiovascular diseases. Data mining or machine learning is a discovery method for analyzing big data from an assorted perspective and encapsulating it into useful information. “Data Mining is a non-trivial extraction of implicit, previously unknown and potentially useful information about data”. Nowadays, a huge amount of data pertaining to disease diagnosis, patients etc. are generated by healthcare industries. Data mining provides a number of techniques which discover hidden patterns or similarities from data.

Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart disease prediction system which was validated on two open access heart disease prediction datasets. Data mining is the computer based process of extracting useful information from enormous sets of databases. Data mining is most helpful in an explorative analysis because of nontrivial information from large volumes of evidence

1. edical data mining has great potential for exploring the cryptic patterns in the data sets of the clinical domain.

These patterns can be utilized for healthcare diagnosis. However, the available raw medical data are widely distributed, voluminous and heterogeneous in nature .This data needs to be collected in an organized form. This collected data can be then integrated to form a

1

medical information system. Data mining provides a user-oriented approach to novel and hidden patterns in the Data The data mining tools are useful for answering business questions and techniques for predicting the various diseases in the healthcare field. Disease prediction plays a significant role in data mining. This paper analyzes the heart disease predictions using classification algorithms. These invisible patterns can be utilized for health diagnosis in healthcare data.

Data mining technology affords an efficient approach to the latest and indefinite patterns in the data. The information which is identified can be used by the healthcare administrators to get better services. Heart disease was the most crucial reason for victims in the countries like India, United States. In this project we are predicting the heart disease using classification algorithms. Machine learning techniques like Classification algorithms such as DNN Classifications, Logistic Regression are used to explore different kinds of heart based problems.

**2.1. Motivation**

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems.

Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data [12]. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: “How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?” This is the main motivation for this research.

**1.2. Problem statement**

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like “What is the average age of patients who have heart disease?”, “How many surgeries had resulted in hospital stays longer than 10 days?” “Identify the female patients who are single, above 30 years old, and who have been treated for cancer.” However, they cannot answer complex queries like “Identify the important preoperative predictors that increase the length of hospital stay”, “Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?”, and “Given patient records, predict the probability of patients getting a heart disease.”

Clinical decisions are often made based on doctors’intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et alproposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome [17]. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

**1.3. Research objectives**

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes and Neural Network.

IHDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs. To enhance visualization and ease of interpretation, it displays the results both in tabular and graphical forms.

### 2. LITERATURE SURVEY

Machine Learning techniques are used to analyze and predict the medical data information resources. Diagnosis of heart disease is a significant and tedious task in medicine. The term Heart disease encompasses the various diseases that affect the heart. The exposure of heart disease from various factors or symptom is an issue which is not complimentary from false presumptions often accompanied by unpredictable effects. The data classification is based on Supervised Machine Learning algorithm which results in better accuracy. Here we are using the DNN Classifications as the training algorithm to train the heart disease dataset and to predict the heart disease. The results showed that the medicinal prescription and designed prediction system is capable of prophesying the heart attack successfully. Machine Learning techniques are used to indicate the early mortality by analyzing the heart disease patients and their clinical records (Richards, G. et al., 2001). (Sung, S.F. et al., 2015) have brought about the two Machine Learning techniques, k-nearest neighbor model and existing multi linear regression to predict the stroke severity index (SSI) of the patients. Their study show that k- nearest neighbor performed better than Multi Linear Regression model. (Arslan, A. K. et al., 2016) have suggested various Machine Learning techniques such as support vector machine (SVM), penalized logistic regression (PLR) to predict the heart stroke. Their results show that SVM produced the best performance in prediction when compared to other models.Boshra Brahmi et al, [20] developed different Machine Learning techniques to evaluate the prediction and diagnosis of heart disease. The main objective is to evaluate the different classification techniques such as J48, Decision Tree, KNN and Naïve Bayes. After this, evaluating some performance in measures of accuracy, precision, sensitivity, specificity are evaluated .

# Data source

Clinical databases have collected a significant amount of information about patients and their medical conditions. Records set with medical attributes were obtained from the Cleveland Heart Disease database. With the help of the dataset, the patterns significant to the heart attack diagnosis are extracted. The records were split equally into two datasets: training dataset and testing dataset. A total of 303 records with 76 medical attributes were obtained. All the attributes are numeric-valued. We are working on a reduced set of attributes, i.e. only 14 attributes.

All these restrictions were announced to shrink the digit of designs, these are as follows:

* 1. The features should seem on a single side of the rule.
  2. The rule should distinct various features into the different groups.

3

* 1. The count of features available from the rule is organized by medical history of people having heart disease only.

The following table shows the list of attributes on which we are working.

|  |  |  |
| --- | --- | --- |
| S no | Attribute Name | Description |
| 1 | Age | age in years |
| 2 | Sex | (1 = male; 0 = female) |
| 3 | Cp | Chest Pain |
| 4 | Trestbps | resting blood pressure (in mm Hg on admission to the hospital) |
| 5 | Chol | serum cholesterol in mg/dl |
| 6 | Fbs | (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false) |
| 7 | Restecg | resting electrocardiographic results |
| 8 | Thalach | maximum heart rate achieved |
| 9 | Exang | exercise induced angina (1 = yes; 0 = no) |
| 10 | Oldpeak | ST depression induced by exercise relative to rest |
| 11 | Slope | the slope of the peak exercise ST segment |
| 12 | Ca | number of major vessels (0-3) colored by flourosopy |
| 13 | Thal | 3 = normal; 6 = fixed defect; 7 = reversible defect |
| 14 | Target | 1 or 0 |

