ABSTRACT

These days, wellbeing disease are expanding step by step as a result of life vogue, inherited. Particularly, cardiopathy has turned into a great deal of basic as of late .for example lifetime of people is in peril. Each individual has totally extraordinary qualities for power per unit region, cholesterol and essential sign. Anyway per restoratively attempted outcomes the customary estimations of power per unit territory is 120/90, cholesterol is and essential sign is seventy two. This paper gives the study with respect to totally extraordinary arrangement systems utilized for anticipating the opportunity dimension of each individual bolstered age, sexual orientation, constrain per unit zone, cholesterol, beat rate. We will utilize naïve bayes and image processing to predict the heart disease efficiently.

CHAPTER 1

INTRODUCTION

As the largest single cause of death on the planet, cardiovascular disease (CVD) in all its forms is an important and life or death matter. CVD is not a single disease, but a cluster of diseases and injuries that affect the cardiovascular system (the heart and blood vessels). These are most commonly diseases of the heart and of the blood vessels of the heart and brain. In general, they affect people in later life (with incidence rising sharply after the 30- 44 age range), although, according to a leading cardiologist, by around 35 years old, most who will get a form of CVD already have the beginnings of the disease. As mentioned, CVD is actually a collection of diseases affecting the cardiovascular system. These include: coronary heart disease; angina stroke; rheumatic heart disease; congenital heart disease; peripheral arterial disease; aortic aneurysm and dissection; deep vein thrombosis; and other, less common, cardiovascular diseases.

1.1 CORONARY HEART DISEASE

Coronary heart disease (CHD), also called coronary artery disease (CAD) and atherosclerotic heart disease, is the end result of the accumulation of atheromatous plaques5 within the walls of the arteries that supply the myocardium (the muscle of the heart). While the symptoms and signs of coronary heart disease are noted in the advanced state of disease, most individuals with coronary heart disease show no evidence of disease for decades as the disease progresses before the first onset of symptoms, often a "sudden" heart attack, finally arise. After decades of progression, some of these atheromatous plaques may rupture and (along with the activation of the blood clotting system) start limiting blood flow to the heart muscle. The disease is the most common cause of sudden death

1.1.2 ANGINA

The pain associated with very advanced CHD is known as angina, and usually presents as a sensation of pressure in the chest, arm pain, jaw pain, and other forms of discomfort.
The word discomfort is preferred over the word pain for describing the sensation of angina, because it varies considerably among individuals in character and intensity and most people do not perceive angina as painful, unless it is severe. Angina is essentially a cramp in the heart muscle .

1.1.3 STROKE

A stroke is an acute neurological injury whereby the blood supply to a part of the brain is interrupted, either by arterial blockage or rupture (haemorrhage). The part of the brain perfused by a blocked or burst artery can no longer receive oxygen carried by the blood; brain cells are therefore damaged or die (become necrotic), impairing the function of that part of the brain. Stroke can cause permanent neurological damage or death if not promptly diagnosed and treated.

1.1.4 CONGENITAL HEART DISEASE

Congenital heart disease is a broad term that can describe a number of different abnormalities affecting the heart, all of which are abnormalities of the heart's structure and function caused by abnormal or disordered heart development before birth. In some cases, such as coarctation of the aorta, it may not present itself for many years and a few lesions such as a small ventricular septal defect (VSD) may never cause any problems and are compatible with normal physical activity and a normal life span. Some congenital heart diseases can be treated with medication alone, while others require one or more surgeries. The risk of death from congenital heart disease surgery in the USA has dropped from approximately 30% in the 1970s to less than 5% in most cases today. Box 1 shows some of the major congenital heart diseases, taken from the British Heart Foundation (BHF)

1.2 NEED OF STUDY ON HEART DISEASE

CVDs are the number 1 cause of death globally: more people die annually from CVDs than from any other cause. An estimated 17.9 million people died from CVDs in 2016, Representing 31% of all global deaths. Of these deaths, 85% are due to heart attack and stroke. A Report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee

global burden of CHD, over 60% is in developing countries. India, China and Russia have the largest death tolls due to CHD, with Lithuania and Bhutan also having very high rates These

figures are mirrored by those for stroke, with the same three countries having the largest number of deaths per annum.

In terms of global mortality, CVD is the largest single contributor to the 57 million deaths registered by the WHO in 200217. In fact, CVD causes more deaths globally than cancer and HIV combined. In current terms, nearly 100 times as many people a year are killed by CVD than by war

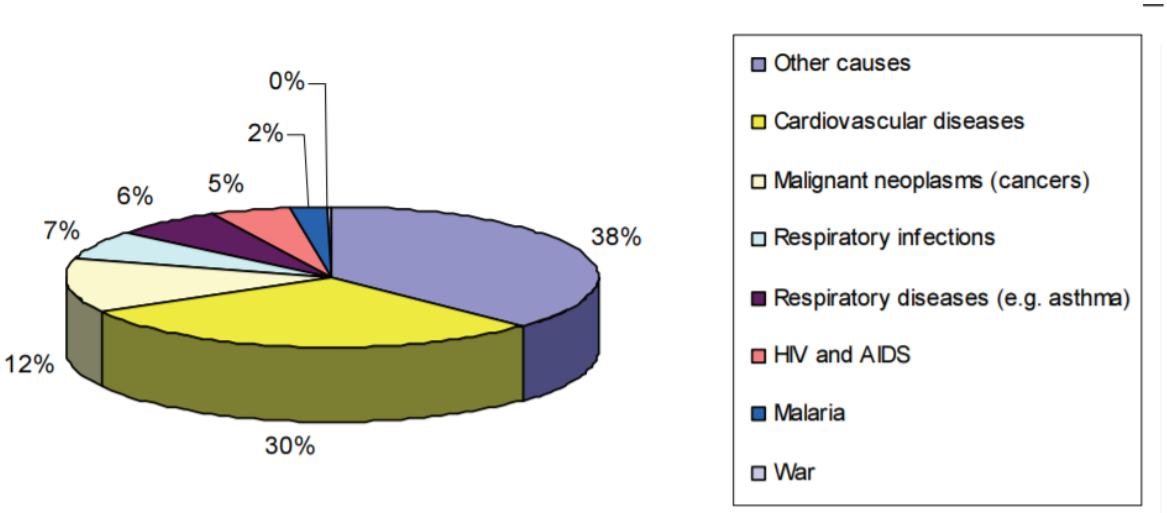


Figure 1.1: Global mortality in 2002

1.3 Reasons for CVD

As CVD is such a complex collection of diseases, unsurprisingly there are a number of risk factors involved. There is also the added complication that many of the risk factors for CVD interact with one another. For example, both obesity and diabetes are major risk factors for the development of CVD, but obesity is also a risk factor for type II diabetes

1.3.1 Smoking

Smoking increases the risk of developing coronary heart disease is by 2–4 times. Smoking is a powerful independent risk factor for sudden cardiac death in patients with coronary heart disease; where smokers have about twice the risk of non-smokers.22 Exposure to other people's smoke increases the risk of heart disease even for non-smokers; the British Heart Foundation estimates that regular exposure to second hand smoke can increase the risk of CHD by up to 25%.23 Cigarette smoking interacting with other risk factors greatly increases CHD risk.

1.3.2 Obesity

Obesity, particularly in those with excess fat around the waist, increases the chance of developing CHD even if no other risk factors are present. Excess weight increases strain on the heart, raises blood pressure and blood cholesterol and triglyceride levels, and lowers HDL cholesterol levels. All of these factors can increase the risk of atherosclerosis and thrombolytic embolism. It also increases the risk of developing type II diabetes, another risk factor for CVD. The WHO estimated that in 2005, around 1.6 billion people were overweight, with 400 million people obese

1.3.3 Diabetes

Diabetes is a disease that affects an individual's ability to maintain an appropriate blood glucose level. The disease has two forms, Type I and Type II diabetes. Type I diabetes is also known as insulin dependent diabetes and is caused by the body not producing any insulin. Type II diabetes is the more common form of the disease, and occurs when the body either does not produce enough insulin or cells do not process the insulin properly.28 Both of these forms can lead to an increased risk of CVD via increased cholesterol levels, hypertension and atherosclerosis.29 Insulin resistance is also related to CHD

1.3.4 High LDL cholesterol

The basic method by which LDL cholesterol (low density lipoprotein cholesterol) increases the risk of CVD is by increasing the fatty deposits in blood vessels, leading to atherosclerosis. Recently, research has suggested that this is more of an active process than previously thought, with LDL cholesterol actually activates endothelial cells to express adhesion molecules that speed the process of atherosclerosis

1.3.5 High blood pressure

High blood pressure (hypertension) is strongly linked to both cardiac diseases and those of the vascular system. High blood pressure affects the heart by causing it to thicken and stiffen as it has to work harder to pump blood; this can lead to heart attacks. The effect on the vascular system is one of pressure on vasculature walls, leading to aneurisms and stroke.

High blood pressure is a very common problem in developed countries, with one in four adults in the USA diagnosed with hypertension, although this represents a drop in the numbers since the 1980s when prevalence was around one in two.33 It is expected that global hypertension rates will rise in the next 20 years, reaching 1.5 billion adults globally, up to 1 /3 from ¼ of the world population

CHAPTER II
LITERATURE SURVEY

[1] Evaluation of Artery Visualizations for Heart Disease Diagnosis – 2011

- Michelle A. Borkin

Heart disease is the number one killer in the United States, and finding indicators of the disease at an early stage is critical for treatment and prevention. In this paper we evaluate visualization techniques that enable the diagnosis of coronary artery disease. A key physical quantity of medical interest is endothelial shear stress (ESS). Low ESS has been associated with sites of lesion formation and rapid progression of disease in the coronary arteries. Having effective visualizations of a patient's ESS data is vital for the quick and thorough non-invasive evaluation by a cardiologist. We present a task taxonomy for hemodynamics based on a formative user study with domain experts. Based on the results of this study we developed HemoVis, an interactive visualization application for heart disease diagnosis that uses a novel 2D tree diagram representation of coronary artery trees. We present the results of a formal quantitative user study with domain experts that evaluates the effect of 2D versus 3D artery representations and of color maps on identifying regions of low ESS. We show statistically significant results demonstrating that our 2D visualizations are more accurate and efficient than 3D representations, and that a perceptually appropriate color map leads to fewer diagnostic mistakes than a rainbow color map.

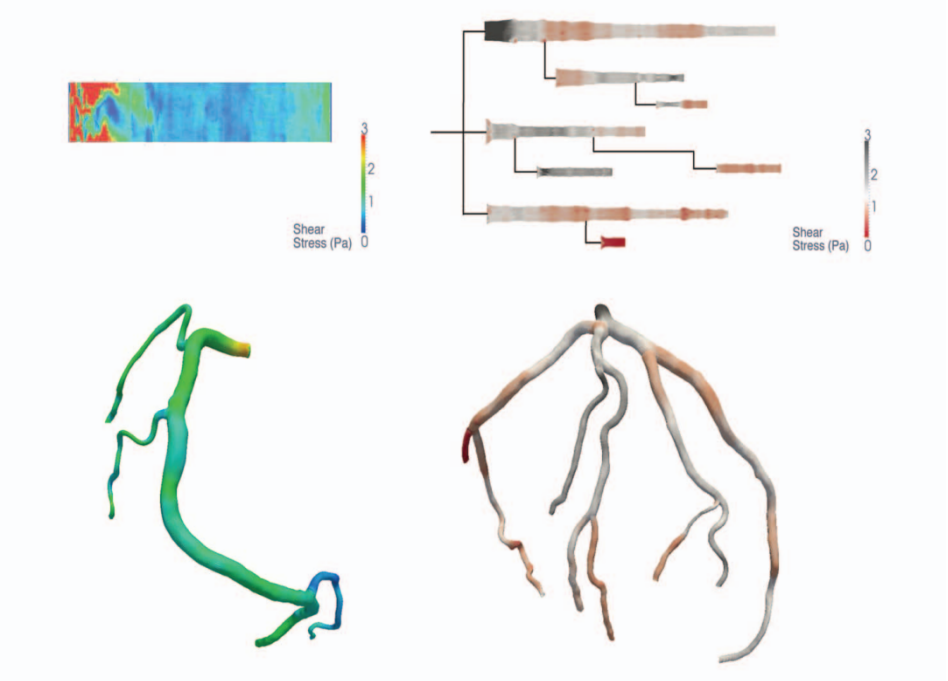


Figure 2.1 :Left: Traditional 2D projection (A) of a single artery, and 3D representation (C) of a right coronary artery tree with a rainbow color map. Right: 2D tree diagram representation (B) and equivalent 3D representation (D) of a left coronary artery tree with a diverging color map.

