IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⪪ Impact Factor 7.39 ⪪ Vol. 11, Issue 9, September 2022

DOI: 10.17148/IJARCCE.2022.11915

Prediction of Cardiac Disease Using Machine Learning

Dr. Chethan Chandra S Basavaraddi [1], Dr. Vasanth G[2], Sapna S Basavaraddi[3], Nandini K R[4], Pallavi T[5], Spandana D S[6], Spandana T R[7]

Research Scholar-Department of Computer Science and Engineering, Research Centre - Government Engineering College Krishnarajapete, Associate Professor, Dept. of CSE, KIT-Tiptur, Visvesvaraya Technological University, Belagavi-590018[1]

Professor and Head, Computer Science and Engineering, Government Engineering College, Ramanagara-562159 Visvesvaraya Technological University, Belagavi-590018[2]

Research Scholar-Department of Computer Science and Engineering, Research Centre - Sri Siddhartha Institute of Technology, Tumkur Assistant Professor, Dept. of CSE, Kalpataru Science College-Tiptur[3]

Project Associates-Department of Computer Science and Engineering, Kalpataru Institute of Technology, Tiptur-572201, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India[4,5,6,7]

Abstract: Most countries face high and increasing rates of heart disease or cardiovascular disease. Even though, modern medicine is generating huge amount of data every day, little has been done to use this available data to solve the challenges that face a successful interpretation of echocardiography examination results. To design a predictive model for heart disease detection using data mining techniques from Transthoracic Echocardiography Report dataset that is capable of enhancingthe reliability of heart disease diagnosis using echocardiography. Knowledge Discovery in Database (KDD) methodology consisting of nineiterative and interactive steps was adopted to extract significant patterns from a dataset containing 7,339 echocardiography examination reports of patients. The data used for this study was collected by Hospital. The findings of this study revealed all the models built from J48 Decision Treeclassifier, Naïve Bayes classifier and Neural Network have high classification accuracy andare generally comparable in predicting heart disease cases. However, comparison that is based on True Positive Rate suggests that the J48 model performs slightly better in predictingheart disease with classification accuracy of 95.56%. This study showed that data mining techniques can be used efficiently to model and predict heart disease cases. The outcome of this study can be used as an assistanttool by cardiologists to help them to make more consistent diagnosis of heart disease.

Keywords: KDD, Data Mining, Decision Tree, Neural Network, Bayesian classifier, Heart Disease.

1. INTRODUCTION

Heart disease or cardiovascular disease is the class of diseases that involve the heart or blood vessels (arteries and veins). Today most countries face high and increasing rates of heart disease and it has become a leading cause of debilitation and death worldwide in men and women over age sixty-five and today in many countries heart disease is viewed as a "second epidemic," replacing infectious diseases as the leading cause of death (Gale Nutrition Encyclopedia, 2011). Traditionally, heart disease was thought to be the problem of developed countries, but now it is becoming a headache for developing countries too and it is especially devastating for the developing countries since they do not have adequate health care. Asnoted by Office on Women's Health (2006) heart disease was considered to be a man's problem, but now it is recognized as number one killer of women, just as it is of men.

Making a diagnosis of heart disease includes taking a complete medical evaluation and history and physical examination and early diagnosis of heart disease can help reduce the rate of mortality (Thaksin University, 2006). One of the best ways to diagnose a heart disease is by using echocardiography. Echocardiography, or echo, is a painless test thatuses sound waves to create pictures of the heart. The test gives information about the size and shape of the heart and how well the heart chambers and valves are working. Echo also can be done to detect heart problems in infants and children.

DOI: 10.17148/IJARCCE.2022.11915

The test also can identify areas of heart muscle that aren't contracting normally due to poor blood flow or injury from a previous heart attack. In addition, a type of echo called Doppler ultrasound shows how well blood flows through the chambers and valves of the heart (Joel and Robert, 1976).

The analysis of Echo data by experts is time consuming and this is in concomitant with the shortage of experts possessing knowledge on the analysis of Echo data. Thus, methods to automate the interpretation of Echo recordings by minimizing human efforts are important for diagnosis of heart disease in patients. In order to solve this and many other problems in the health sector related to disease diagnosis, one has to come up with a way to extract hidden information from enormous datasets that are collected in the past. Data mining can be a solution by generating rules from those enormous datasets which can be used in echo readings.

Data mining can be a useful tool in the health sector and healthcare. Organizations that perform data mining are better positioned to meet their long-term needs, Benko and Wilson (2003) argue that data can be a great asset to healthcare organizations, but they have to be first transformed into information.

Predicting the outcome of a disease is one of the most interesting and challenging tasks in which to develop data mining applications. In recent years new research avenues such as knowledge discovery in databases (KDD), which includes data mining techniques, has become a popular research tool for medical researchers who seek to identify and exploit patterns and relationships among large number of variables, and be able to predict the outcome of a disease using the historical cases stored within datasets.