Table 2.1: List of attributes

1. Arti Gupta, Maneesh Shreevastava IJETAE, 2011. **Medical Diagnosis using Back Propagation Algorithm.** In this paper, feed forward Back Propagation algorithm is described which is used as a classifier to distinguish between infected and non-infected person in medical diagnosis. The back propagation algorithm presented in this paper used for training depends on a multilayer neural network with a very small learning rate, especially when using a large training set size**.**
2. Shraddha Subhash Shirsath, Prof. Shubhangi Patil IJIRSET, June 2018. **Disease Prediction using Machine Learning over Big Data.** This paper discusses about machine learning algorithm which is used for the accurate disease prediction. Here to achieve the incomplete data latent factor model is used. DNN algorithm is used for clarification of large volume of data from hospital and then Convolutional Neural Network Based Multimodal Disease Prediction (DNN-MDRP) algorithm helps to provide result of a disease prediction.
3. Nikita Kamble, Manjiri Harmalkar, Manali Bhoir, Supriya Chaudhary**,** IJSRCSEIT, 2017. **Smart Health Prediction System Using Machine Learning.** The paper presents an overview of the Machine Learning techniques with its applications, medical and educational aspects of Clinical Predictions. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. Such a large amount of data cannot be processed by humans in a short time to make diagnosis, and treatment schedules. A major objective is to evaluate Machine Learning techniques in clinical and health care applications to develop accurate decisions. It also gives a detailed discussion of medical Machine Learning techniques which can improve various aspects of Clinical Predictions. It is a new powerful technology which is of high interest in computer world. It is a sub field of computer science that uses already existing data in different databases to transform it into new researches and results. It makes use of machine learning and database management to extract new patterns from large data sets and the knowledge associated with these patterns. The actual task is to extract data by automatic or semi-automatic means. The different parameters included in Machine Learning include clustering, forecasting, path analysis and predictive analysis.
4. Nilesh Borisagar, Dipa Barad, Priyanka Raval, Conference paper (PICCN), April 2017. **Chronic Kidney Disease Prediction using Back Propagation Neural Network Algorithm**. In this paper, various training algorithms like Levenberg, Bayesian regularization, Scaled Conjugate and Resilient back propagation algorithm are discussed. After neural network is trained using back propagation algorithms, this trained neural network system is used for detection of kidney disease in the human body. The back propagation algorithms presented here have capacity for distinguishing amongst infected patients or non-infected person.
5. Sellappan Palaniappan, Rafiah Awang IEEE, 2008**. Intelligent Heart Disease Prediction System Using Machine Learning Technique.** This paper discusses about the development of prototype using Machine Learning techniques, namely, DNN Classification, DNN and Neural Network. It can answer complex “what if “queries which traditional decision support system cannot.it is web-based, user-friendly, scalable, reliable and expandable.
6. M.A. Nishara Banu, B Gomathy, IJTRA, Dec 2013. **Disease Prediction System Using Machine Learning Techniques.** This paper analyzes the heart disease predictions using different classification algorithms. Here medical Machine Learning techniques like Association Rule Mining, Clustering and Classification Algorithms such as DNN Classification, C4.5 Algorithm are implemented to analyze the different kinds of heart based problems. Maximal Frequent Itemset Algorithm (MAFIA) is used for mining maximal frequent item sets from a transactional database and C4.5 Algorithm and Clustering Algorithm like K-Means are used to draw a DNN Classification. With the help of this technique, the accuracy of disease can be validated.

Following is the details of the literature survey conducted in tabular form, given in Table 2.1

**Table 2.1 Summary of Literature Survey**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No.** | **Paper** | **Author** | **Published** | **Methodology used** | **Issues** |
| 1 | Medical Diagnosis using Back propagation Algorithm | Arti Gupta, Maneesh Shreevastava | IJETAE,  2011 | feed forward Back propagation algorithm | very small learning rate when using a large training set size |
| 2 | Disease Prediction Using Machine  Learning Over Big Data | Shraddha Subhash Shirsath, Prof. Shubhangi Patil | IJIRSET,  June 2018 | Naive  Bayesian,  DNN-MDRP | Not a good  way to describe  complex disease,  algorithm performs poor for the numeric data. |
| 3 | Smart Health Prediction System Using Machine Learning | Nikita Kamble, Manjiri Harmalkar, Manali Bhoir, Supriya Chaudhary | IJSRCSEIT, 2017 | Naive Bayesian | No more data reduction method is used, limited datasets |
| 4 | Chronic kidney disease prediction using Back Propagation Neural Network Algorithm | Nilesh Borisagar, Dipa Barad, Priyanka Raval | Conference paper  (PICCN),  April 2017 | Levenberg, Bayesian regularization, Scaled  Conjugate and Resilient back propagation algorithm | Slow learning rate and Execution time is more. |
| 5 | Intelligent Heart Disease Prediction System  Using Machine Learning Technique | Sellappan Palaniappan, Rafiah Awang | IEEE,  2008 | Multilayer Perceptron  Neural Network with Back Propagation Algorithm | Less accuracy of output |
| 6 | Disease Predicting System Using Machine Learning Techniques | M.A.Nishara Banu,  B Gomathy | IJTRA,  Dec 2013 | MAFIA Algorithm, C4.5 Algorithm,  K-means clustering | Not used in large real time health datasets to predict the diseases |

### CHAPTER 3

### SYSTEM DESIGN

## EXISTING SYSTEM

Clinical decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. There are many ways that a medical misdiagnosis can present itself. Whether a doctor is at fault, or hospital staff, a misdiagnosis of a serious illness can have very extreme and harmful effects. The National Patient Safety Foundation cites that 42% of medical patients feel they have had experienced a medical error or missed diagnosis. Patient safety is sometimes negligently given the back seat for other concerns, such as the cost of medical tests, drugs, and operations. Medical Misdiagnoses are a serious risk to our healthcare profession. If they continue, then people will fear going to the hospital for treatment. We can put an end to medical misdiagnosis by informing the public and filing claims and suits against the medical practitioners at fault.

## Disadvantages:

* + - Prediction is not possible at early stages.
    - In the Existing system, practical use of collected data is time consuming.
    - Any faults occurred by the doctor or hospital staff n predicting would lead to fatal incidents.
    - Highly expensive and laborious process needs to be performed before treating the patient to find out if he/she has any chances to get heart disease in future.