Current visualizations techniques, as shown in Fig. 1 (left), use either a 2D cylindrical projection of a single artery or a 3D representations of the coronary artery tree. In both cases, ESS is mapped to the surface using a color encoding, typically with a rainbow (or "spectrum") color map. Both representations have their advantages and disadvantages: 2D allows one to see all the data at once, but anatomical information is lost both in the shape of a vessel and in how each individual vessel connects to other branches. 3D preserves the anatomical structure, but introduces issues of occlusion and requires human interaction to rotate the model in order to see all the data. The fundamental visualization issue is how to display a scalar quantity

[2] Algorithm of Heart Sound Segmentation and Feature Extraction -2013

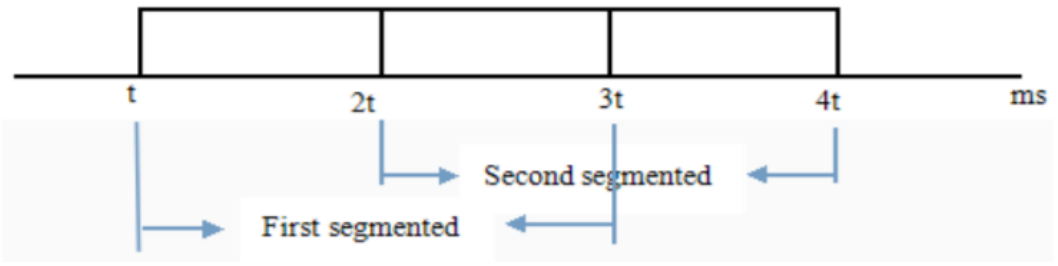
- Liping Liu, Weilian Wang

In this paper, an algorithm about heart sound signals segmentation was described, in which only heart sound signals were used. The algorithm was based on the normalized average Shannon energy of heart sound signals. 20 cases of normal and 20 cases of abnormal heart sound signals of heart valve disease selected randomly from the database of the Institute of Research Clinique of Montréal were used for testing. The features of each component, such as AR spectrum energy in systolic period were extracted after segmentation. The energy of high frequency is generally less than -95db to normal signal. Reversely, the energy of high frequency is generally more than -95db for abnormal signal. This feature is valuable in diagnosis of congenital heart diseases and other cardiac diseases. One of the most common methods is using electrocardiogram (ECG signal) as reference to identify first and second sounds. The segmentation of heart sound by this method is very accurate, but large amount of tasks could be done. Based on the algorithm, every cycle of the PCG signals is separated into four parts, such as the first heart sound, the systolic period, the second heart sound and the diastolic period. Both normal and abnormal heart sound recordings are investigated through calculating the energy of the systolic period.

x_{norm} (i) according to the (1)

x_{norm} (i)=x(i)/max(x(i))

Secondly, the signal x_{norm} (i) are segmented in continues 2t time in ms. As shown in Fig. 1, the signals of two sections, first and second part, are overlap in one t.



After segmentation, each section NASE is computed according to (2) ,

$$E_{norm} = (-1/N) * \sum_{i=1}^{N} x_{norm}^{2}(i) \log x_{norm}^{2}(i) \quad (2)$$

Where N is the number of signal points in a window.

Thirdly, according to (2) the normalized Shannon energy can be obtained normalized average Shannon energy P(t) according to (3).

$$P(t) = (E_{norm}(t) - M (E_{norm}(t))) / S (E_{norm}(t)) \quad (3)$$

Where E_{norm} (t) is the average Shannon energy for window t,

M (E_{norm} (t)) is the mean value of E_{norm} (t) ,and S (E_{norm} (t)) is the standard deviationof E_{norm} (t)

[3] Cardiac Cycle Phase Detection in Echocardiography Images Using ANN

- S. Sundaramurthy and L. P. Devi

The cardiac cycle is the combination of two phases of the heart i.e. the actions that happen in between the pumping of the heart. This could also be said as two main states of heart, “diastole and systole”. The diastole is the process of blood filled into a chamber of the heart and the systole is the process of blood flowing out from the chamber of the heart. The aim here is to identify the states of the heart and the volume during the cardiac cycle function occurs. In this paper the left ventricle is considered for the project as due to its importance in pumping the oxygenated blood (pure blood) to all parts of the body. This had done by identifying the anatomical information of the heart with the dataset of both the normal and infant cardiac pathology images of the heart.

This extracts the information about the given image and also differentiates them under the two categories either the heart left ventricle is in the diastolic state or under the systole state. For identifying this sate the mitral valve position of the heart is considered .The image here consists of noise and is removed by using the median filter as the first scenario and wavelet transform for the edge detection in the second scenario. To extract the data from the image is the third scenario, the mean and SD has been calculated here. To classify the two states of the heart, the Artificial Neural Network (ANN) is used. This is fourth scenario. By training the Neural network classifiers the heart images are classified as diastole and systole .

The main reason of underlying this study is, if the bicuspid valve or the mitral valve annulus fails to function properly then the functioning of the heart will be compromised in resulting in the heart attack, which is the main reason of the occurrence of the cardiovascular disease [maximum reason for the cause of the death], Heart failure, Endocarditis, Rheumatic fever, Stenosis, Infective endocarditis . In order to effectively work on the project, the mitral valve position of the heart is taken under the study. These images are the frames of the ultrasound images of the heart of the patients were used for the case. The dataset of 130 images were used, in which the both normal and abnormal cases of the images were also taken under consideration. Mostly the image with the infract pathology has been chosen as because it will

cause the mitral value to not close properly and make the mitral valve regurgitation. Also the dead tissue will show the compromise of the mitral valve position and the process.

[4] **Pulmonary Artery-Vein Classification in CT Images Using Deep Learning**

- Pietro Nardelli,Daniel Jimenez-Carretero,David Bermejo-Pelaez

Recent studies show that pulmonary vascular diseases may specifically affect arteries or veins
through different physiologic mechanisms. To detect changes in the two vascular trees, physicians manually analyze the chest computed tomography (CT) image of the patients in search of abnormalities. This process is time-consuming, difficult to standardize and thus not feasible for large clinical studies or useful in real-world clinical decision making. Therefore, automatic separation of arteries and veins in CT images is becoming of great interest, as it may help physicians accurately diagnose pathological conditions. In this work, we present a novel, fully automatic approach to classifying vessels from chest CT images into arteries and veins. The algorithm follows three main steps: first, a scale-space particles segmentation to isolate vessels; then a 3D convolutional neural network (CNN) to obtain a first classification of vessels; finally, graph-cuts (GC) optimization to refine the results. To justify the usage of the proposed CNN architecture, we compared different 2D and 3D CNNs that may use local information from bronchus- and vessel-enhanced images provided to the network with different strategies. We also compared the proposed CNN approach with a Random Forests (RF) classifier.

The methodology was trained and evaluated on the superior and inferior lobes of the right lung of eighteen clinical cases with non-contrast chest CT scans, in comparison with manual
classification. The proposed algorithm achieves an overall accuracy of 94%, which is higher than the accuracy obtained using other CNN architectures and RF. Our method was also validated with contrast-enhanced CT scans of patients with Chronic Thromboembolic

Pulmonary Hypertension (CTEPH) to demonstrate that our model generalizes well to contrast-enhanced modalities.

[5] Detection and Localization of Coronary Artery Stenotic Segments using Image Processing

-N. Mohan,Vishnukumar S

Coronary Artery Disease (CAD) is one of the emerging causes of death all over the world. Coronary angiography techniques for the detection of CAD lead to various complications like artery-dissection, arrhythmia and even death. In this paper we propose an image processing method for detecting and localizing the stenosis regions in the artery. The method is applied on the Digital Imaging and Communications in Medicine (DICOM)) image of the heart for the detection of stenosis. The region of interest is considered as the coronary arteries and is segmented using vessel enhancement diffusion filter along with dilation and erosion morphology. Centerlines of the segmented arteries are extracted using a fast marching based method. Detection of the stenosis is done by estimating the vessel diameter at each centerline location. The abrupt reduction in the diameter of the arteries is a sign of stenosis.

The proposed system works on a DICOM image which gives a view of the coronary arteries. The image actually has the arteries and other portions of the heart. In order to process the arteries for detecting the stenosis, the arteries need to be segmented. The segmentation of the coronary arteries is done through vessel enhancement diffusion filter along with the application of morphological operations. As a preprocessing step the input image undergoes segmentation and vessel enhancement. The segmentation process consists of segmenting the region of interest which is considered as the arteries from the original image. After segmentation, image enhancement is done to enhance the vessels in the image.

Figure 2.2: The Original Image

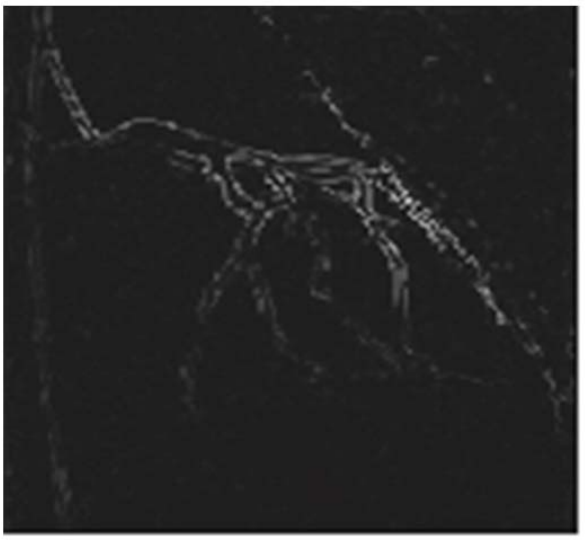


Figure 2.3: Image after segmentation

Figure 2.4:Enhanced vessel image

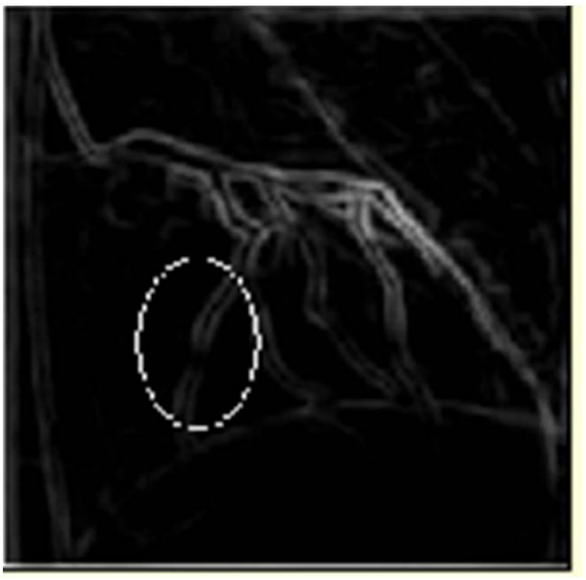


Figure 2.5 : Detected Stenosis in theenhanced vessel image

[6] Hidden pattern analysis for heart disease classification

-Chavan Krushna D,Kale Abhijeet A ,Kulkarni Swapnil P, Sayyed Ajmeer D

Prediction of Heart Disease using technique of Data Mining is effective but there is loss of Accuracy by using the image processing as additional processing for more Accuracy . In Proposed System we are using the algorithms like Decision Tree, Nueral Network and the Naive Bayes in the data mining and in the image Processing we are using the popular algorithms like Local Binary Pattern. The research result shows prediction accuracy of 99 Percent. Data mining enable the health sector to predict patterns in the datasets.