Kangwanariyakul et al., (2010), Patil and Kumaraswamy, (2009) have tried to apply data mining techniques in the diagnosis of heart disease. Different classification methods such as Neural Networks and Decision Trees were applied to predict the presence of heart disease and to identify the most significant factor which contributes for the cause of the disease, while association rule discovery was used to identify the effect of diet, lifestyle, and environment on the outcome of the disease. Clustering algorithms like the k-means algorithm were used on heart disease data warehouse which contains screening clinical data of patients to identify instances which are more relevant to heart attack. The results showed a bright future of data mining in the diagnosis of heart disease.

1.1.1 statement of the problem

WHO, (2011) reported Cardiovascular Diseases (CVDs) are the number one cause of death globally: more people die annually from CVDs than from any other cause. An estimated
17.1 million people died from CVDs in 2004, representing 29% of all global deaths, of these deaths, an estimated 7.2 million were due to coronary heart disease which is one of the most common types of heart disease and 5.7 million were due to stroke.

Low- and middle-income countries are disproportionally affected, 82% of CVD deaths take place in low- and middle-income countries and occur almost equally in men and women. By 2030, almost 23.6 million people will die from CVDs, mainly from heart disease and stroke. These are projected to remain the single leading causes of death. The largest percentage increase will occur in the Eastern Mediterranean Region. The largest increase in number of deaths will occur in the South-East Asia Region due to change in lifestyle, work culture and food habits. Hence, more careful and efficient methods of cardiac diseases and periodic examination are of high importance (WHO, 2011).

Another report by WHO shows that Cardiovascular Disease and Ischemic heart disease together account 6% of total deaths in Ethiopia for all ages, which makes them the 7th and 8th deadliest diseases in Ethiopia and persons dying from heart disease are expected to growdrastically partly as a result of increasing longevity, urbanization, lifestyle changes, work culture changes and food habits changes (WHO, 2006). In order to decrease mortality from heart diseases there should be a fast and effective detection method especially, in developing countries like Ethiopia where there is a shortage of specialists and wrongly diagnosed cases are high. Data mining can be a convenient tool to assist physicians in detecting the disease by obtaining knowledge and information regarding the disease from patient's data.

Deciding on doctors' intuition and experience rather than on the knowledge-rich data hidden in the database leads to unwanted biases, errors and excessive medical costs which affect the quality of service provided to patients. Wu et al. (2002) proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, and decrease unwanted practice variation. 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.

DOI: 10.17148/IJARCCE.2022.11915

The purpose of this study is, therefore to apply data mining techniques for extracting hidden patterns, which are significant to heart diseases, from a data collected by International Cardiovascular Hospital.

1.2 Objective

1.2.1 General Objectives

The general objective of this study is to design a predictive model for heart diseasedetection using data mining techniques from Transthoracic Echocardiography Report dataset that is capable of enhancing the reliability of heart disease diagnosis using echocardiography.

1.2.2 Specific Objectives

- To identify key patterns or features from the dataset.
- To identify and select attributes that are more relevant in relation to heart diseasediagnosis.
- To compare Decision Tree, Neural Network and Bayesian classifiers in predictingheart disease cases.
- To interpret and analyze the results of the selected model with the help of domainexpert.

1.3 Research methodology

1.3.1 Research Design

The process starts with determining the KDD goals, and ends with the implementation ofthe discovered knowledge. This methodology is selected for this specific study because of three reasons. First, the KDD methodology is best suited for academic researches. Second, as an outsider for the domain using KDD methodology reduces the skill required for the knowledge discovery. Third, the KDD methodology is also independent from any tools and techniques, so one can follow any desired technique during the study. So, based on the KDD methodology, for this specific data mining project, the following nine steps were undertaken.

2. UNDERSTANDING THE APPLICATION DOMAIN

To define the problem and determine medical goals, the researcher has worked closely with the hospital’s Echo Department head, who is also a consultant cardiologist at the hospital. The discussions with the domain expert have helped the researcher to learn about the problem and to know about current solutions to those problems. Since the knowledge gained from the domain expert is a high-level description of the problem from the medical point of view, a literature review was carried out and relevant works related to data mining and heart disease have been reviewed to have more knowledge about the domain. Furthermore, a real time observation of the system was performed to understand the business process of the hospital.

3. LITERATURE SURVEY

Due to a wide availability of huge amount of data and a need to convert this available huge amount of data to useful information necessitates the use of data mining techniques. Data Mining and KDD have become popular in recent years. The popularity of data mining and KDD shouldn’t be a surprise since the size of the data collections that are available are far too large to be examined manually and even the methods for automatic data analysis based on classical statistics and machine learning often face problems when processing large, dynamic data collections consisting of complex objects.

The abundance of data, coupled with the need for powerful data analysis tools, has been described as a data rich but information poor situation. The fast-growing, tremendousamount of data, collected and stored in large and numerous data repositories, has far exceeded our human ability for comprehension without powerful tools.