## PROPOSED SYSTEM

This section depicts the overview of the proposed system and illustrates all of the components, techniques and tools are used for developing the entire system. To develop an intelligent and user-friendly heart disease prediction system, an efficient software tool is needed in order to train huge datasets and compare multiple machine learning algorithms. After choosing the robust algorithm with best accuracy and performance measures, it will be implemented on the development of the smart phone-based application for detecting and predicting heart disease risk level. Retinal fundus photographs can be used to detect a range of retinal conditions. Here we show that deep-learning models trained instead on external photographs of the eyes can be used to detect diabetic retinopathy (DR), diabetic macular oedema and poor blood glucose control. We developed the models using eye photographs from patients with diabetes from 301 DR screening sites and evaluated the models on four tasks and four validation datasets with a patients from 198 additional screening sites. For all four tasks, the predictive performance of the deep-learning models was significantly higher than the performance of logistic regression models using self-reported demographic and medical history data, and the predictions generalized to patients with dilated pupils, to patients from a different DR screening programme and to a general eye care programme that included diabetics and non-diabetics. We also explored the use of the deep-learning models for the detection of elevated lipid levels. The utility of external eye photographs for the diagnosis and management of diseases should be further validated with images from different cameras and patient populations.

5

## ALGORITHMS

* + 1. **Logistic Regression**

A popular statistical technique to predict binomial outcomes (y = 0 or 1) is Logistic Regression. Logistic regression predicts categorical outcomes (binomial / multinomial values of y). The predictions of Logistic Regression (henceforth, LogR in this article) are in the form of probabilities of an event occurring, i.e. the probability of y=1, given certain values of input variables x. Thus, the results of LogR range between 0-1.

LogR models the data points using the standard logistic function, which is an S- shaped curve also called as sigmoid curve and is given by the equation:



## Logistic Regression Assumptions:

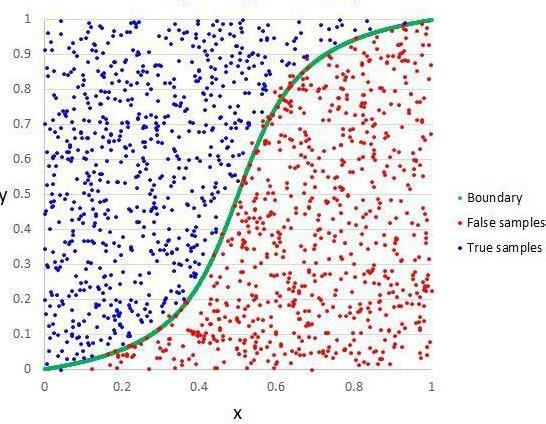
* + - * Logistic regression requires the dependent variable to be binary.
      * For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.
      * Only the meaningful variables should be included.
      * The independent variables should be independent of each other·
      * Logistic regression requires quite large sample sizes.
      * Even though, logistic (**logit**) regression is frequently used for binary variables (2 classes), it can be used for categorical dependent variables with more than 2 classes.
      * In this case it’s called Multinomial Logistic Regression. 

Fig 3.1: logistic regression

## DNN Classifications

DNN Classifications is a supervised learning algorithm which is used for both classification as well as regression .But however ,it is mainly used for classification problems .As we know that a forest is made up of trees and more trees means more robust forest .

Similarly ,DNN Classifications creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting .It is ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result .

Working of DNN Classifications with the help of following steps:

* First ,start with the selection of random samples from a given dataset.
* Next ,this algorithm will construct a decision tree for every sample .Then it will get the prediction result from every decision tree .
* In this step, voting will be performed for every predicted result.
* At last ,select the most voted prediction results as the final prediction result. The following diagram will illustrates its working-

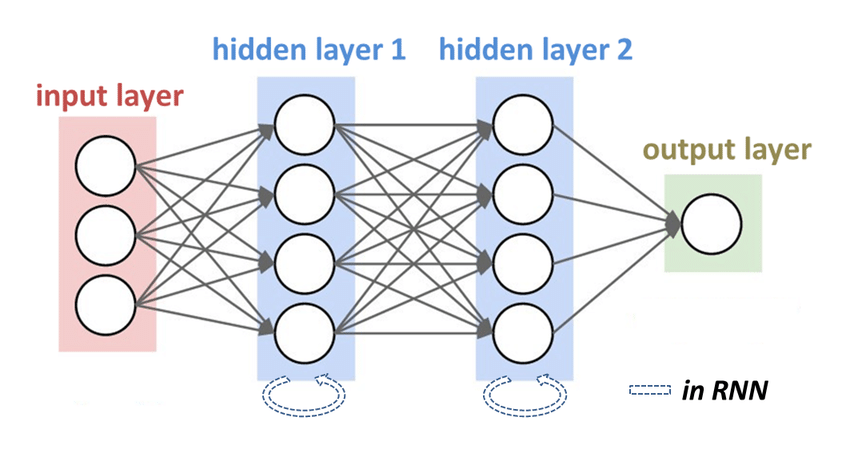


Fig 3.2: DNN Classifications

## FEASIBILITY STUDY

A Feasibility Study is a preliminary study undertaken before the real work of a project starts to ascertain the likely hood of the projects success. It is an analysis of possible alternative solutions to a problem and a recommendation on the best alternative.

## Economic Feasibility:

It is defined as the process of assessing the benefits and costs associated with the development of project. A proposed system, which is both operationally and technically

8

feasible, must be a good investment for the organization. With the proposed system the users are greatly benefited as the users can be able to detect the fake news from the real news and are aware of most real and most fake news published in the recent years. This proposed system does not need any additional software and high system configuration. Hence the proposed system is economically feasible.

## Technical Feasibility:

The technical feasibility infers whether the proposed system can be developed considering the technical issues like availability of the necessary technology, technical capacity, adequate response and extensibility. The project is decided to build using Python. Jupyter Note Book is designed for use in distributed environment of the internet and for the professional programmer it is easy to learn and use effectively. As the developing organization has all the resources available to build the system therefore the proposed system is technically feasible.

## Operational Feasibility:

Operational feasibility is defined as the process of assessing the degree to which a proposed system solves business problems or takes advantage of business opportunities. The system is self-explanatory and doesn’t need any extra sophisticated training. The system has built-in methods and classes which are required to produce the result. The application can be handled very easily with a novice user. The overall time that a user needs to get trained is 14 less than one hour. As the software that is used for developing this application is very economical and is readily available in the market. Therefore the proposed system is operationally feasible.

* 1. **EFFORT, DURATION AND COST ESTIMATION USING COCOMO MODEL** The Cocomo (Constructive Cost Model) model is the most complete and thoroughly documented model used in effort estimation. The model provides detailed formulas for determining the development time schedule, overall development effort, and effort breakdown by phase and activity as well as maintenance effort.