Here we use image processing for comparing the ECG, CT scan, Angiography, etc, reports and finding the more accurate results. The main objective of this paper is to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. It can solve complicated queries for detecting heart disease and thus assist medical practitioners to make smart clinical decisions which traditional decision support systems were not able to. By providing efficient treatments, it can help to reduce costs of treatment.

[7] An Integrated Architecture for Prediction of Heart Disease from the Medical Database

-Gunasekar Tangarasu, Kayalvizhi Subramanian and P. D. D. Dominic

A database is a collection of data organized for storage, access and retrieval. With increasing growth in big data, especially in healthcare and biomedical communities, the techniques to analyze the medical data tends to benefit the patients by detecting the disease early. However, with the advent of incomplete data in such medical datasets, the quality tends to reduce. In addition to this, each region has its own unique disease characteristics, which further reduces the prediction quality. To overcome the difficulty in processing the medical datasets with incomplete data, the proposed method initially reconstructs the missing or incomplete data. To improve the processing capability of automated disease prediction in an uncertain environment, an integrated architecture is proposed. It controls the processing capability of medical datasets in an uncertain environment. This integrated diagnostic model generates the hesitant fuzzy based decision tree algorithm using genetic classification. The architecture is designed to process both the structure and unstructured data sets. The new and innovative prediction methods are projected in this research to predict heart disease from the medical database in a faster manner.

Our current research work provides the comprehension into the design of new framework using combined of Hesitant Fuzzy based Decision tree Algorithm and Genetic algorithm for predicting heart disease. The vision of this research is to identify and develop methods and procedures for predicting heart disease that assist medical practitioners in an efficient way for people getting longer life in this world.

Limitations

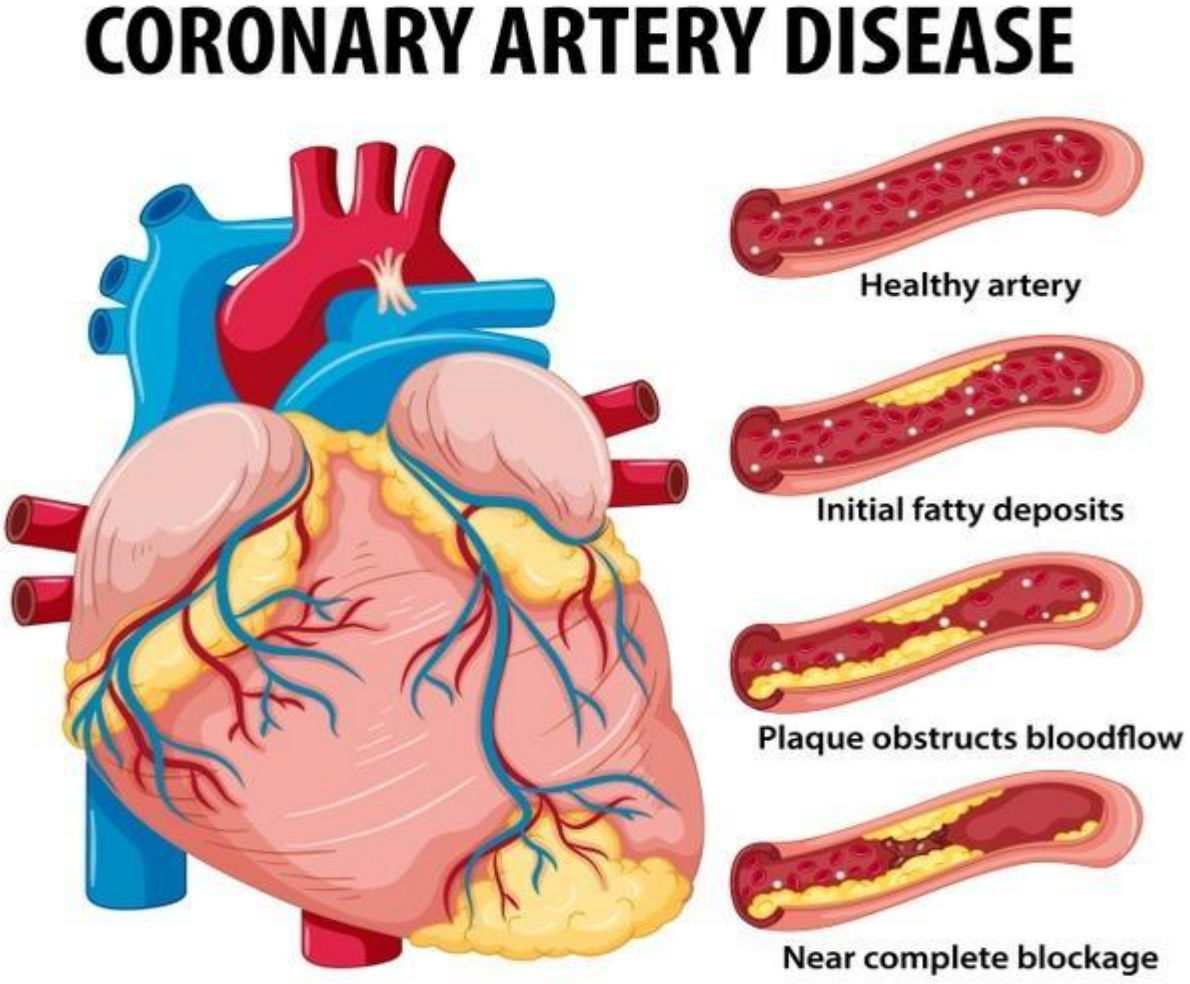
Hesitant Fuzzy based Decision tree Algorithm and other techniques are proposed in this research capable of tolerating a certain level of noise data. However, the noise data sometimes lead to misleading the result when people health conditions are more complex.

CHAPTER 3
HEART DISEASE PREDICTION TECGNIQUE

INTRODUCTION

Naive Bayes is a machine learning algorithm we use to solve classification problems. It is based on the Bayes Theorem. It is one of the simplest yet powerful ML algorithms in use and finds applications in many industries. Suppose you have to solve a classification problem and have created the features and generated the hypothesis, but your superiors want to see the model. You have numerous data points (lakhs of data points) and many variables to train the dataset. The best solution for this situation would be to use the Naive Bayes classifier, which is quite faster in comparison to other classification algorithms.

Cardiovascular diseases are one of the most common diseases of the modern world. There are certain things that increase a person's chances of getting cardiovascular disease. Cardiovascular disease (CVD) refers to any condition that affects the heart. Many CVD patients have symptoms such as chest pain (angina) and fatigue. which occur when the heart isn't receiving adequate oxygen. As per a survey nearly 50 percent of patients, however, have no symptoms until a heart attack occurs. A number of factors have been shown to increase the risk of developing CVD. Some of these are family history of cardiovascular disease, High levels of LDL (bad) cholesterol, Low level of HDL (good) cholesterol, Hypertension, High fat diet, Lack of regular exercise and Obesity. With so many factors to analyze for a diagnosis of cardiovascular disease, physicians generally make a diagnosis by evaluating a patient's current test results. Previous diagnoses made on other patients with the same results are also examined by physicians. These complex procedures are not easy. Therefore, a physician must be experienced and highly skilled to diagnose cardiovascular disease in a patient.Data mining has been heavily used in the medical field, to include patient diagnosis records to help identify best practices. The difficulties posed by prediction problems have resulted in a variety of problem-solving techniques. For example, data mining methods comprise Artificial Neural Networks and Clustering Techniques (K-Means Clustering). It is difficult, however, to compare the accuracy of the techniques and determine the best one because their performance is data dependent. A few studies have compared data mining and statistical approaches to solve prediction problems. The comparison studies have mainly considered a specific data set.



RISKS IN EXISTING HEART DISEASE PREDICTION METHODS

Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.

This algorithm faces the 'zero-frequency problem' where it assigns zero probability to a categorical variable whose category in the test data set wasn't available in the training dataset. It would be best if you used a smoothing technique to overcome this issue.

Its estimations can be wrong in some cases, so you shouldn't take its probability outputs very seriously.If your test data set has a categorical variable of a category that wasn't present in the training data set, the Naive Bayes model will assign it zero probability and won't be able to make any predictions in this regard. This phenomenon is called 'Zero Frequency,' and you'll have to use a smoothing technique to solve this problem.

CHAPTER 4

HYBRID COHERENT TECHNIQUE OF PREDICTING CARDIO VASCULAR DISEASE USING NAIVE BAYES AND CNN WITH IMAGE PROCESSING

The life of a person is conformed with the heart beat, heart is located in between the lungs. This is the important organ in the body due to is importance of making all parts of the body to function with the pure blood and collect the impure blood from those organs . The cardiac heart failure is the most commonly affecting problem in the human body due to various reasons like, blood pressure, cholesterol, diabetes, obesity, etc,. The global cause of death among all people is now-a-days is due to the cardiovascular diseases, than others causes the most of the people die due to this diseases. Finding indicators of the disease at an early stage is critical for treatment and prevention. Some of the cardiovascular diseases that appear in body are: coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis and pulmonary embolism.

Today, almost all the sectors and in other fields get the aid of computerized systems. In the field of medical science there is a great demand for computer aided tools to facilitate many tasks. Many things that were done manually using traditional equipment have been replaced with automated systems. Modern medical science is looking for solution which could assist the doctors with any aspect of work using the new technology. So an efficient method is proposed to predict heart disease. The proposed system introduces an app called “Heart care” by which patients can scan and upload their angiogram results or enter ECG values and they will get the results .And also can consult near by doctors if diseased. So that they can send their results to the doctor by a chat with him and can find a solution. The proposed system eliminate the drawbacks of the existing system to a great extend, which includes Naive bayes

as the classification algorithm and CNN is used to train the user given data and produce accurate output which gives more accurate result.

Fig 4.1 HEART CARE logo

4.1 PROPOSED SYSTEM ARCHITECTURE

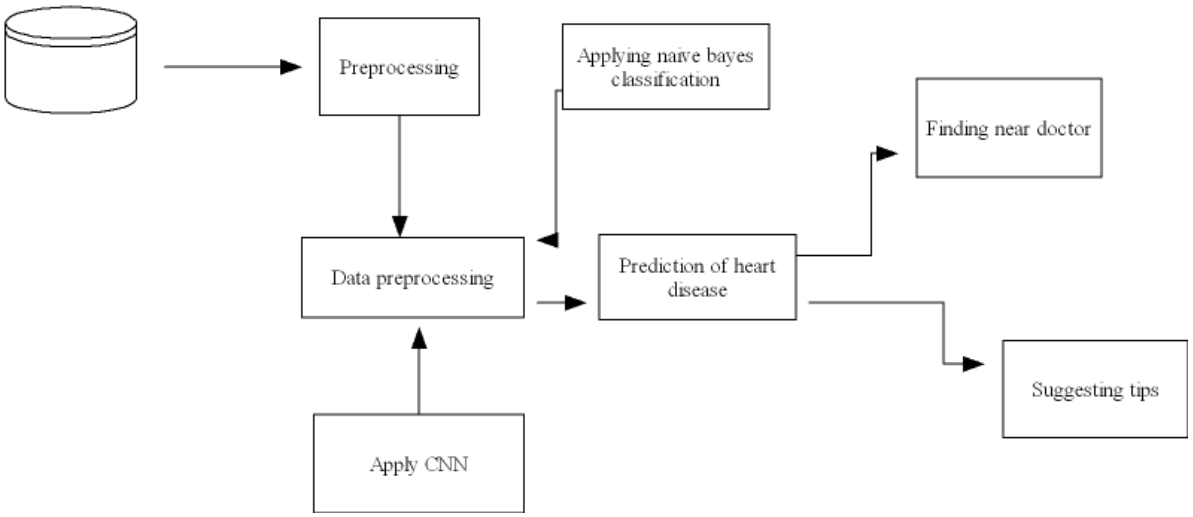


Image processing is the use of computer algorithms to perform image processing on digital images. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing.