Fayyad et al. (1996) historically the notation of finding useful patterns in data have beengiven a variety of names including data mining, knowledge extraction, information discovery, information harvesting, data archaeology and data pattern processing but recently the terms data mining and KDD are dominating in the Management Information Science (MIS) communities and database fields.

KDD is an automatic, exploratory analysis and modelling of large data repositories. KDD is the organized process of identifying valid, novel, useful, and understandable patterns from large and complex datasets. Data Mining is the core of the KDD process, involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns. The model is used for understanding phenomena from the data, analysis and prediction (Oden and Lior, 2005).

Just a few short years ago, few people had even heard of the term data mining. Though data mining is the evolution of a field with a long history, the term itself was onlyintroduced relatively recently, in the 1990s.

As Han and Kamber (2006) described the building blocks of today's data miningtechniques date back to 1960s where database and information technology has been evolving systematically from primitive file processing systems to sophisticated and powerful database systems. Since the 1970s database systems has progressed from early hierarchical and network database systems to the development of relational databasesystems (where data are stored in relational table structures), data modelling tools, and indexing and accessing methods. In addition, users gained convenient and flexible data access through query languages, user interfaces, optimized query processing, and transaction management.

As Hand et al. defined it "Data mining is the analysis of (often large) observational datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner". The definition above refers to "observational data," as opposed to "experimental data." Data mining typically deals with data that have already been collected for some purpose other than the data mining analysis. This means that the objectives of the data mining exercise play no role in the data collection strategy. This is one way in which data mining differs from much of statistics, in which data are often collected by using efficient strategies to answer specific questions. For this reason, data mining is often referred to as "secondary" data analysis.

According to Connolly et al. Data mining is "a process of extracting valid, previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions".

The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. In this regard, Thearling (2000) notes that data mining tools can answer business questions that traditionally were too time-consuming to resolve. These tools scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

To summarize, data mining is a way to find previously unknown, valid patterns and relationships from huge amount of data represented in qualitative, textual or multimedia formats by applying different data analysis tools and also most of the time the datasets are collected for other purposes.

SEMMA, CRISP-DM and KDD are popular data mining process models that are used in data mining projects.

KDD process, as presented in Fayyad et al. (1996), is the process of using data mining methods to extract what is deemed knowledge according to the specification of measures and thresholds, using a database along with any required preprocessing, sub sampling, and transformation of the database. The KDD process model is adopted for this study.

According to Fayyad et al. The KDD process is interactive and iterative, involving numerous steps with many decisions being made by the user. Each step attempts to complete a particular discovery task and each accomplished by the application of a discovery method. Knowledge discovery concerns the entire knowledge extraction process, including how data are stored and accessed, how to use efficient and scalable algorithms to analyze massive datasets, how to interpret and visualize the results, and how to model and support the interaction between human and machine. It also concerns support for learning and analyzing the application domain.

Fayyad et al. defined KDD as a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. According to this definition, data is a set of facts that is somehow accessible in electronic form. The term "patterns" indicates models and regularities which can be observed within the data. Patternshave to be valid, i.e they should be true on new data with some degree of certainty.

KDD and data mining are often used interchangeably in some literatures, according to Chen et al. data mining, which is also referred to as knowledge discovery in databases (KDD), is defined as a process of extracting nontrivial, implicit, previously unknown and potentially useful information (such as knowledge, rules, constraints, regularities) from datain databases.