## Product Attributes

* + - * **Required reliability (RELY):** It is used to express an effect of software faults ranging from slight inconvenience (VL) to loss of life (VH). The nominal value (NM) denotes moderate recoverable losses.
      * **Data bytes per DSI (DATA):** The lower rating comes with lower size of a database. Complexity (CPLX): The attribute expresses code complexity again ranging from straight batch code (VL) to real time code with multiple resources scheduling (XH)

## Computer Attributes

* + - * **Execution time (TIME) and memory (STOR) constraints:** This attribute identifies the percentage of computer resources used by the system. NM states that less than 50% is used; 95% is indicated by XH.]
      * **Virtual machine volatility (VIRT):** It is used to indicate the frequency of changes made to the hardware, operating system, and overall software environment. More frequent and significant changes are indicated by higher ratings.
      * **Development turnaround time (TURN):** This is a time from when a job is submitted until output becomes received. LO indicated a highly interactive environment, VH quantifies a situation when this time is longer than 12 hours.

## Personal Attributes:

* + - * Analyst capability(ACAP) and programmer capability (PCAP):
      * This describes skills of the developing team. The higher the skills, the higher the rating.
      * Application experience (AEXP), language experience (LEXP), and virtual machine experience (VEXP):
* These are used to quantify the number of experience in each area by the development team; more experience, higher rating.

## Project Attributes:

* + - * **Modern development practices (MODP):** deals with the amount of use of modern software practices such as structural programming and object oriented approach.
        + **Use of software tools (TOOL):** is used to measure a level of sophistication of automated tools used in software development and a degree of integration among the tools being used. Higher rating describes levels in both aspects.
        + **Schedule effects (SCED):** concerns the amount of schedule compression (HI or VH), or schedule expansion (LO or VL) of the development schedule in comparison to a nominal (NM) schedule

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | VL | LO | NM | HI | VH | XH |
| RELY | 0.75 | 0.88 | 1.00 | 1.15 | 1.40 |  |
| DATA |  | 0.94 | 1.00 | 1.08 | 1.16 |  |
| CPLX | 0.70 | 0.85 | 1.00 | 1.15 | 1.30 | 1.65 |
| TIME |  |  | 1.00 | 1.11 | 1.30 | 1.66 |
| STOR |  |  | 1.00 | 1.06 | 1.21 | 1.56 |
| VIRT |  | 0.87 | 1.00 | 1.15 | 1.30 |  |
| TURN |  | 0.87 | 1.00 | 1.15 | 1.30 |  |
| ACAP | 1.46 | 1.19 | 1.00 | 0.86 | 0.71 |  |
| AEXP | 1.29 | 1.13 | 1.00 | 0.91 | 0.82 |  |
| PCAP | 1.42 | 1.17 | 1.00 | 0.86 | 0.70 |  |
| LEXP | 1.14 | 1.07 | 1.00 | 0.95 |  |  |
| VEXP | 1.21 | 1.10 | 1.00 | 0.90 |  |  |
| MODP | 1.24 | 1.10 | 1.00 | 0.91 | 0.82 |  |
| TOOL | 1.24 | 1.10 | 1.00 | 0.91 | 0.83 |  |
| SCED | 1.23 | 1.08 | 1.00 | 1.04 | 1.10 |  |

Table 3.2:Project Attributes

12

### SYSTEM REQUIREMENT: HARDWARE REQUIREMENTS:

Processor : PENTIUM IV

Ram : 1 GB SD RAM

Monitor : 15” COLOR

Hard Disk : 80 GB

Keyboard : STANDARD 102 KEYS

Mouse : 3 BUTTON

### SOFTWARE CONFIGURATION:

Operating System : Windows Environment : Python, HTML

### EXISTING SYSTEM:

The diagnosis of Vitamin Deficiency detection systems. The common approach of designing is

* + Image Acquisition
  + Image preprocessing

### Image Acquisition

The database images we use contain digital images taken by means of a digital camera. These images are fed into a computer system for further processing.

The images mainly dealt for the research are RGB images. Since color is a powerful descriptor the RGB images are considered for the paper. The database images are obtained from different sources and the size of the images is non-standard.

### Image preprocessing

Preprocessing as the fundamental stage of detection system helps to enhance the quality of an image by removing hairs, noise and air bubbles on the Skin Tissue.

The enhanced image is used for feeding the next step. In preprocessing of an image, there are many existing techniques which can be classified into two groups; binary and gray color images. The common images chosen for research here are color images.

### Disadvantages:

An image is segmented band allocation of each pixel of the image to existing classes that would finally lead to the lesion segregation from the healthy Skin Tissue.

**CHAPTER 4**

**SOFTWARE REQUIREMENTS**

**SOFTWARE ENVIRONMENT**

**Python:**

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

* **Python is Interpreted** − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
* **Python is Interactive** − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
* **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
* **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

**History of Python**

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

**Python Features**

Python's features include −

* **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
* **Easy-to-read** − Python code is more clearly defined and visible to the eyes.
* **Easy-to-maintain** − Python's source code is fairly easy-to-maintain.
* **A broad standard library** − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
* **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
* **Portable** − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
* **Extendable** − You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
* **Databases** − Python provides interfaces to all major commercial databases.
* **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
* **Scalable** − Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below −

* It supports functional and structured programming methods as well as OOP.
* It can be used as a scripting language or can be compiled to byte-code for building large applications.
* It provides very high-level dynamic data types and supports dynamic type checking.
* It supports automatic garbage collection.
* It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

**Getting Python**

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python [https://www.python.org](https://www.python.org/).

Windows Installation

Here are the steps to install Python on Windows machine.

* Open a Web browser and go to <https://www.python.org/downloads/>.
* Follow the link for the Windows installer python-XYZ.msifile where XYZ is the version you need to install.
* To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.
* Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

The Python language has many similarities to Perl, C, and Java. However, there are some definite differences between the languages.

**First Python Program**

Let us execute programs in different modes of programming.

**Interactive Mode Programming**

Invoking the interpreter without passing a script file as a parameter brings up the following prompt −

$ python

Python2.4.3(#1,Nov112010,13:34:43)

[GCC 4.1.220080704(RedHat4.1.2-48)] on linux2

Type"help","copyright","credits"or"license"for more information.