• The main steps of proposed methodology to skin disease recognition are:

(1) Processing of the original image
(2) Feature extraction
(3) Classification based on Naive bayes
(4) Training CNN with image
(5) Predicting Disease
(6) If diseased find nearest doctor
(7) Else provide some health tips

4.2 ALGORITHM DESCRIPTION

In our project we use two algorithms, Naïve Bayes algorithm for heart disease prediction by value and Convolutional Neural Network for prediction by images.

4.2.1 Naïve Bayes algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems . It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.**
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles**.**

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

- **Naïve**: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the

fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

- **Bayes**: It is called Bayes because it depends on the principle of Bayes' Theorem.

Bayes' Theorem:

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

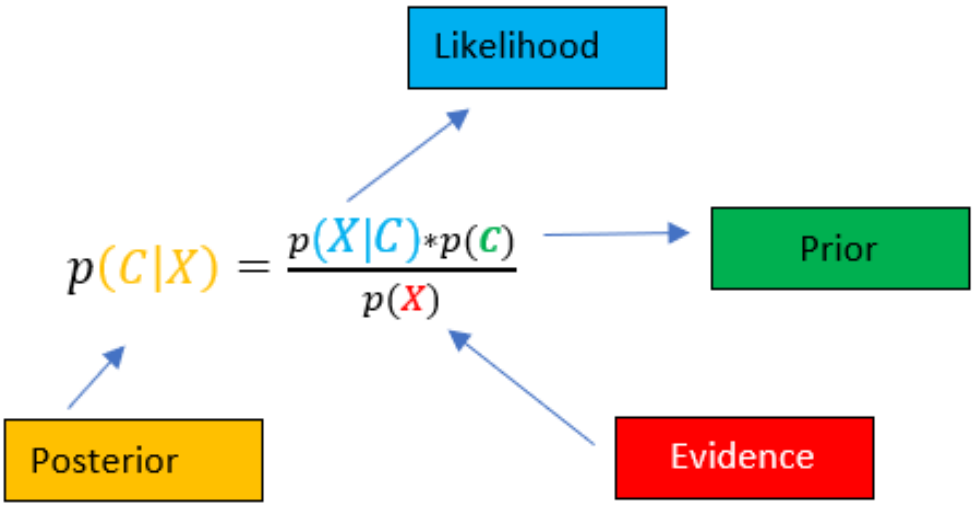


Figure 4.2 : Naive Bayes Classifier

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

Working of Naïve Bayes' Classifier:

The naıve Bayesian classifier, or simple Bayesian classifier, works as follows:

1. Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector, X = (x1, x2,..., xn), depicting n measurements made on the tuple from n attributes, respectively, A1, A2,..., An.

2. Suppose that there are m classes, C1, C2,..., Cm. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naıve Bayesian classifier predicts that tuple X belongs to the class Ci if and only if P(Ci/ X) > P(Cj/ X) for $1 \leq j \leq m, j \neq i$. Thus, we maximize P(Ci /X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

$$P(Ci\ /X) = \frac{P(X|Ci)P(Ci)}{P(X)}$$

3. As P(X) is constant for all classes, only P(X|Ci)P(Ci) needs to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is,

P(C1) = P(C2) = ⋯ = P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci). Note that the class prior probabilities may be estimated by P(Ci) = |C i,D|/|D|, where |C i,D| is the number of training tuples of class Ci in D.

4. Given data sets with many attributes, it would be extremely computationally expensive to compute

P(X|Ci). To reduce computation in evaluating P(X|Ci), the naıve assumption of class-conditional independence is made. This presumes that the attributes' values are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes).

Thus, $P(X|Ci) = \prod_{n} k = 1\ P(xk\ |Ci)$ ---------- --1

= P(x1|Ci) × P(x2|Ci) × ⋯ × P(xn|Ci).

We can easily estimate the probabilities P(x1|Ci), P(x2|Ci),..., P(xn|Ci) from the training tuples. Recall that here xk refers to the value of attribute Ak for tuple X.

For each attribute, we look at whether the attribute is categorical or continuous-valued. For instance, to compute P(X|Ci), we consider the following:

(a) If Ak is categorical, then P(xk |Ci) is the number of tuples of class Ci in D having the value xk for Ak , divided by |Ci,D|, the number of tuples of class Ci in D.

(b) If Ak is continuous-valued, then we need to do a bit more work, but the calculation is pretty straightforward. A continuous-valued attribute is typically assumed to have a Gaussian distribution with a mean μ and standard deviation σ, defined by

These equations may appear daunting, but hold on! We need to compute μCi and σCi , which are the mean (i.e., average) and standard deviation, respectively, of the values of attribute Ak for training tuples of class Ci . We then plug these two quantities into Eq. 1, together with xk , to estimate P(xk |Ci).

4.2.2 Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.CNN is a type of Neural Network where the mathematical operation used to find the relationship between the data is Convolution .Traditional neural network fails when coming to complex problems such as image classification, video classification, pattern recognition, etc. but CNN has achieved great success in these applications, yielding good accuracy.

CNN consists of mainly 4 Layers, viz. Convolutional layer (CONV), pooling layer (POOL), dropout and fully connected layers (FC). These layers together extract the features from the input data. The algorithm learns from the feature, where the features of interest are represented by each convolution filter.

Convolutional Layer

The main building block of CNN is the convolutional layer.The convolutional layer consists of small patches, which transforms the entire image based on the filter values. Equation (1) is the formula to create feature maps, i.e. the output from the convolutional layer.

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k]\ f[m-j, n-k] \qquad \text{-----(1)}$$

where f is the input image, h is the filter, *(m,n)* is the size of the resulting matrix generated.

Pooling Layer

The output from the convolution layer is passed on to a pooling layer where its size gets reduced without any loss of information. These 2-dimensional arrays are converted to a single dimensional vector using the flatten layer so that it can be fed to the neural network for classification.

Dropout

Dropout is a technique used to prevent a model from overfitting. Dropout works by randomly setting the outgoing edges of hidden units (neurons that make up hidden layers) to 0 at each update of the training phase.

Overfitting indicates that your model is too complex for the problem that it is solving, i.e. your model has too many features in the case of regression models and ensemble learning, filters in the case of Convolutional Neural Networks.

Fully Connected Layers

After the convolution and pooling layers we add a couple of fully connected layers to wrap up the CNN architecture.Fully Connected Layer is simply, feed forward neural networks. Fully Connected Layers form the last few layers in the network.The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

Back-Propagation

The neural network uses the back-propagation algorithm where the errors are back propagated to adjust the weights, thereby reducing the error (loss) function.

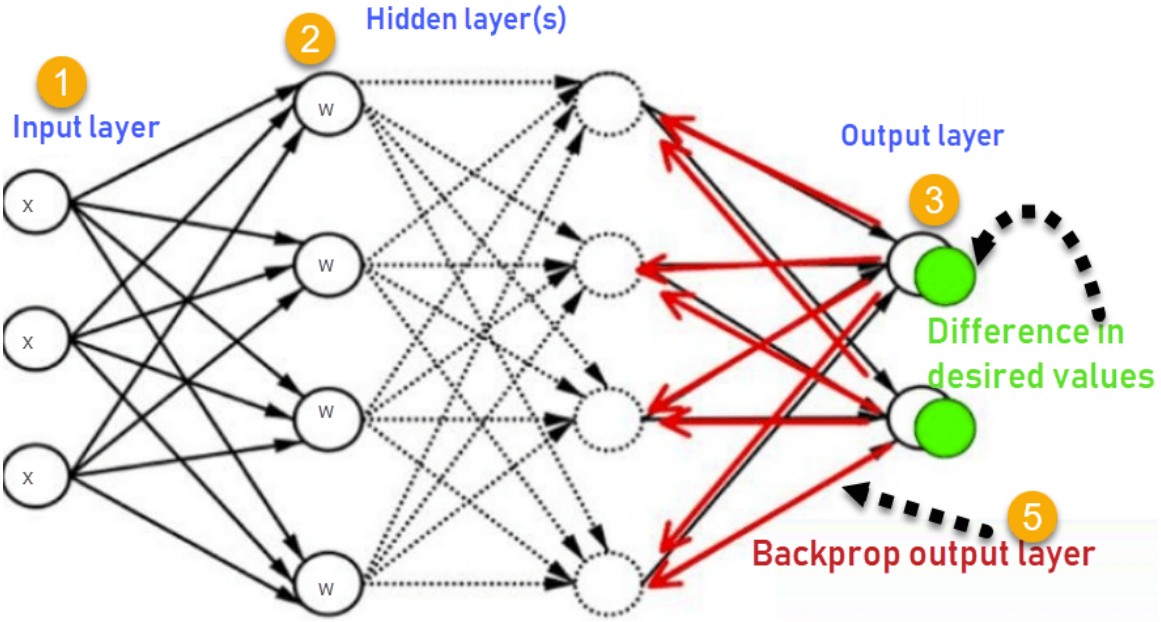


Figure 4.3: Backpropagation In Neural Network

The weight updating is done using the equation (2)

$$W_i = W_i + \Delta W_i \qquad \text{----------(2)}$$

Where W_i is the weight and ΔW_i is given by the delta rule as in (3)

$$\Delta W_i = n\frac{dE}{dW_i} x_i \qquad \text{----------(3)}$$

where n is the learning rate is the error function and x_i is the input.

4.3 REQUIREMENT ANALYSIS

Requirements analysis understands the user's requirements within the framework of the organization objective and the environment in which the system being installed. Requirements analysis is a software engineering task that bridges the gap between system level software allocation and software design. The first step in designing a system was to identify the underlying cause behind the problems in the existing system. Requirement study has been done to the gather the required information. Requirements analysis enables the system engineer to specify software function and performance, indicate software's interface with other system elements, and establish constraints that software must meet. Requirements analysis allows the software engineer to refine the software allocation and build models of the data, Requirements analysis provides the software designer with models that can be translated into data, architectural, interface and procedural design. Finally, the requirement specification provides the developer and the customer with the means to assess quality once software is built and behavioural domains that will be treated by software.

4.4 SOFTWARE LIFECYCLE

Requirements analysis enables the system engineer to specify software function and performance, indicate software's interface with other system elements, and establish constraints that software must meet. Requirements analysis allows the software engineer to refine the software allocation and build models of the data, functional and behavioral domains that will be treated by software. Requirements analysis provides the software designer with models that can be translated into data, architectural, interface, and procedural design. Finally, the requirement

specification provides the developer and the customer with the means to assess quality once software is built.