DOI: 10.17148/IJARCCE.2022.11915

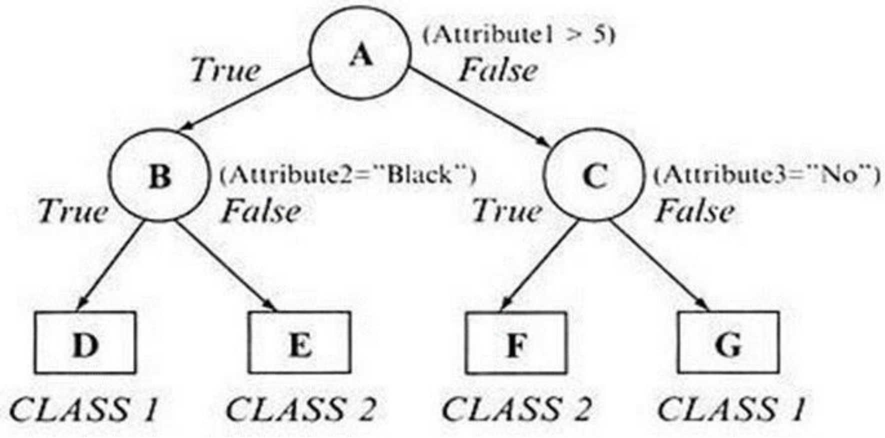
4. SYSTEM ARCHITECTURE

This study aims at applying classification techniques for heart disease detection. An attempt was made to construct prediction model using Decision Tree, Neural Network and Bayesian Classifier. After constructing the models, performance of each models were evaluated, and also their performances were compared to each other. In this section the algorithms used to build the models and matrices used for performance measure and comparison are discussed in detail. To define the problem and determine medical goals, the researcher has worked closely with the hospital's Echo Department head, who is also a consultant cardiologist at the hospital. The discussions with the domain expert have helped the researcher to learn about the problem and to know about current solutions to those problems. Since the knowledge gained from the domain expert is a high-level description of the problem from the medical point of view, a literature review was carried out and relevant works related to data mining and heart disease have been reviewed to have more knowledge about the domain. Furthermore, a real time observation of the system was performed to understand the business process of the hospital.

4.1 Decision Trees

Han and Kamber (2006) defined decision tree as a flowchart like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node.\

Figure 4.1: A simple Decision Tree



As the construction of decision tree classifiers does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery they have become popular. Decision trees can handle high dimensional data.

Their representation of acquired knowledge in tree form is intuitive and generally easy to assimilate by humans. The learning and classification steps of decision tree induction are simple and fast. In general, decision tree classifiers have good accuracy. However, successful use may depend on the data at hand.

Decision tree induction algorithms have been used for classification in many application areas, such as medicine, manufacturing and production, financial analysis, astronomy,and molecular biology (Han and Kamber, 2006).

4.2 J48 Classifier Algorithm

During the late 1970s and early 1980s, J. Ross Quinlan, a researcher in machine learning, developed a decision tree algorithm known as ID3 (Iterative Dichotomiser). This work expanded on earlier work on concept learning systems, Quinlan later presented C4.5 (a successor of ID3), which became a benchmark to which newer supervised learning algorithms are often compared.

C4.5 algorithm is an improvement of IDE3 algorithm. It is based on Hunt's algorithm andalso like IDE3, it is serially implemented. Pruning takes place in C4.5 by replacing the internal node with a leaf node thereby reducing the error rate. Unlike IDE3, C4.5 accepts both continuous and categorical attributes in building the decision tree. It has an enhanced method of tree pruning that reduces misclassification errors due noise or too- muchdetails in the training data

DOI: 10.17148/IJARCCE.2022.11915

set. Like IDE3 the data is sorted at every node of the tree in order to determine the best splitting attribute. It uses gain ratio impurity method to evaluate the splitting attribute (Quinlan, 1993).

C4.5 uses this concept of entropy as follows. Suppose that we have a candidate split S,which partitions the training dataset T into several subsets, T1, T2, Tk . The mean information requirement can then be calculated as the weighted sum of the entropies forthe individual subsets, as follows:

Where Pi represents the proportion of records in subset i. We may then define our information gain to be gain(S) = H(T) − Hs(T), that is, the increase in information produced by partitioning the training data T according to this candidate split S. At each decision node, C4.5 chooses the optimal split to be the split that has the greatest information gain, gain(S).
Han and Kamber stated that, when a decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Tree pruning methods address this problem of overfitting the data. Such methods typically use statistical measures to remove the least reliable branches.

The second and more common approach is post pruning, which removes sub trees from a “fully grown” tree. A sub tree at a given node is pruned by removing its branches and replacing it with a leaf. The leaf is labelled with the most frequent class among the sub tree being replaced.

As described by Quinlan (1993) J48 Decision tree classifier follows the following simple algorithm. J48 builds decision trees from a set of labelled training data using the concept of information entropy. It uses the fact that each attribute of the data can be used to makea decision by splitting the data into smaller subsets. J48 examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. To make the decision, the attribute with the highest normalized information gain is used. Then the algorithm recurs on the smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then a leaf node is created in the decision tree telling to choose that class. But it can also happen that none ofthe features give any information gain. In this case J48 creates a decision node higher up in the tree using the expected value of the class.

4.3 Bayesian Classifier

Han and Kamber (2006) stated that Bayesian classifiers are statistical classifiers that are based on Bayes’ theorem. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Bayes’ theorem is named after Thomas Bayes, a nonconformist English clergyman who did early work in probability and decision theory during the 18th century. Let X be a data tuple. In Bayesian terms, X is considered “evidence.” As usual, it is described by measurements made on a set of n attributes. Let H be some hypothesis, such that the data tuple X belongs to a specified class C. For classification problems, we want to determine (|) the probability that the hypothesis H holds given the “evidence” or observed data tuple

X. In other words, we are looking for the probability that tuple X belongs to class C, giventhat we know the attribute description of X.

The probabilities P(H) and P(X) may be estimated from the given data. Bayes’ theorem isuseful in that it provides a way of calculating the posterior probability, from P(H) and P(X) as shown in equation 3.3.