>>>

Type the following text at the Python prompt and press the Enter −

>>>print"Hello, Python!"

If you are running new version of Python, then you would need to use print statement with parenthesis as in **print ("Hello, Python!");**. However in Python version 2.4.3, this produces the following result −

Hello, Python!

**Script Mode Programming**

Invoking the interpreter with a script parameter begins execution of the script and continues until the script is finished. When the script is finished, the interpreter is no longer active.

Let us write a simple Python program in a script. Python files have extension **.py**. Type the following source code in a test.py file −

print"Hello, Python!"

We assume that you have Python interpreter set in PATH variable. Now, try to run this program as follows −

$ python test.py

This produces the following result −

Hello, Python!

**Flask Framework:**

Flask is a web application framework written in Python. Armin Ronacher, who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

Http protocol is the foundation of data communication in world wide web. Different methods of data retrieval from specified URL are defined in this protocol.

The following table summarizes different http methods −

|  |  |
| --- | --- |
| **Sr.No** | **Methods & Description** |
| 1 | **GET**  Sends data in unencrypted form to the server. Most common method. |
| 2 | **HEAD**  Same as GET, but without response body |
| 3 | **POST**  Used to send HTML form data to server. Data received by POST method is not cached by server. |
| 4 | **PUT**  Replaces all current representations of the target resource with the uploaded content. |
| 5 | **DELETE**  Removes all current representations of the target resource given by a URL |

By default, the Flask route responds to the **GET** requests. However, this preference can be altered by providing methods argument to **route()** decorator.

In order to demonstrate the use of **POST** method in URL routing, first let us create an HTML form and use the **POST** method to send form data to a URL.

Save the following script as login.html

<html>

<body>

<formaction="http://localhost:5000/login"method="post">

<p>Enter Name:</p>

<p><inputtype="text"name="nm"/></p>

<p><inputtype="submit"value="submit"/></p>

</form>

</body>

</html>

Now enter the following script in Python shell.

from flask importFlask, redirect,url\_for, request

app=Flask(\_\_name\_\_)

@app.route('/success/<name>')

def success(name):

return'welcome %s'% name

@app.route('/login',methods=['POST','GET'])

def login():

ifrequest.method=='POST':

user=request.form['nm']

return redirect(url\_for('success',name= user))

else:

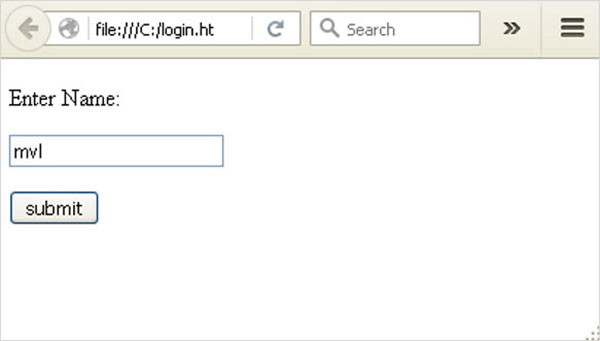
user=request.args.get('nm')

return redirect(url\_for('success',name= user))

if \_\_name\_\_ =='\_\_main\_\_':

app.run(debug =True)

After the development server starts running, open **login.html** in the browser, enter name in the text field and click **Submit**.

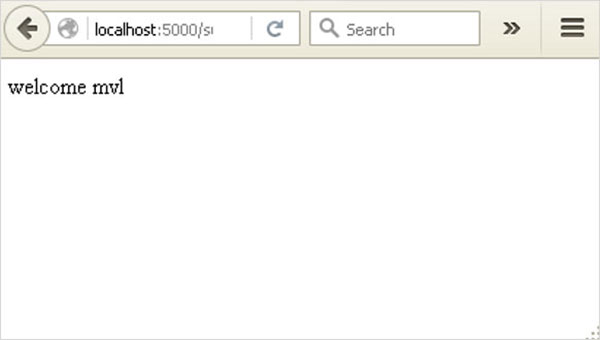


Form data is POSTed to the URL in action clause of form tag.

**http://localhost/login** is mapped to the **login()** function. Since the server has received data by **POST** method, value of ‘nm’ parameter obtained from the form data is obtained by −

user = request.form['nm']

It is passed to **‘/success’** URL as variable part. The browser displays a **welcome** message in the window.



Change the method parameter to **‘GET’** in **login.html** and open it again in the browser. The data received on server is by the **GET** method. The value of ‘nm’ parameter is now obtained by −

User = request.args.get(‘nm’)

Here, **args** is dictionary object containing a list of pairs of form parameter and its corresponding value. The value corresponding to ‘nm’ parameter is passed on to ‘/success’ URL as before.

**What is Python?**

Python is a popular programming language. It was created in 1991 by Guido van Rossum.

It is used for:

* web development (server-side),
* software development,
* mathematics,
* system scripting.

**What can Python do?**

* Python can be used on a server to create web applications.
* Python can be used alongside software to create workflows.
* Python can connect to database systems. It can also read and modify files.
* Python can be used to handle big data and perform complex mathematics.
* Python can be used for rapid prototyping, or for production-ready software development.

**Why Python?**

* Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
* Python has a simple syntax similar to the English language.
* Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
* Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
* Python can be treated in a procedural way, an object-orientated way or a functional way.

Good to know

* The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.
* In this tutorial Python will be written in a text editor. It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

Python Syntax compared to other programming languages

* Python was designed to for readability, and has some similarities to the English language with influence from mathematics.
* Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.
* Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

## Python Install

Many PCs and Macs will have python already installed.

To check if you have python installed on a Windows PC, search in the start bar for Python or run the following on the Command Line (cmd.exe):

C:\Users\Your Name>python --version

To check if you have python installed on a Linux or Mac, then on linux open the command line or on Mac open the Terminal and type:

python --version

If you find that you do not have python installed on your computer, then you can download it for free from the following website: <https://www.python.org/>

## Python Quickstart

Python is an interpreted programming language, this means that as a developer you write Python (.py) files in a text editor and then put those files into the python interpreter to be executed.

The way to run a python file is like this on the command line:

C:\Users\Your Name>python helloworld.py

Where "helloworld.py" is the name of your python file.

Let's write our first Python file, called helloworld.py, which can be done in any text editor.

helloworld.py

print("Hello, World!")