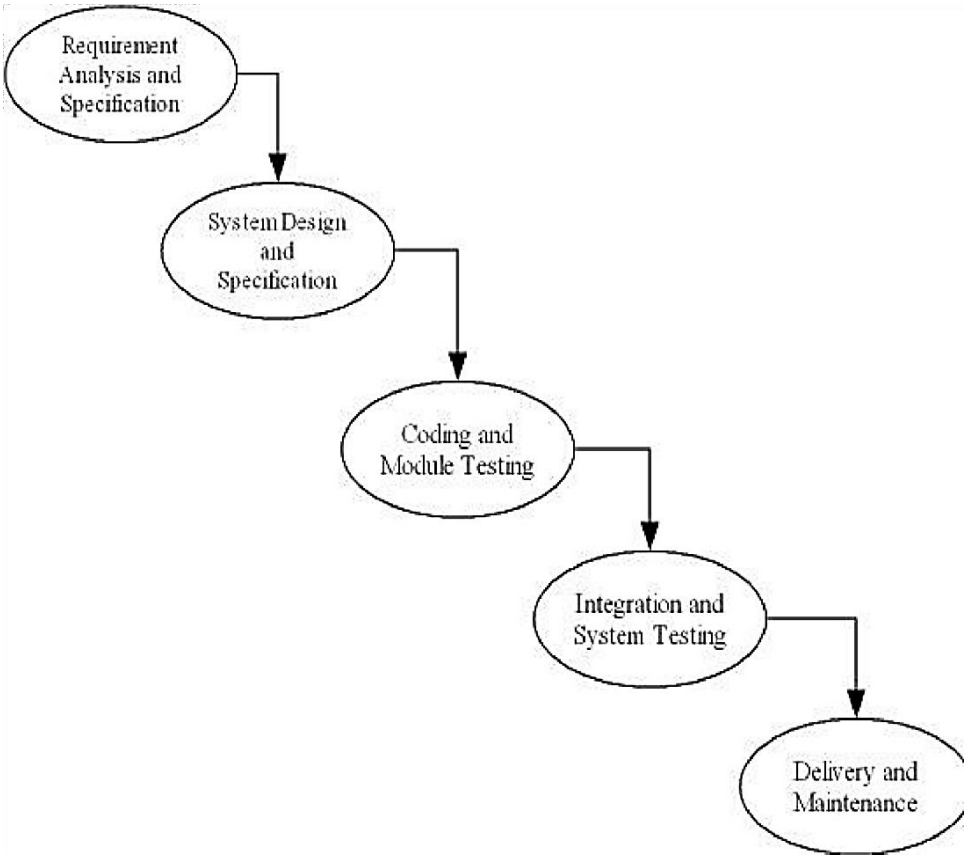


Fig 4.4: Software Lifecycle

It is It is a four-step process, which includes –

1. Feasibility Study
2. Requirement Elicitation and Analysis
3. Software Requirement Specification
4. Software Requirement Validation

4.5 FEASIBILTY STUDY

Feasibility is defined as the practical extent to which a project can be performed successfully. To evaluate feasibility, a feasibility study is performed, which determines whether the solution considered to accomplish the requirements is practical and

workable in the software. Information such as resource availability, cost estimation for software development, benefits of the software to the organization after it is developed and cost to be incurred on its maintenance are considered during the feasibility study. The objective of the feasibility study is to establish the reasons for developing the software that is acceptable to users, adaptable to change and conformable to established standards. Various other objectives of feasibility study are listed below.

To analyze whether the software will meet organizational requirements.

To determine whether the software can be implemented using the current technology and within the specified budget and schedule.

To determine whether the software can be integrated with other existing software.

4.5.1 Technical feasibility

Technical feasibility assesses the current resources (such as hardware and software) and technology, which are required to accomplish user requirements in the software within the allocated time and budget. For this, the software development team ascertains whether the current resources and technology can be upgraded or added in the software to accomplish specified user requirements. Technical feasibility also performs the following tasks.

Analyzes the technical skills and capabilities of the software development team members.

Determines whether the relevant technology is stable and established.

Ascertains that the technology chosen for software development has a large number of users so that they can be consulted when problems arise or improvements are required.

4.5.2 Operational Feasibility

Operational feasibility assesses the extent to which the required software performs a series of steps to solve business problems and user requirements. This feasibility is dependent on human resources (software development team) and involves visualizing whether the software will operate after it is developed and be operative once it is installed. Operational feasibility also performs the following tasks.

Determines whether the problems anticipated in user requirements are of high priority.

Determines whether the solution suggested by the software development team is acceptable.

Analyzes whether users will adapt to a new software.

Determines whether the organization is satisfied by the alternative solutions proposed by the software development team.

4.5.3 Economic Feasibility

Economic feasibility determines whether the required software is capable of generating financial gains for an organization. It involves the cost incurred on the software development team, estimated cost of hardware and software, cost of performing feasibility study, and so on. For this, it is essential to consider expenses made on purchases (such as hardware purchase) and activities required to carry out software development. In addition, it is necessary to consider the benefits that can be achieved by developing the software. Software is said to be economically feasible if it focuses on the issues listed below.

Cost incurred on software development to produce long-term gains for an organization.

Cost required to conduct full software investigation (such as requirements elicitation and requirements analysis).

Cost of hardware, software, development team, and training.

4.5.4 Behavioral Feasibility

Normal human psychology of human beings indicate that people are resistant to change and computers are known to facilitate change. Any project formulations should consider this factor also. Before the development of the Project titled "Delhi Metro", the need to study the feasibility of the successful execution of the project was felt and thus the following factors are considered for a Feasibility Study.

1. Need Analysis.
2. Provide the users information pertaining to the preceding requirement.

4.6 FUNCTIONAL REQUIREMENTS

The statement of services the system should provide, how the system should react to particular inputs and how the system should behave in particular situation is analysed here.

The functional requirements are:

4.6.1 Predicting android app

User can login with mail and the password

User Upload image of angiogram

User can type the test result value

Find nearest doctor for user if diseased

4.6.2 Admin and Doctor software

Admin need to login to the website by Username and Password.

The Admin can approve or reject the doctor and view the details of the doctor .

- The admin can also view user details, complaints and feedback.

The Doctor need to login to the website by email and phone number.

The doctor can view the details of the consultant.

The doctor can chat with consultant.

4.7 NON-FUNCTIONAL REQUIREMENTS

Non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. Non-functional requirements are often called qualities of a system. The Non-functional requirements includes the digital data of the organization. The memory requirement is minimum 2GB RAM and 20GB of secondary storage.

4.8 COST/BENEFIT ANALYSIS

Cost-benefit analysis compares the expected financial gain derived from a particular set of actions with the expected cost of providing each action to determine the most profitable option. The projected benefits of a plan or program are divided by its estimated total long-term cost. Additionally, subjective variables such as customer satisfaction are often quantified and factored in. The program must take into account the fact that the impacts of costs and the values of benefits might extend far into the future.

4.9 PROCESS ANALYSIS

Process analysis is a detailed study of process that looks at everything that is going on in the process. Process Analysis is the documentation and detailed understanding of how work is performed Goals of process Analysis are:

- Reduce the time, cost and effort of the process.
- Improve process efficiency.

4.9.1 Problem Definition

The purpose of the project entitled “Hybrid coherent technique of predicting cardio vascular disease using Naive Bayes and CNN with Image processing” is to predict whether a patient have heart disease or not.

4.9.2 Module Description

The main modules in the task include:

1. Admin
2. Doctor
3. User

4.9.3 Admin

Admin need to login to the website by Username and Password.

The Admin can approve or reject the doctor and view the details of the doctor .

The admin can also view user details, complaints and feedback.

4.9.4 Doctor

The Doctor need to login to the website by email and password

The doctor can view the details of the consultant.

The doctor can chat with consultant.

4.9.5 User

The user can register and login to app .

The user can upload image and see prediction. Also can enter report values to see prediction

The user can book doctor .

4.10 SYSTEM DESIGN

Once the requirements document for the software to be developed is available, the software design phase begins. While the requirement specification activity deals entirely with the problem domain, design is the first phase of transforming the problem into a solution. In the design phase, the customer and business requirements and technical considerations all come together to formulate a product or a system. The design process comprises a set of principles, concepts and practices, which allow a software engineer to model the system or product that is to be built. This model, known as design model, is assessed for quality and reviewed before a code is generated and tests are conducted. The design model provides details about software data structures, architecture, interfaces and components which are required to implement the system. This chapter discusses the design elements that are required to develop a software design model. It also discusses the design patterns and various software design notations used to represent a software design.

4.10.1 Data Flow Diagram

Data flow diagrams (DFDs) reveal relationships among and between the various components in a program or system. DFDs are an important technique for modeling a system's high-level detail by showing how input data is transformed to output results through a sequence of functional transformations. DFDs consist of four major components: entities, processes, data stores, and data flows. The symbols used to depict how these components interact in a system are simple and easy to understand; however, there are several DFD models to work from, each

having its own symbology. DFD syntax does remain constant by using simple verb and noun

When it comes to conveying how information data flows through systems (and how that data is transformed in the process), data flow diagrams (DFDs) are the method of choice over technical descriptions for three principal reasons.

1. DFDs are easier to understand by technical and nontechnical audiences.
2. DFDs can provide a high level system overview, complete with boundaries and connections to other systems.
3. DFDs can provide a detailed representation of system components.

DFDs help system designers and others during initial analysis stages visualize a current system or one that may be necessary to meet new requirements. Systems analysts prefer working with DFDs, particularly when they require a clear understanding of the boundary between existing systems and postulated systems. DFDs represent the following:

1. External devices sending and receiving data.
2. Processes that change that data.
3. Data flows themselves.
4. Data storage locations.

The hierarchical DFD typically consists of a top-level diagram (Level 0) underlain by cascading lower level diagrams (Level 1, Level 2…) that represent different parts of the system.

External Entity

An external entity, which are also known as terminators, sources, sinks, or actors, are an outside system or process that sends or receives data to and from the diagrammed system. They're either the sources or destinations of information, so they're usually placed on the diagram's edges.

Process

Process is a procedure that manipulates the data and its flow by taking incoming data, changing it, and producing an output with it. A process can do this by performing computations and using logic to sort the data or change its flow of direction. Processes usually start from the top left of the DFD and finish on the bottom right of the diagram.

Data store

Data stores hold information for later use, like a file of documents that's waiting to be processed. Data inputs flow through a process and then through a data store while data outputs flow out of a data store and then through a process.

Data flow

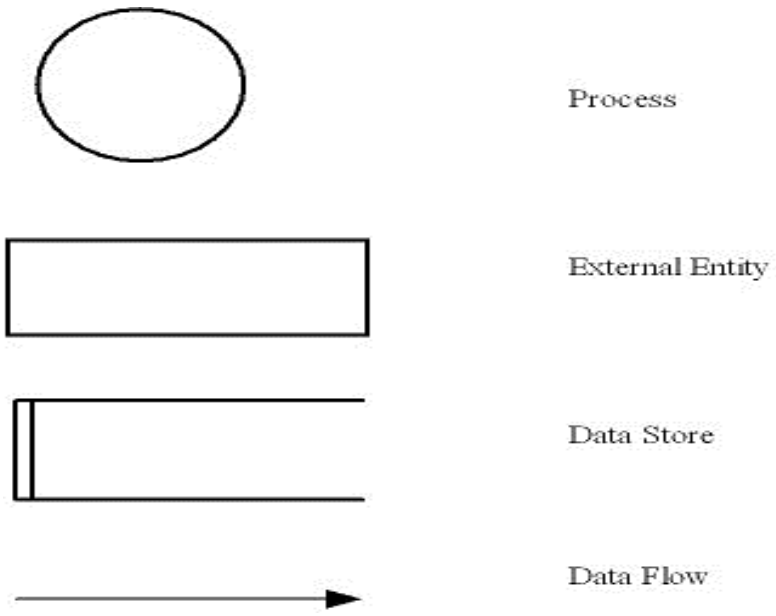


Fig 4.5: DFD symbols

Data flow is the path the system's information taken from external entities through processes and data stores. With arrows and succinct labels, the DFD can show you the direction of the data flow.

Level 0 DFDs, also known as context diagrams, are the most basic data flow diagrams. They provide a broad view that is easily digestible but offers little detail. Level 0 data flow diagrams show a single process node and its connections to external entities.