4.4 Naïve Bayes

The Naïve Bayes classifier is based on the Bayes rule of conditional probability. It makes use of all the attributes contained in the data, and analyses them individually as though they are equally important and independent of each other. Various empirical studies of this classifier in comparison to decision tree and neural network classifiers have found it to be comparable in some domains. In theory, Bayesian classifiers have the minimum error rate in comparison to all other classifiers. However, in practice this is not always the case, owing to inaccuracies in the assumptions made for its use, such as class conditional independence, and the lack of available probability data (Han and Kamber, 2006).
As described by Han and Kamber (2006) the naïve Bayesian classifier, or simple Bayesian classifier, works as follows: 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.

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 naïve Bayesian classifier predicts that tuple X

DOI: 10.17148/IJARCCE.2022.11915

belongs to the class Ci if and only if As P(X) is constant for all classes, only (|) ( ) need 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 (|) ( ). Note that the class prior probabilities may be estimated by number of training tuples of class Ci in D. Given datasets with many attributes, it would be extremely computationally expensive to compute In order to reducecomputation in evaluating, the naive assumption of class conditional independence is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes).

4.5 Performance Measures

In order to minimize the bias associated with the random sampling of the training and test data samples k-Fold Cross Validation was adopted. In k-fold cross-validation, the initial data are randomly partitioned into k mutually exclusive subsets or “folds,” D1, D2, : : : , Dk, each of approximately equal size. Training and testing is performed k times (Han and Kamber, 2006).

5. SYSTEM DESIGN

This study is based on a data collected by the International Cardiovascular Hospital. So, to understand the medical domain, the researcher has worked closely with the hospital’s administration and Echo department. In addition, real time observation of the business process was performed to gain an insight how the hospital functions.

International Cardiovascular Hospital is one of the two privately owned cardiac hospitals in Addis Ababa. The hospital is established with the aim of serving the needs of the local community related to cardiac diseases by providing the highest quality health care to patients and their families. International Cardiovascular Hospital aspires to transform the future of healthcare, through filling the gap created by the limited ability of government owned hospitals to cope up with the growing cardiovascular disease.

To understand the medical problem domain, the first step was identifying the current procedures that are used to diagnose heart disease, after the procedures are identified the researcher continued to define the problems which occur during the diagnosis of the disease. Finally medical goals were determined and data mining goals were set to identify and prepare data required for the study.

5.1 Business Process Description

Basically there are three types of patients who come to the hospital: new patients, referred patients and patients who had examination at the hospital at least once. The process of medical examination varies depending on the patient type, thus the process of medical examinations are discussed separately for each type of patient.

If the patient is new he/she directly requests for the service at the reception desk and the receptionist fills all the required fields on the patients’ medical chart and after completing the registration the receptionist will issue an index card and a receipt upon payment for the patient and will send the patient’s medical chart to the nurses’ station.

After taking vital signs from patients the nurses will send the patient to a waiting roomand the patient’s medical chart to the physician. When the physician is ready to examine the patient one of the nurses will take the patient to examination room and the physician diagnosis the patient, depending on the symptoms, then the physician may advices and discharges the patient if he/she is disease free or otherwise may request laboratory tests (including echocardiography examination), after the patient undergoes a laboratory testthe result will be sent to the physician from the laboratory directly.

Referred patients come to the hospital for two reasons; the first is for only laboratory tests’ including ECG, Echocardiography and stress test, the second is for specialized medical examination. Patients that come for laboratory tests don’t need to go through registration and medical examination like new patients all they have to do is pay the fee required for the laboratory test, undergo the test and take the result with them to the health institution that referred them, but patients that are referred for specialized medical examination might need to go through all the steps like new patients.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 9, September 2022

DOI: 10.17148/IJARCCE.2022.11915

5.2 Heart Diseases Diagnosis

There are many methods that can be used to diagnose for possible heart disease. Usually the recognition of the disease starts with identifying the patient’s history, habits such as smoking, life style and by preforming the physical examination including blood pressure, blood chemistries, ECG, and Echocardiography. The healthcare provider will try to choose the testing modality that will best provide the diagnosis.

To understand the medical problem related to heart disease diagnosis using echocardiography test the focus will be on echocardiography, processes of the echocardiography examination and different types of echocardiography tests.

5.3 Echocardiography

Echocardiography is one of the most widely used diagnostic tests for heart disease.The term echocardiography refers to a group of tests that utilize ultrasound to examinethe heart and record information in the form of echoes, i.e. reflected sonic. In addition to providing single-dimension images, known as M-mode echo that allows accurate measurement of the heart chambers, the echocardiogram also offers far more sophisticated and advanced imaging. This is known as two- dimensional (2-D) Echo andis capable of displaying a cross-sectional "slice" of the beating heart, including thechambers, valves and the major blood vessels that exit from the left and right ventricle (Joel and Robert, 1976).