Simple as that. Save your file. Open your command line, navigate to the directory where you saved your file, and run:

C:\Users\Your Name>python helloworld.py

The output should read:

Hello, World!

Congratulations, you have written and executed your first Python program.

## The Python Command Line

To test a short amount of code in python sometimes it is quickest and easiest not to write the code in a file. This is made possible because Python can be run as a command line itself.

Type the following on the Windows, Mac or Linux command line:

C:\Users\Your Name>python

From there you can write any python, including our hello world example from earlier in the tutorial:

C:\Users\Your Name>python  
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:04:45) [MSC v.1900 32 bit (Intel)] on win32  
Type "help", "copyright", "credits" or "license" for more information.  
>>> print("Hello, World!")

Which will write "Hello, World!" in the command line:

C:\Users\Your Name>python  
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:04:45) [MSC v.1900 32 bit (Intel)] on win32  
Type "help", "copyright", "credits" or "license" for more information.  
>>> print("Hello, World!")  
Hello, World!

Whenever you are done in the python command line, you can simply type the following to quit the python command line interface:

exit()

Execute Python Syntax

As we learned in the previous page, Python syntax can be executed by writing directly in the Command Line:

>>> print("Hello, World!")  
Hello, World!

Or by creating a python file on the server, using the .py file extension, and running it in the Command Line:

C:\Users\*Your Name*>python myfile.py

Python Indentations

Where in other programming languages the indentation in code is for readability only, in Python the indentation is very important.

Python uses indentation to indicate a block of code.

Example

if 5 > 2:  
  print("Five is greater than two!")

Python will give you an error if you skip the indentation:

Example

if 5 > 2:  
print("Five is greater than two!")

Comments

Python has commenting capability for the purpose of in-code documentation.

Comments start with a #, and Python will render the rest of the line as a comment:

Example

Comments in Python:

#This is a comment.  
print("Hello, World!")

Docstrings

Python also has extended documentation capability, called docstrings.

Docstrings can be one line, or multiline.

Python uses triple quotes at the beginning and end of the docstring:

Example

Docstrings are also comments:

"""This is a   
multiline docstring."""  
print("Hello, World!")

**CHAPTER**

### SYSTEM DESIGN:

**SYSTEM DESIGN**

## SYSTEM ARCHITECTURE

The below figure shows the process flow diagram or proposed work. First we collected the Cleveland Heart Disease Database from UCI website then pre-processed the dataset and select 16 important features.

Heart disease data base

Data - preprocessing

Features selection

Algorithms

Output

Fig 5.1: System Architecture

For feature selection we used Recursive feature Elimination Algorithm using Chi2 method and get 16 top features. After that applied ANN and Logistic algorithm individually and compute the accuracy. Finally, we used proposed Ensemble Voting method and compute best method for diagnosis of heart disease.

## MODULES

The entire work of this project is divided into 4 modules. They are:

* + 1. Data Pre-processing
    2. Feature
    3. Classification
    4. Prediction

## Data Pre-processing:

This file contains all the pre-processing functions needed to process all input documents and texts. First we read the train, test and validation data files then performed some preprocessing like tokenizing, stemming etc. There are some exploratory data analysis is performed like response variable distribution and data quality checks like null or missing values etc.

## Feature:

Extraction In this file we have performed feature extraction and selection methods from sci- kit learn python libraries. For feature selection, we have used methods like simple bag-of- words and n-grams and then term frequency like tf-tdf weighting. We have also used

20

word2vec and POS tagging to extract the features, though POS tagging and word2vec has not been used at this point in the project.

## Classification:

Here we have built all the classifiers for the breast cancer diseases detection. The extracted features are fed into different classifiers. We have used Naive-bayes, Logistic Regression, Linear SVM, Stochastic gradient decent and DNN Classifications classifiers from sklearn. Each of the extracted features was used in all of the classifiers. Once fitting the model, we compared the f1 score and checked the confusion matrix.

After fitting all the classifiers, 2 best performing models were selected as candidate models for heart diseases classification. We have performed parameter tuning by implementing GridSearchCV methods on these candidate models and chosen best performing parameters for these classifier.

Finally selected model was used for heart disease detection with the probability of truth. In Addition to this, we have also extracted the top 50 features from our term-frequency tfidf Vectorizer to see what words are most and important in each of the classes.

We have also used Precision-Recall and learning curves to see how training and test set performs when we increase the amount of data in our classifiers.

## Prediction:

Our finally selected and best performing classifier was algorithm which was then saved on disk with name final\_model.sav. Once you close this repository, this model will be copied to user's machine and will be used by prediction.py file to classify the Heart diseases

. It takes a news article as input from user then model is used for final classification output that is shown to user along with probability of truth.

## DATA FLOW DIAGRAM

The data flow diagram (DFD) is one of the most important tools used by system analysis. Data flow diagrams are made up of number of symbols, which represents system components. Most data flow modeling methods use four kinds of symbols: Processes, Data stores, Data flows and external entities.

These symbols are used to represent four kinds of system components. Circles in DFD represent processes. Data Flow represented by a thin line in the DFD and each data store has a unique name and square or rectangle represents external entities.