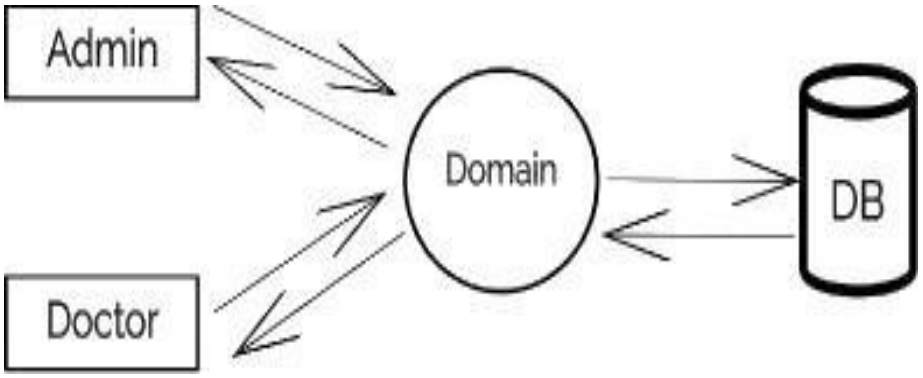


Fig 4.6: LEVEL 0 DFD

Level 1 DFDs are still a general overview, but they go into more detail than a context diagram. In a level 1 data flow diagram, the single process node from the context diagram is broken down into subprocesses. As these processes are added, the diagram will need additional data flows and data stores to link them together.

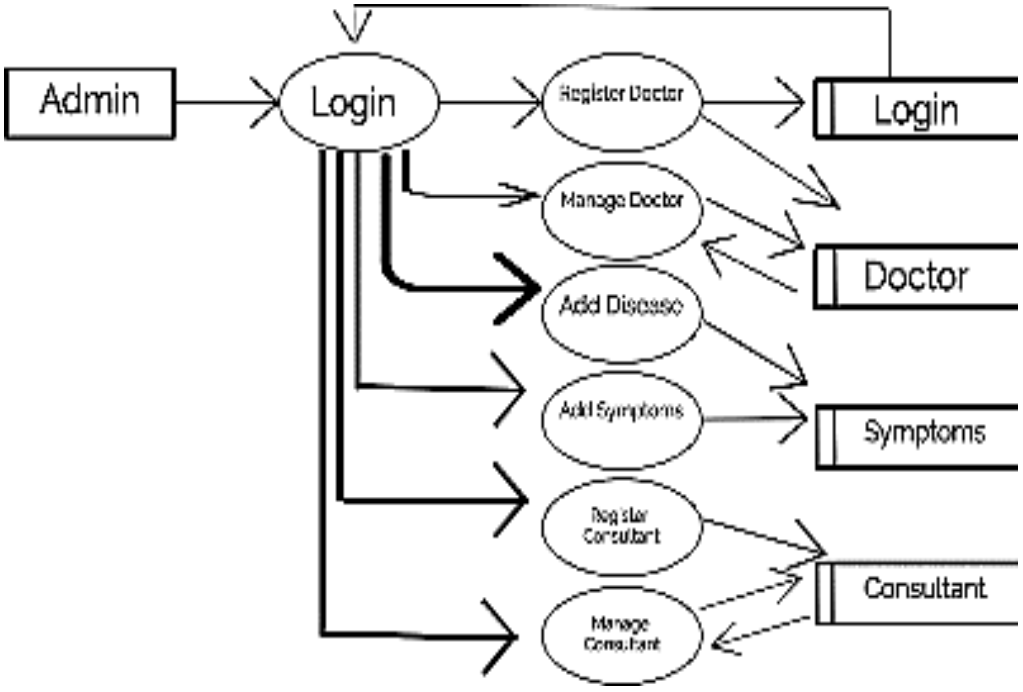


Fig 4.7: LEVEL 1.0 DFD

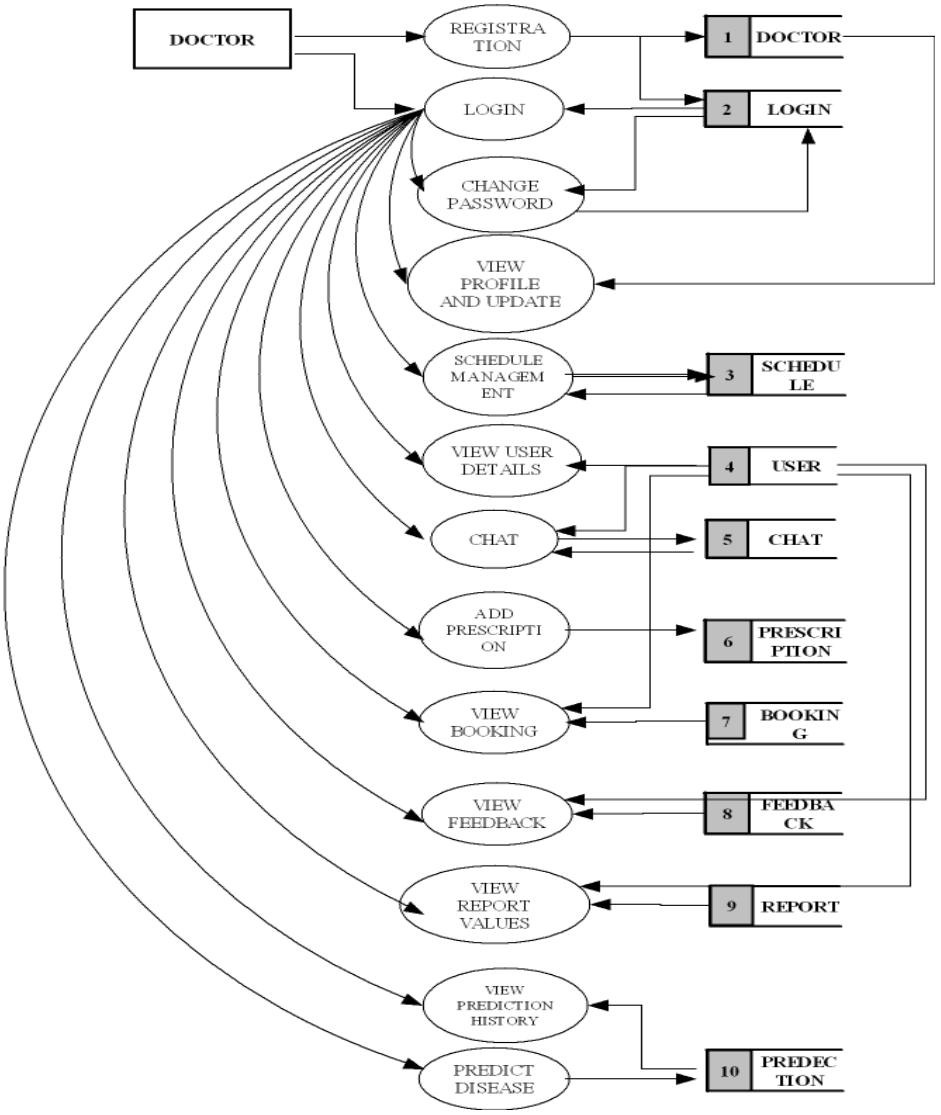
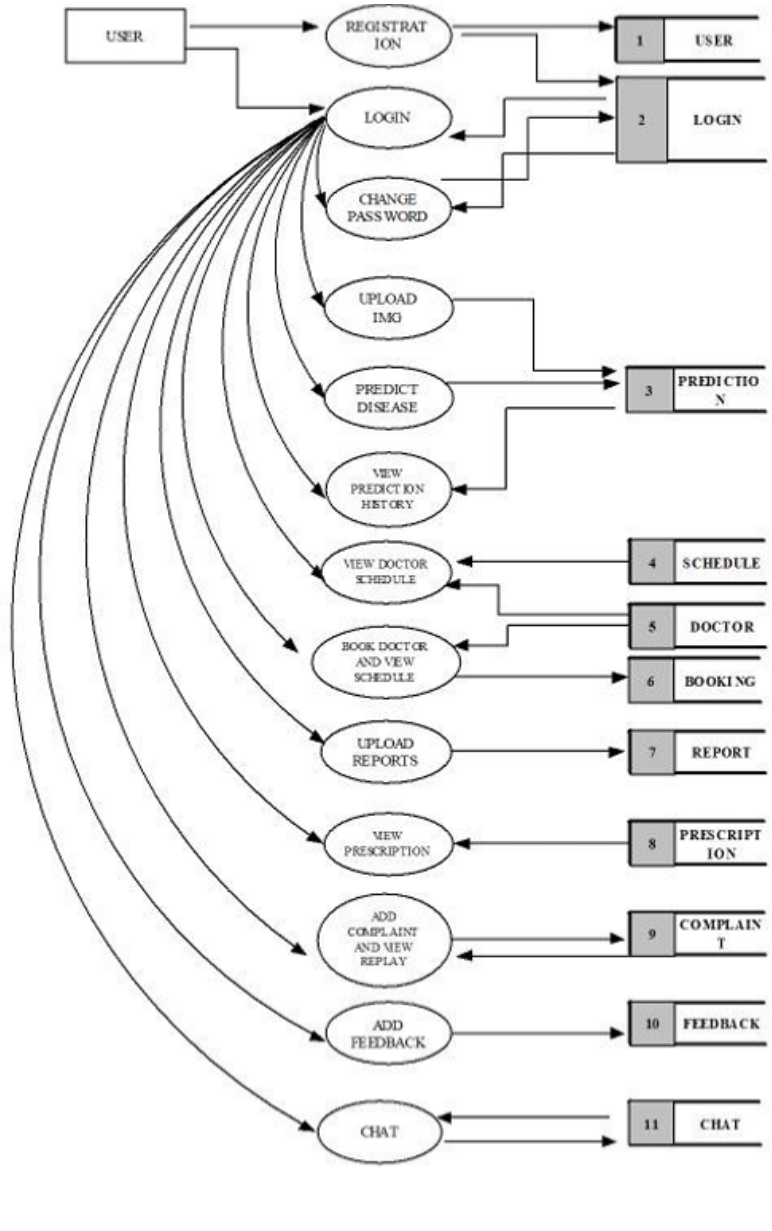


Fig 4.8: LEVEL 1.1 DFD



4.10.2Database Design

The database design is a collection of stored data organized in such a way that the data requirements are satisfied by the database, The general objective is to make information access easy, quick inexpensive and flexible for the user. There are also some specific objectives like controlled redundancy from failure, privacy, security, and performance. A collection of relative records make up a table. To design and store data to the needed forms database tables are prepared. Two essential settings for a database are:

Primary key: The field that is unique for all the record occurrences.

Foreign key: The field used to set relation between tables.

Normalization is a technique to avoid redundancy in the tables.

Database design requires managing the large bodies of information. The management of data involves both the definition of structure of the storage of information and provision of mechanism for the manipulation of information. In addition the DB system must provide for the safety of information handled, despitethe system crashes due to attempts art unauthorized access. For developing an efficient db, we will have to fulfil certain condition such as:

Control redundancy: Redundancy is the duplication of critical components or functions of a system with the intention of a increasing reliability of the system. There are two types of redundancy; known as passive redundancy and active redundancy. Both functions prevent performance decline from exceeding specification limits without human intervention using extra capacity.

Data independence: Data independency is the type of data transparency that matters for a centralized DBMS. It refers to the immunity of user applications to changes made in the definition and organization of data. The physical data independence deals with hiding the details of the storage structure from user applications. The data independence and operation independence together gives the feature of data abstraction.

Accuracy and integrity: Data integrity refers to maintaining and assuring the accuracy and consistency of the data over its entire life cycle, and is a critical aspect to the design, implementation, and usage of any system which stores, processes, or retrieves the data. Accuracy refers to getting the extracted data.

Avoiding inordinate delays: Difficulties in managing DBMS delays managing performance are more challenging. DBMS stores information on hard disks. This has different types of delays.