A standard echocardiography or Transthoracic Echocardiography which is the most common type of echocardiogram examination generally lasts between 15–30 minutes. The patient lies bare-chested on an examination table. A special gel is spread over the chest to help the transducer make good contact and slide smoothly over the skin. The transducer, also called a probe, is a small handheld device at the end of a flexible cable.

The transducer, essentially a modified microphone, is placed against the chest and directs ultrasound waves into the chest. Some of the waves get echoed (or reflected) back to the transducer. Since different tissues and blood reflect ultrasound waves differently, these sound waves can be translated into a meaningful image of the heart that can be displayed on a monitor or recorded on paper or tape (See Figure 4.1).

Finally, the physician reviews the examination prior to completion of the final report. Thepatient does not feel the sound waves, and the entire procedure is painless (Lee *et al.*, 2001).

Echocardiography is usually performed in the cardiology department at a hospital, but may also be performed in a cardiologist's office or an outpatient imaging center. Because the ultrasound scanners used to perform echocardiography are portable (handheld) or mobile, echocardiography can be performed in the hospital's emergency department or at the bedside of patients who cannot be transported to the cardiology department.

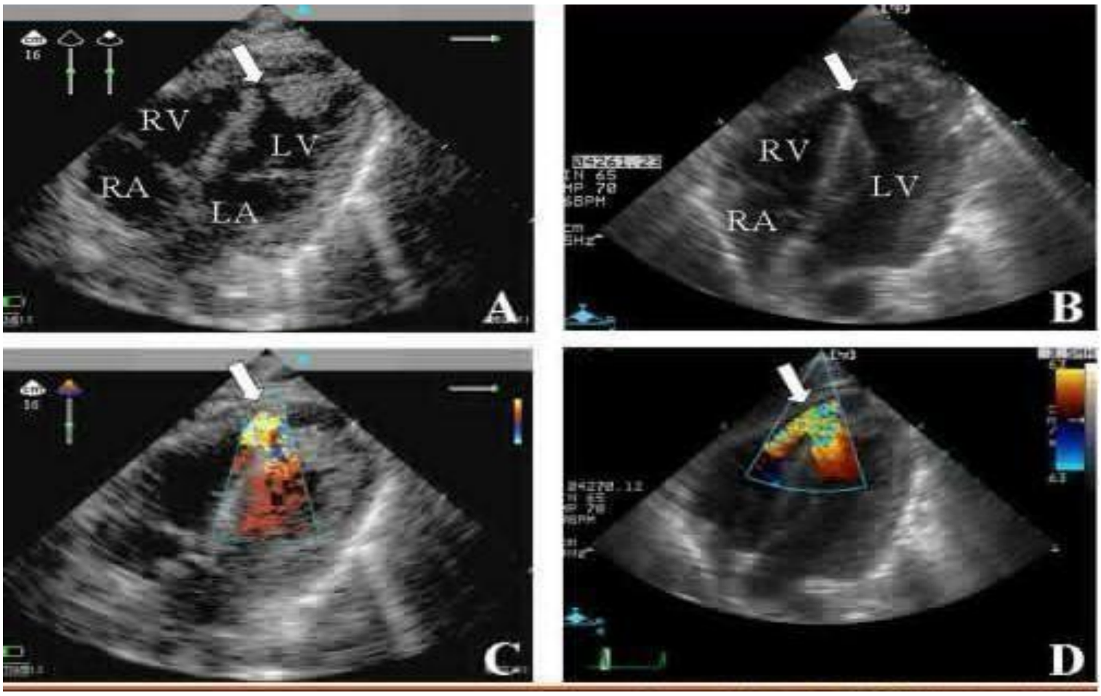


Figure 5.1: Images from echocardiography examination

DOI: 10.17148/IJARCCE.2022.11915

Occasionally, variations of the echocardiography test are used. This includes Doppler echocardiography and Stress echocardiography.

Doppler echocardiography: Doppler is a special part of the ultrasound examination that assesses blood flow (direction and velocity). In contrast, the M-mode and 2-D Echo evaluates the size, thickness and movement of heart structures (chambers, valves, etc.).
During the Doppler examination, the ultrasound beams will evaluate the flow of blood as it makes it way though and out of the heart. This information is presented visually on the monitor (as color images or gray scale tracings and also as a series of audible signals witha swishing or pulsating sound).

Stress echocardiography: or exercise echo, is an echocardiogram performed during exercise, when the heart muscle must work harder to supply blood to the body. This allows physicians to detect heart problems that might not be evident when the body is at rest and needs less blood. For patients who are unable to exercise, certain drugs can be used to mimic the effects of exercise by dilating the blood vessels and making the heart beat faster. During the examination the sonographer can take measurements and, usingthe ultrasound scanner's computer, make calculations, including measuring blood flow speed.
Most ultrasound scanners are equipped with videotape recorders or digital imaging/archiving devices to record the real-time examination, and with medical image printers to print out hard copies of still images.