21

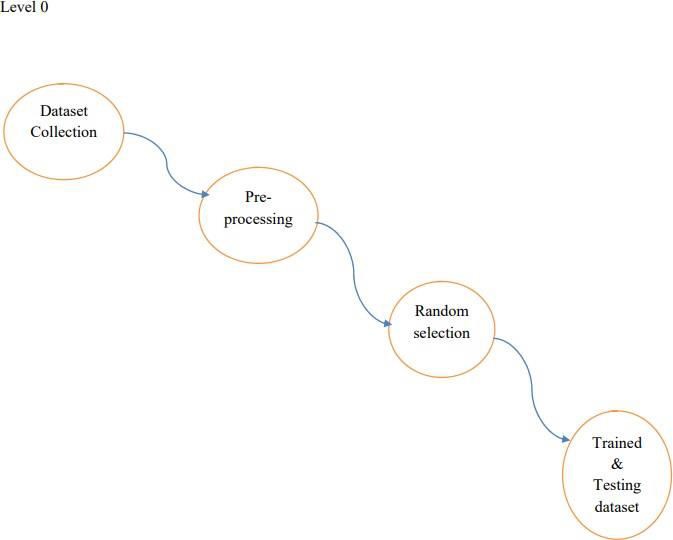


Fig 5.2: Data Flow diagram Level 0

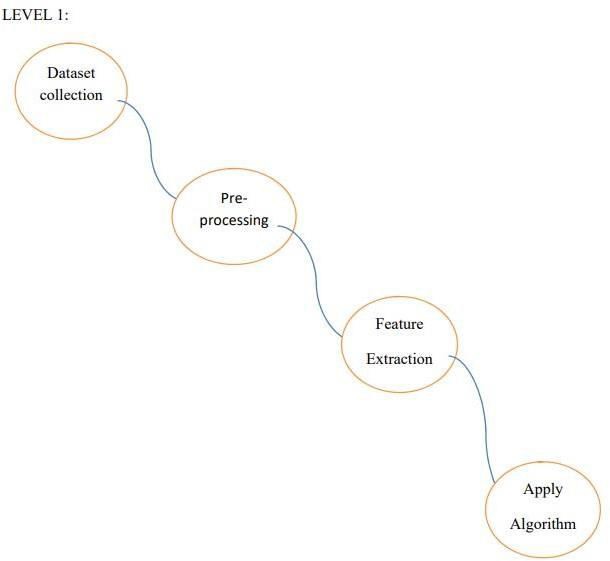


Fig 5.3: Data Flow Diagram Level 1

## UML DIAGRAMS

## Use-Case Diagram

A use case diagram is a diagram that shows a set of use cases and actors and their relationships. A use case diagram is just a special kind of diagram and shares the same common properties as do all other diagrams, i.e a name and graphical contents that are a projection into a model. What distinguishes a use case diagram from all other kinds of diagrams is its particular content.

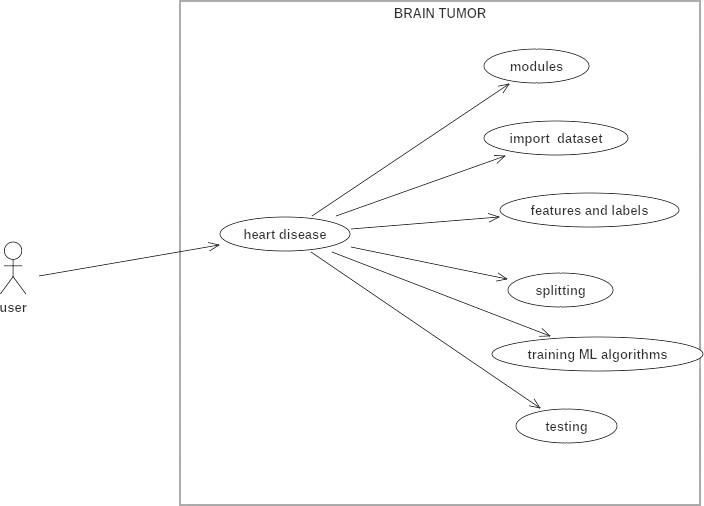
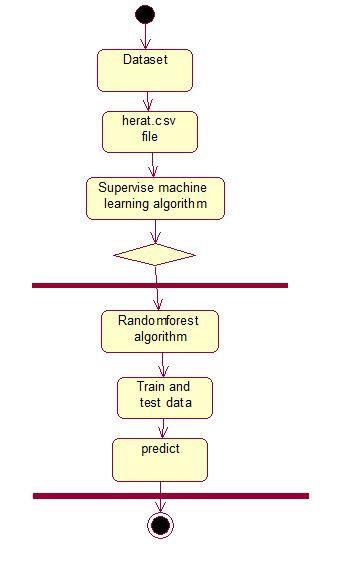


Fig 5.4: Use case Diagram

## Activity Diagram

An activity diagram shows the flow from activity to activity. An activity is an ongoing non- atomic execution within a state machine. An activity diagram is basically a projection of the elements found in an activity graph, a special case of a state machine in which all or most states are activity states and in which all or most transitions are triggered by completion of activities in the source.



## Sequence Diagram

Fig 5.5 Activity Diagram

A sequence diagram is an interaction diagram that emphasizes the time ordering of messages. A sequence diagram shows a set of objects and the messages sent and received by those objects. The objects are typically named or anonymous instances of classes, but may also represent instances of other things, such as collaborations, components, and nodes. We use sequence diagrams to illustrate the dynamic view of a system.

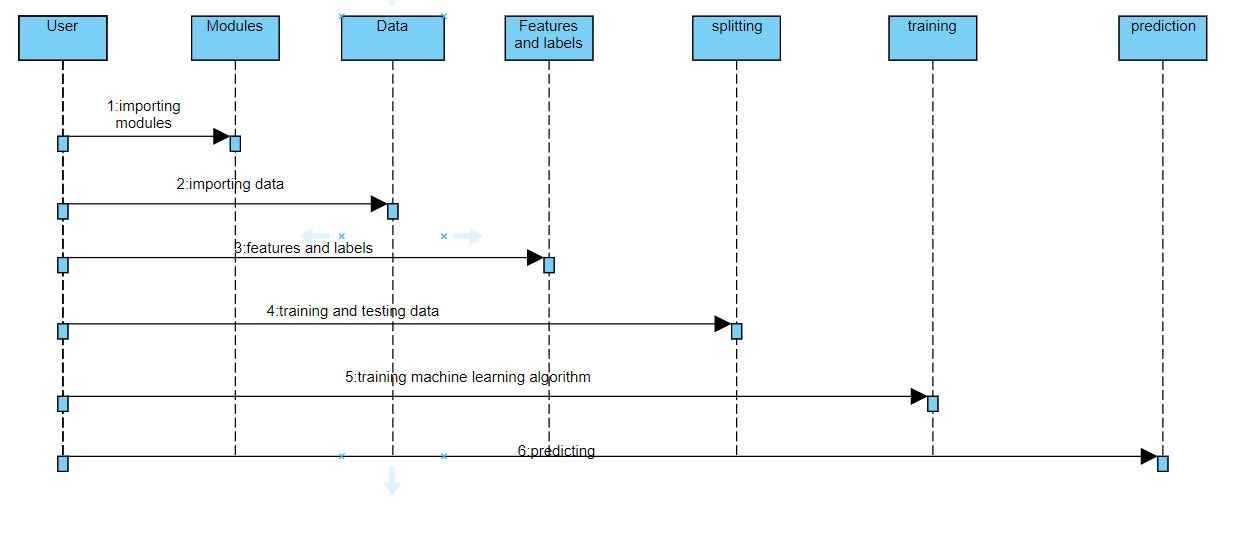


Fig 5.6 Sequence Diagram

**INPUT DESIGN AND OUTPUT DESIGN**

**INPUT DESIGN**

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

* What data should be given as input?
* How the data should be arranged or coded?
* The dialog to guide the operating personnel in providing input.
* Methods for preparing input validations and steps to follow when error occur.