Ease of use: It is also referred to as usability. It is the ease of use and learn ability of a human made object. It also includes methods of measuring usability, such as needs analysis and the study of the principles behind an object's perceived efficiency

Recovery from failure: In DBMS failures are generalized into various categories. They are transaction failures, system crash, disk failures etc. When a DBMS recovers from a crash, it should maintain some characteristics. It should check the states of all the transactions, which were being executed, it should check whether the transaction can be completed now or it needs to be rolled back etc.

Private and security: Database security concerns the use of a broad range of information security controls to protect databases against compromises of their confidentiality, integrity and availability. Security risks to database system include unauthorized or unintended activity or misuse by authorized database users, database administrators or network/systems managers or by unauthorized users or hackers etc.

List of tables used in this project:

Login, doctor, consultant, disease, patient, symptoms

Login table

id	UserName	Password	type

Doctor table

lid	Name	Gender	Address	Longiude	Latitude	Experience	Area of Specification

pincode	Email	Qualification

Consultant table

id	Name	Gender	Age	Address	Longiude	Latitude	Email

Disease table

id	Disease name	Description

Patient table

id	Pid	Disease

Prediction table

id	did	Prediction

Fig 4.10: Table schema

CHAPTER V

SYSTEM IMPLEMENTATION AND TESTING

5.1 SOFTWARE REQUIREMENT

Language: Python

Operating system: Windows 10 (64 bit)

IDE: Android Studio 4.0, Pycharm

Database: MySQL Server

5.1.1 MySQL

MySQL is an Oracle-backed open source relational database management system (RDBMS) based on Structured Query Language (SQL). MySQL runs on virtually all platforms, including Linux, UNIX and Windows. Although it can be used in a wide range of applications, MySQL is most often associated with web applications and online publishing. MySQL is an important component of an open source enterprise stack called LAMP. LAMP is a web development platform that uses Linux as the operating system, Apache as the web server, and MySQL as the relational database management system and PHP as the object-oriented scripting language. (Sometimes Perl or Python is used instead of PHP).

Features:

- Relational Database Management System (RDBMS): MySQL is a relational database management system.
- Easy to use: MySQL is easy to use. You have to get only the basic knowledge of SQL. You can build and interact with MySQL with only a few simple SQL statements.

It is secure: MySQL consist of a solid data security layer that protects sensitive data from intruders. Passwords are encrypted in MySQL.

Client/ Server Architecture: MySQL follows a client /server architecture. There is a database server (MySQL) and arbitrarily many clients (application programs), which communicate with the server; that is, they query data, save changes, etc.

Free to download: MySQL is free to use and you can download it from MySQL official website.

It is scalable: MySQL can handle almost any amount of data, up to as much as 50 million rows or more. The default file size limit is about 4 GB. However, you can increase this number to a theoretical limit of 8 TB of data.

Compatible on many operating systems: MySQL is compatible to run on many operating systems, like Novell NetWare, Windows* Linux*, many varieties of UNIX* (such as Sun* Solaris*, AIX, and DEC* UNIX), OS/2, FreeBSD*, and others. MySQL also provides a facility that the clients can run on the same computer as the server or on another computer (communication via a local network or the Internet).

Allows roll-back: MySQL allows transactions to be rolled back, commit and crash recovery.

High Performance: MySQL is faster, more reliable and cheaper because of its unique storage engine architecture.

High Flexibility: MySQL supports a large number of embedded applications which makes MySQL very flexible.

High Productivity: MySQL uses Triggers, Stored procedures and views which allows the developer to give a higher productivity.

5.1.2 PyCharm IDE

PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as Data Science with Anaconda.

FEATURES:

Coding assistance and analysis, with code completion, syntax and error highlighting, linter integration, and quick fixes.

Project and code navigation: specialized project views, file structure views and quick jumping between files, classes, methods and usages.

Python refactoring: includes rename, extract method, introduce variable, introduce constant, pull up, push down and others.

Support for web frameworks: Django, web2py and Flask [professional edition only.

Integrated Python debugger

Integrated unit testing, with line-by-line code coverage.

Google App Engine Python development [professional edition only. Support for scientific tools like matplotlib, numpy and scipy professional edition only.For developing the Android application “Android studio” and for working with database MySql “SQLyog” tool is used. Due to its simple UI, It is very easy to built and manage database using queries.

5.1.3 SQLyog

MySQL manager and admin tool:

SQLyog is the most powerful manager, admin and GUI tool for MySQL, combining the features of MySQL Query Browser, Administrator, phpMyAdmin and other MySQL Front Ends and MySQL GUI tools in a single intuitive interface. SQLyog is a fast, easy to use and compact graphical tool for managing your MySQL databases. SQLyog was developed for all who use MySQL as their preferred RDBMS. Whether you enjoy the control of handwritten SQL or prefer to work in a visual environment, SQLyog makes it easy for you to get started and provides you with tools to enhance your MySQL experience.

Session Restore: SQLyog is a powerful MySQL manager that restores your previous session the way you left it. Be it a system crash or accidentally closing your MySQL client.

Schema and Data Sync: Find and fix schema mismatching while syncing data from one database to another. Replicate data periodically to a different server at scheduled intervals.

Compressed backups with scheduling: Find the plethora of options in mysqldump daunting? Just back up with a single click interface in SQLyog MySQL GUI tool.
Auto complete and SQL formatting: Save hours of typing. Write queries 10x faster.

Inherited spaghetti SQL? Make it neat with a click. Improve readability and maintenance of SQL.

Query Builder: Build complex queries using drag-n-drop interface. Visually create SQL statements without the need to remember column names.

Schema Optimizer and Index Analyzer: Determine the optimal column types. Fit more data into memory thereby improving response times. Eliminate Redundant Indexes by improving INSERTs and UPDATEs.

Object Browser: The Object browser window gives you complete details of the server you are connected to in a tree format. It allows you to perform different operations on selected objects such as Open Table, Create Table, Alter Table, Manage Indexes, Relationships/Foreign Keys, Backup/Export, Import and more.

Favorites : SQLyog allows easy access to frequently used SQL scripts. The script files can be stored as files from the SQLyog interface or link to an existing file anywhere where windows can access it - on a local drive or a shared network drive. The SQL scripts and the file links can be organized in folder and subfolders. SQLyog gives you a choice between four different connectivity options. No matter whatever your firewall/proxy/ISP settings are, you should be able to smartly manage your MySQL Server(s).

The various connectivity options (in the order of responsiveness and network efficiency) are:

• Direct connection using MySQL C API: This is by far the fastest way to connect and manage your MySQL server. This is the preferred method if you are not concerned about your MySQL traffic being spoofed. However, this requires direct access to your MySQL server. Most ISPs block the MySQL port or disallow remote access.

• SSH Tunneling: Use this option if you need to encrypt the contents of your MySQL traffic, but still want the efficiency of the MySQL C API. If your host supports SSH tunneling, SQLyog can use the tunnel to communicate with MySQL.

• HTTP Tunneling: This option saves your day if your ISP/MySQL provider disallows remote access, either by blocking the MySQL port or by giving access to "local host" only. The response is slow compared to the previous options since the data is XML encoded and HTTP is stateless by nature. However, you can use all the cool features of SQLyog.

• HTTPS Tunneling: This option is similar to HTTP Tunneling but adds another layer of security by establishing a secure channel.

• SSL Encryption: With this option all communication between SQLyog and the MySQL server is encrypted. The MySQL server must support this option. Not all servers do.

5.1.4 Android Studio IDE

Android Studio is the official Integrated Development Environment (IDE) for Android app development, based on IntelliJ IDEA. On top of IntelliJ's powerful code editor and developer tools, Android Studio offers even more features that enhance your productivity when building Android apps, such as:

- A flexible Gradle-based build system.
- A fast and feature-rich emulator.
- A unified environment where you can develop for all Android devices.
- Instant Run to push changes to your running app without building a new APK.
- Code templates and GitHub integration to help you build common app features and import sample code.
- Lint tools to catch performance, usability, version compatibility, and other problems.
- C++ and NDK support.
- Built-in support for Google Cloud Platform, making it easy to integrate
- Google Cloud Messaging and App Engine.

Android features:

1. Instant App Run

It is an advanced technology in which it cleverly understands the transmutations done in the applications and delivers it instantly without taking time to rebuild the apk and installations make. So, quick that you can see the changes in app immediately. This is done by launching the URL on Android application installing the native libraries with Android Instant apps. Instant App sometimes is known as "New Module Wizard".

2. Visual Layout Editor

Layout editor helps to build the layout quickly by adding different attributes either by hard-code or drag and drop. The preview of the codes can be seen easily on the visual editor screen and changes can be made accordingly by resizing it dynamically. This will make testing the application process more facile and more exhaustive.

3. Fast Emulator

Android has a great feature of Emulator which is exactly like the android phones to test how the application looks like in physical devices. It gives real-time experience to the Android applications. It allows you to test your applications faster and on different-different configuration devices like tablet, android phone etc. It helps you to make your application development life cycle shorter and more efficient.

4. Intelligence Code Editor

Android Studio provides you with the intelligent and quick code editor. This will help you and guide you with the accurate code. It helps you to complete code in advance and analyses your code in advance before building. Android studio has the special feature of the development of code by the drop-down list with suggesting the code you can integrate.

5. Addition of New Activity as a Code Template

Yes, Android also has the feature of templates built-in. If you know to build that accordingly that makes your task easier. it has both pros and cons, you don't find every template in Android Studio. It's an additional feature which helps the developer to build an application efficiently and effectively which provide effective solutions.

6. Help to Build Up App for All Devices

Android studio builds applications for every screen size, for wear and gear devices etc. It also can stimulate the various type of features which a hardware has like GPS location tracker, multi-touch.

7. Help to Connect with Firebase

Android Studio helps to give real-time experience with IOT based project development with dynamic upgrades in the application. Firebase connectivity help to create direct updates and provide databases connectivity. To build high-quality applications we must use Firebase connectivity it helps to build the scalable infrastructure for building the application. You can create chat applications by using firebase connectivity it helps you to do happy chat experience.

8. Support KOTLIN

Kotlin the official language can be considered for Android. It is a language without having any new restrictions and has various advantages in it. The great feature of Kotlin is it run without any disturbances in older android versions as well that means no issues of specific android versions. It runs fast and equivalent to Java. Java developers can easily hands-on Kotlin with no-issues because it based on automated Java only.

9. Color Previews

Android studio helps to see the code XML part in a preview to know that how perfectly we are designing the application according to the need before launching the application. It provides powerful functionality and enhanced features of drag and drops or resizes the application. It contains drag and drop features but not support for every function, that's why be careful while doing that.

10. Maven Repository

In Android Studio, Maven integration of its repository can be done, within SDK manager support libraries of IDE is used. It's a kind of a repository which is a directory in which various jar files like project jars, Plug-in are stored.

5.2 HARDWARE REQUIREMENTS

Processor: Intel Core i3

Processor Speed: 2.4GHZ

RAM: 4GB

5.3 INSTALLATION REQUIREMENTS

Smart Phone with Android Operating System

Android version: 4.4 and above (i.e., kitkat and above).

5.3.1 Android Operating System

Android is an open source and Linux based operating system for mobile devices such as Smartphone and tablet computers. Android was developed by the open handset Alliance, led by Google, and other companies. Google wanted Android to be open and free; hence most of the Android code was released under the open source Apache license, which means that anyone who wants to use Android can do so by downloading the full Android source code. Vendors (i.e. hardware manufactures) can add their own proprietary extensions to Android and customize Android to differentiate their product from others. The main advantage of adopting Android is that it offers a unified approach to application development. Developer need only develop for Android, and their application should be able to run numerous different devices as long as these devices are powered using android. Application is the most important part of success chain in the world of Android.