5.4 Medical Problem Definition

After discussing the current issues related to heart disease diagnosis with the hospital's Echo Department head, who is a consultant cardiologist the researcher was ableto understand that interpretation of echo recordings remains a challenge since there are noavailable precise rules that are deduced from databases. The rules they are following now are gained through experience and professional learning.
There are limited number of specialists and a high number of patients, because of this the hospital is forced to use junior cardiologists which lead to inconsistencies and errors, currently there is only one specialist who does all the echocardiography examinations facing 400 up to 700 examinations per month in addition to his regular duties.
These inconsistencies and errors arise from being solely dependent on experience rather than linking experience with other useful tools like rules that can be generated from existing data. Thus, there is a need for a tool to assist in the diagnosis process that isbased on rules that are generated from the data collected by the hospital.

5.4.1 Defining Medical Goals

To solve the problems that exist in the current system, the following medical goals were set.The medical goals set for this study are:

- To create a tool that could be used in heart disease diagnostic process by assistingcardiologists to interpret echo recordings.
- To make the interpretation procedure easier, more consistent, and efficient based on the rules that are reformulated by the system.

5.4.2 Defining Data Mining Goals

The data mining goals are translations of the medical goals, here the goals are settowards the technical part of the solution.
The main data mining goals are:

- Given patients' echo result for each attribute, classify patients into two categories; thosewho are diagnosed with heart disease and those who are free from the disease.
- To identify key features or patterns from the dataset.
- To Identify and select attributes that are significant in relation to the predictable state –heart disease.

To achieve these data mining goals it was required to use different classificationalgorithms such as Neural Network, Decision Trees and Bayesian classifiers.

Selecting and Creating the Target Dataset

After establishing data mining goals and the project plan, the next step was selecting and creating a target dataset

DOI: 10.17148/IJARCCE.2022.11915

which is suitable for the study. This includes finding out what data is available, obtaining additional necessary data, and then integrating all the data for the knowledge discovery into one dataset, including the attributes that will be consideredfor the process. This process is very important because the Data Mining learns and discovers from the available data.

Data Understanding

Data collected from patients include personal information, medical history, and laboratory test results. These data is usually generated and stored in the context of making
a medical decision, for the purpose of augmenting choices about further testing and/or treatment and also for future access when needed.

Selecting the Target Dataset

The first stage of this step is to select the related data from many available datasets to correctly describe the given medical task. The transthoracic echocardiography report of 7,708 patients from the October, 2008 to March, 2011 with a size of 300 Megabytes was selected as a target dataset for this study and the hospital kindly provided the data. The report contains different measurements that are taken during the echocardiography examination, including information on 20 variables.

Creating the Database

As the hospital keeps the record of each patient in a separate Microsoft Word file, in order to integrate the data it was needed to create a database with variables of interest andrecord the values of each variable into the new database.
After removing these redundant and corrupted files by deleting from the original dataset the researcher was left with 7,339 files. Before creating the new database the researcher turned to a domain expert to select the most valuable attributes which should be included in the new database and the domain expert selected 15 of the most important variables and removed the remaining 5 because of ethical issues.

Description of the Dataset

Each record in the dataset corresponds to a single patient's exam results which are collected during the echocardiography examination. The information measured by the echo from a number of views around the patient is used to construct the report.
The variables in the dataset include Age, Sex, Aortic root – diameter, Left atrium: (sys) diameter, Left ventricle in: diastole, Left ventricle in systole, Posterior wall of LV,Interventricular septum, LV- ejection fraction, Main Pulmonary Artery diameter, Pericardial effusion, TR Velocity, Em/Am velocity ratio, Rhythm, and Diagnosis. The attributes and their description are presented in Table 4.1.

Table 5.1: Attributes and their description

No	Attributes	Description	Type
1	Age	Age of the patient in years	Numeric
2	Sex	Sex of the patient (Male / Female)	Nominal
3	Aortic root – diameter	Size of Aortic root – diameter in mm	Numeric
4	Left atrium: (systole) diameter	Size of Left atrium: (sys) diameter in mm	Numeric
5	Left ventricle in: diastole	Left ventricle in: diastole in mm	Numeric
6	Left ventricle in systole	Size of Left ventricle in systole in mm	Numeric
7	Posterior wall of LV	Size of Posterior wall of LV in mm	Numeric
8	Interventricular septum	Size of in Interventricular septum in mm	Numeric
9	LV- ejection fraction	Fraction of blood pumped out of ventricles with each heart beat in percentage	Numeric
10	Main Pulmonary Artery diameter	Size of Main Pulmonary Artery diameter in cm	Numeric
11	Pericardial effusion	Presence of an abnormal amount and/or character of fluid in the pericardial space	Ordinal
12	TR Velocity	Tricuspid Regurgitation Velocity in cm/sec	Numeric
13	Em/Am velocity ratio	The ratio between myocardial early and atrial peak velocities	Numeric
14	Rhythm	Type of the heart rhythm observed	Nominal
15	Diagnosis	Does the patient has a heart disease (Yes or No)	Nominal

5.1 Descriptive Data Visualization

The dataset includes 3,813 (51.96%) females and 3,526 (48.04%) males age ranging from 4 years to 102 years.