**OBJECTIVES**

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

**OUTPUT DESIGN**

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system’s relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

* Convey information about past activities, current status or projections of the
* Future.
* Signal important events, opportunities, problems, or warnings.
* Trigger an action.
* Confirm an action.

**SYSTEM STUDY**

**FEASIBILITY STUDY**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

* ECONOMICAL FEASIBILITY
* TECHNICAL FEASIBILITY
* SOCIAL FEASIBILITY

**ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

### TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**SOCIAL FEASIBILITY**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

**OPERATIONAL FEASIBILITY**

Operational feasibility is the measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.[[14]](https://en.wikipedia.org/wiki/Feasibility_study#cite_note-SAD-Global_Enterprise-14)

The operational feasibility assessment focuses on the degree to which the proposed development project fits in with the existing business environment and objectives with regard to development schedule, delivery date, [corporate culture](https://en.wikipedia.org/wiki/Corporate_culture) and existing business processes.

To ensure success, desired operational outcomes must be imparted during design and development. These include such design-dependent parameters as reliability, maintainability, supportability, usability, producibility, disposability, sustainability, affordability and others. These parameters are required to be considered at the early stages of design if desired operational behaviours are to be realised. A system design and development requires appropriate and timely application of engineering and management efforts to meet the previously mentioned parameters. A system may serve its intended purpose most effectively when its technical and operating characteristics are engineered into the design. Therefore, operational feasibility is a critical aspect of systems engineering that needs to be an integral part of the early design phases.

### SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the

Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

**TYPES OF TESTS**

**Unit testing**

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

**Integration testing**

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. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

**Functional test**

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

**System Test**

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

**White Box Testing**

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

**Black Box Testing**

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

**6.1 Unit Testing:**

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

**Test strategy and approach**

Field testing will be performed manually and functional tests will be written in detail.

**Test objectives**

* All field entries must work properly.
* Pages must be activated from the identified link.
* The entry screen, messages and responses must not be delayed.

**Features to be tested**

* Verify that the entries are of the correct format
* No duplicate entries should be allowed
* All links should take the user to the correct page.

# 6.2 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

**6.3 Acceptance Testing**

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

**CONCLUSION**

In this project, we introduce about the heart disease prediction system with different classifier techniques for the prediction of heart disease. The techniques are DNN Classifications and Logistic Regression: we have analyzed that the DNN Classifications has better accuracy as compared to Logistic Regression. Our purpose is to improve the performance of the DNN Classifications by removing unnecessary and irrelevant attributes from the dataset and only picking those that are most informative for the classification task.

# CHAPTER FUTURE SCOPE

**FUTURE SCOPE**

As illustrated before the system can be used as a clinical assistant for any clinicians. The disease prediction through the risk factors can be hosted online and hence any internet users can access the system through a web browser and understand the risk of heart disease. The proposed model can be implemented for any real time application .Using the proposed model other type of heart disease also can be determined. Different heart diseases as rheumatic heart disease, hypertensive heart disease, ischemic heart disease, cardiovascular disease and inflammatory heart disease can be identified.

Other health care systems can be formulated using this proposed model in order to identify the diseases in the early stage. The proposed model requires an efficient processor with good memory configuration to implement it in real time. The proposed model has wide area of application like grid computing, cloud computing, robotic modeling, etc. To increase the performance of our classifier in future, we will work on ensembling two algorithms called DNN Classifications and Adaboost. By ensembling these two algorithms we will achieve high performance.

# CHAPTER

**REFERENCES**

**REFERENCES**

[1] P .K. Anooj, ―Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rulesǁ; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40. Computer Science & Information Technology (CS & IT) 59

[2] Nidhi Bhatla, Kiran Jyoti”An Analysis of Heart Disease Prediction using Different Data Mining Techniques”.International Journal of Engineering Research & Technology

[3] Jyoti Soni Ujma Ansari Dipesh Sharma, Sunita Soni. “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”.

[4] Chaitrali S. Dangare Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques” International Journal of Computer Applications (0975 – 888)

[5] Dane Bertram, Amy Voida, Saul Greenberg, Robert Walker, “Communication,

Collaboration, and Bugs: The Social Nature of Issue Tracking in Small, Collocated Teams”.

[6] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, ―Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithmǁ; International Journal of Engineering Science and Technology, Vol. 2(10), 2010.

[7] Ankita Dewan, Meghna Sharma,” Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification”, 2nd International Conference on Computing for Sustainable Global Development IEEE 2015 pp 704-706. [2].

[8] R. Alizadehsani, J. Habibi, B. Bahadorian, H. Mashayekhi, A. Ghandeharioun, R. Boghrati, et al., "Diagnosis of coronary arteries stenosis using data mining," J Med Signals Sens, vol. 2, pp. 153-9, Jul 2012.

[9] M Akhil Jabbar, BL Deekshatulu, Priti Chandra,” Heart disease classification using

nearest neighbor classifier with feature subset selection”, Anale. Seria Informatica, 11, 2013

[10]Shadab Adam Pattekari and Asma Parveen,” PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES”, International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624, Vol 3, Issue 3, 2012, pp 290-294.

[11]C. Kalaiselvi, PhD, “Diagnosis of Heart Disease Using K-Nearest Neighbor Algorithm

of Data Mining”, IEEE, 2016

[12] Keerthana T. K., “ Heart Disease Prediction System using Data Mining Method”, International Journal of Engineering Trends and Technology”, May 2017.

[13] Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, ELSEVIER.

Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Prediction Using Machine Learning and Data Mining July 2017, pp.2137-2159.