Features of Android

Android is open source and freely available to manufacturers for customization therefore there are no fixed hardware or software configurations. However Android itself support the following features;

- Storage – Uses MySQL for data storage.
- Connectivity - Support GSM/EDGE, IDEN, CDMA, EV-DO, UMTS, Bluetooth(includes A2DP and AVRCP), Wi-Fi, LTE and WiMAX.
- Messaging - Supports both SMS and MMS.
- Web browser - Based on open source web kit together with chrome's Vs. JavaScript engine.
- Media Support- Includes support for the following media: H.263, H.264 (in 3GP 3GP or MP4 container), MPEG-4 SP, AMR-WB(in 3GP container), AAC, HEAAC(in MP4 or 3GP container), MP3, MIDI, WAV, JPEG, PNG, GIF and BMP.
- Hardware Support - Accelerometer sensor, Camera, Digital Compass, Proximity sensor and GPS.
- Multi-touch - Supports multi-touch screens
- Multi-tasking - Support multi-tasking applications.
- Flash Support - Android 2.3 supports flash 10.1.
- Tethering - supports sharing of internet connection as a wired/wireless Hotspot

The main Android devices in the market include Smartphone, Tablet, E-reader devices, net books, MP4 players, internet TVs and so on.

Android Versions	*Release Date*	*Code Name*
1.1	9th February 2009	
1.5	30th April 2009	Cupcake
1.6	15th September 2009	Donut
2.0/2.1	26th October 2009	Éclair
2.2	20th May 2010	Froyo
2.3	6th December 2010	Gingerbread
3.0/3.1/3.2	22nd February 2011	Honeycomb
4.0/4.0.4	19th October 2011	Ice Cream Sandwich
4.1-4.3.1	9th July 2014	Jelly Bean
4.4-4.44,4.4W-4.4W.2	31st October 2013	Kitkat
5.0	12th November 2014	Lollipop
M(Developer Preview 2)		

Table 5.1: Android Versions

5.4 SYSTEM IMPLEMENTATION

It is the stage of implementation, which ensures that system works accurately and effectively before the live operation commences. It is confirmation that all are correct and opportunity to show the users that the system must be tested with test data show that the system will operate successfully and produce expected result under expected conditions. Before implementation, the proposed system must be tested with raw data to ensure that the modules of the system work correctly and satisfactorily. The system must be tested with valid data to achieve its objective. The purpose of system testing is to identify and correct errors in the candidate system. As important as this phase is, it is one that is frequently compromised. Typically, the project the schedule or the user is eager to go directly to conversion. Actually, testing is done to achieve the system goal. Testing is vital to the parts of

the system are correct, the goal will be successfully achieved. Inadequate testing or non-testing leads to errors that may not appear until months later. This creates two problems:

The time lag between the cause and appearance of the problem.

The effect of system errors on files records within the system. A small system errors can conceivably exploded into much larger problem. Effectively early

in the process translates directly into long term cost saving from a reduced numbers of errors.

In our project we properly installed our system for our client in the following way:

Hardware configuration installed properly.

Front-end and back-end software installed properly.

Additional softwares are installed properly.

Path settings are corrected

Run the software

The system worked efficiently and properly after proper implementation.

5.5 SYSTEM TESTING

The scope of system testing is not only limited to the design of the system but also to the behavior and believed expectations of the business. In accordance with the software test cycle, system testing is performed before acceptance testing and after integration testing. Independent users or testers are given the tasks to perform in the system testing phase. Importance aspects of system testing:

In accordance to the software development lifecycle, system testing is considered as the first level of testing where the entire system is checked or tested.

Proper evaluation of the system meeting the functional requirements is done in system testing.

Validation, verification and testing of business requirements and application architecture is done during the system testing phase.

System testing provides users with an effective environment which more or less resembles the live or production environment. As such any testing done provides more reliable and efficient results.

5.5.1 Testing Techniques

Testing is an activity to verify that a correct system is being built and is performed with the intent of finding fault in the system. However it is not restricted to being performed after the development phase is complete. But this is to carry out in parallel with all stages of system development, starting with requirement specification. Testing results, once gathered and evaluated, provide a qualitative indication of software quality and reliability and serve as a basics for design modification if required. A project is said to be incomplete without proper testing. There are two types of testing techniques:

White Box Testing: White Box Testing is also known as open, transparent or glass box testing. In white box testing, the tester has prior knowledge of the code and accordingly prepares the test case. The tester has the knowledge of the internals of a system and knows how the system is implemented. The tester uses this knowledge to develop test cases that will examine the control flow, information flow, data flow, exception and error handling as well as coding practices of the system.

Working of White Box Testing

1. The first step for the tester is to understand the source code.
2. White Box testing then involves testing of internal functions of the application, so knowledge of source code is crucial.
3. The tester should be aware of the secure coding practices as security is the most important factor in testing.
4. Tester can then write code for testing the application or can prepare certain test cases with suitable inputs.

White Box Testing Techniques

Code Coverage Analysis: It eliminates gaps in test case suite by identifying the program which cannot be examined by test cases. In addition, you can

create test cases for untested part of the program which improves the quality of the software.

Statement Coverage: This technique checks every statement of the code at least once during the test cycle.

Branch Coverage: This technique tests every possible path in the code like If-else loops and other conditional loops of the software.

Advantages

It optimizes the code as it tests every statement of the code.

Automated testing is supported.

Tests and test scripts can be reused.

Testing is supported at early development stages.

Black Box Testing

Testing is broadly based on software requirements and specifications. Black Box Testing is a technique in which tester is unaware about the internal structure or code of the software. The focus is on inputs and outputs ignoring the internal knowledge of the code. Using black box testing, one can test operating systems like Windows, websites like Google and even our own customized applications, as the core knowledge about these operating systems are not required.

Working of Black Box Testing

The first step is to thoroughly examine the requirements and specifications of the system.

The tester explores the system's UI and functionality to understand how the processes on the system are expected to work.

On later stage, the tester checks efficiency of the software by determining expected outputs with their corresponding inputs.

Finally, the developer fixes the bug detected and the output undergoes retesting.

Black Box Testing Techniques

Equivalent Class Testing: It is used to reduce the number of possible test cases to an ideal level to maintain a reasonable test coverage.

Boundary Value Testing: It determines whether certain range of values are accepted by the software or not. This helps in reducing number of test cases.

Decision Table Testing: A decision table puts conditions and their outcomes in a matrix. There is a unique combination in every segment.

Advantages

Suitable for large code segments

Increased Efficiency

Prior knowledge of code is not required

Black box testing is all about enhancing the user experience even if they are from a non-technical background. On the other hand, for technical support and precise coding, White box testing is an excellent approach for organizations to employ.

5.5.2 Types of Testing

Unit testing: Unit testing includes testing of smallest piece of software to verify its behavior. It ensures that the code should satisfy the requirements. It is done by developers. Unit testing is done manually or by using tools also like JUnit, NUnit. It is done to reduce future cost due to early detection of errors. It tests smaller components so that it is easy to find out errors. Enhancement can be performed. **Integration testing**: Integration testing is done to ensure that individually tested components can work together to perform the intended task. Integration testing is important because modules work individually but they may not work together

when they are integrated It is used to uncover the problems which are occur in interfaces between different modules.

Following approaches are used in Integration testing :

Bottom up approach: In bottom up approach the modules are combined and tested which belong to the bottom of the application hierarchy starting with modules at the top module. When top module is not ready but bottom modules are developed at that time this approach is used. Driver is used in bottom up approach.

Top down approach: When all the bottom modules are not ready at that time top down approach is used. Stubs are used in top down approach. Stub is a special purpose program which is written to test the integration between modules.

Critical Part First approach: In this approach critical parts of the system are implemented and tested first. When focusing on entire application is not possible due to time constraint then Functionality which is important is concentrated first.

Big bang approach: Big bang approach is a common approach in which all modules are integrated at once. No stub and drivers are required in this approach. But it is difficult to debug the code and find out the location of the defect.

User Acceptance testing: User Acceptance testing is a final testing before the system is accepted for its intended use. It is based on requirement document. UA testing validates both functional as well as nonfunctional specifications. UA testing is done according to user requirement to confirm that application is behaving as expected by the customer. It is usually done by customer or end user.

There are two types of UA testing;

Alpha testing: It is done after system testing. It is done at the developer site by a customer. It is performed in a controlled environment.

Beta testing: It is done after alpha testing In beta testing developer is not present at the location. It is conducted at customer site by an end user.

5.6 SYSTEM MAINTENANCE

Software maintenance is widely accepted part of SDLC now a days. It stands for all the modifications and updations done after the delivery of software product. There are number of reasons, why modifications are required, some of them are briefly mentioned below:

Market Conditions - Policies, which changes over the time, such as taxation and newly introduced constraints like, how to maintain bookkeeping, may trigger need for modification.

Client Requirements - Over the time, customer may ask for new features or functions in the software.

Host Modifications - If any of the hardware and/or platform (such as operating system) of the target host changes, software changes are needed to keep adaptability.

Organization Changes - If there is any business level change at client end, such as reduction of organization strength, acquiring another company, organization venturing into new business, need to modify in the original software may arise.

Reports suggest that the cost of maintenance is high. A study on estimating software maintenance found that the cost of maintenance is as high as 67% of the cost of entire software process cycle.

CHAPTER VI

RESULT AND DISCUSSION

Hybrid coherent technique of predicting cardio vascular disease using naive bayes and CNN with image processing will predict if a person is infected with communicable disease or not. The proposed system consists of mainly – Admin and Doctor website and Prediction android app. The admin website is used by the administrator to imply various operations such as manage doctor, and manage complaints. The prediction android app is used by the user to upload image of ECG report through the app and can easily predict the disease. In the proposed system admin will register all the consultants, but there is a chance that all the consultants may have disease. So the doctor will confirm that whether the visited consultant have heart disease. This app can also predict heart disease by user entering thier report values. The user can also consult near by doctors futher treatment through this app.

The system also performs the task with a good performance, an addition it's easily used requiring uncomplicated configurations to be used, so it's a user-friendly application. The developed system performs the required work with accuracy 94% within the dataset and 85% with the external data, and the integrated system is working properly. .The color feature and texture feature can be used to make up for the weakness of the unidirectional recognition so that the recognition ion rate can reach to 90% and more, which greatly improves the accuracy, therefore improving the overall accuracy of Heart disease identification.

CHAPTER VII

CONCLUSION AND FUTURE SCOPE

Difficulties in the detection of heart disease because of the increasing number of cases all over the world, which make it a challenge to the cardiologist to recognize the different heart diseases easily, a computer aided system is proposed to resolve these difficulties, so a machine learning model with CNN and Naives Bayes as a classifier is used to develop mobile applications for android devices. CNN and Naives Bayes is a best approach for skin disease detection. The method used for extracting useful information from the image by using gray scale image conversion method is powerful tool for extracting relevant information from the user uploaded image. CNN provide an accuracy of 92% of accuracy from the prediction from the image and Naives Bayes provide an 95% of accuracy from the input records of the patients.

Mobile platform is leading technologies in modern day. So we are targeting release Android mobile platform and IOS compatibility in the near future. Also with the usage and the demand of the system we will expand the number of diseases which are to be recognized by the system in to a considerable amount. Most people have very busy schedule so this kind of system very useful for the future. And also to attract and help local community more we are planning to enhance the local language support and also to identify animal skin disease model. That kind of improvement increase user friendliness for the system. Those kind improvements want to be done in future to our system.

CHAPTER VII

CONCLUSION AND FUTURE SCOPE

HYBRID COHERENT TECHNIQUE OF PREDICTING CARDIO VASCULAR DISEASE USING NAIVE BAYES AND CNN WITH IMAGE PROCESSING is an excellent solution for the conventional heart disease prediction methods. The proposed method will predict if a person has heart disease or not. This application will reduce the time and effort of the medical team since it seeks to remedy the tiresome manual prediction of the disease. The user can just upload image of ECG report through the app and can easily predict the disease. This app can also predict heart disease by user entering thier report values. The user can also consult near by doctors futher treatment through this app.

Future scope is that more types of heart diseases can be included in this.