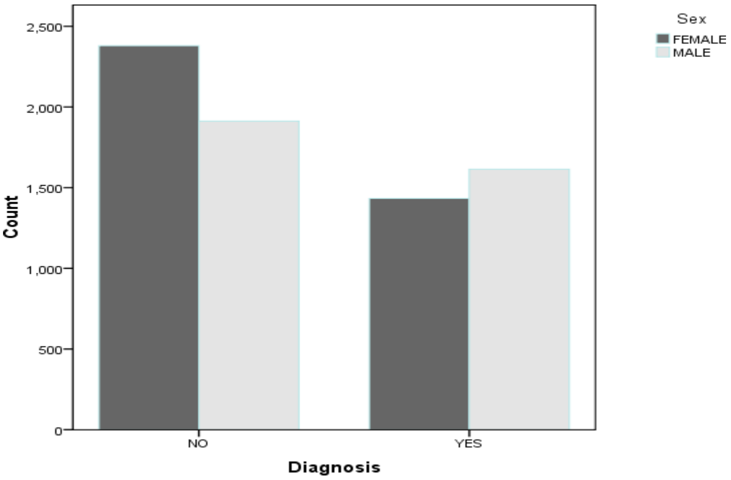


Figure 5.2: Distribution of heart disease among Gender

Figure 5.2 shows that there is a gender gap related to diagnosis. Men who took the testare more likely to have the disease than women who took the test. Out of 3,526 males who were examined 1,614 (45.77%) were diagnosed with the disease while out of 3,813 females who were examined 1,433 (37.85%) were diagnosed with the disease.

The distribution of the disease is more over normal throughout different age groups (See Figure 4.3). However, from all the patients who took the test, patients whose ages were above 60 were slightly susceptible to the disease while patients whose age were less than 40 were free from the disease compared to those who are above 60.

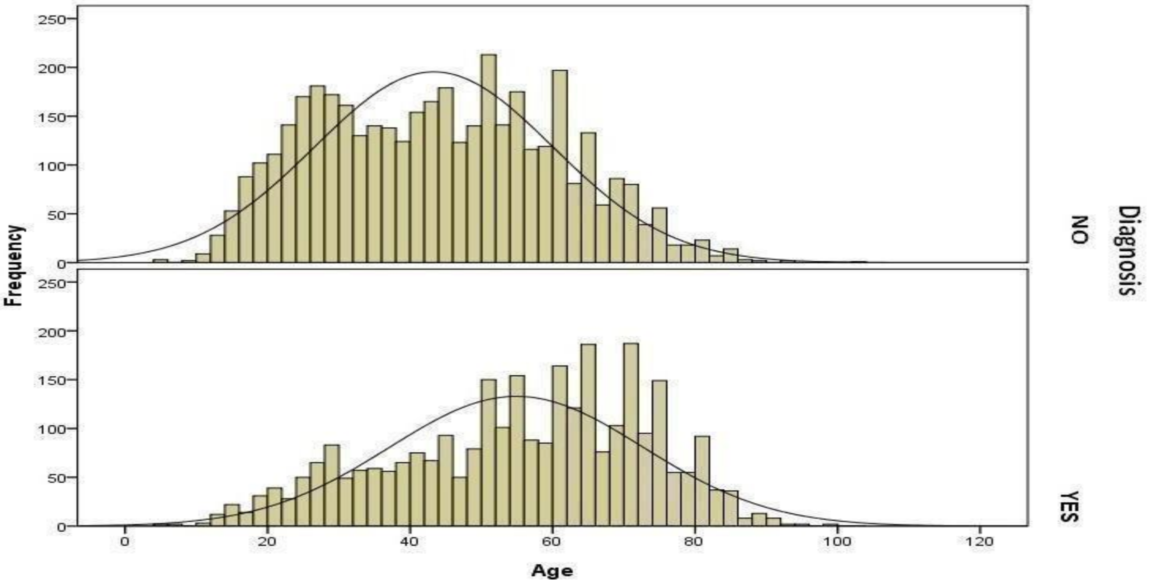


Figure 5.3: Distribution of Heart Disease among different age groups

To summarize attributes with numerical data type measures of centrality and measure of spread were done. To measure centrality, mean (arithmetic average) was used and to measure how the data is spread standard deviation and range (minimum value and maximum value) were used. The result is presented in Table 5.2.

From all the measures of dispersion listed on Table 5.2 standard deviation has a more statistical significance, thus standard deviation is used as a major tool for measure of variation. The age of the patient's included in the dataset has a standard deviation value of 18.128 suggesting there is fairly unique values for age and each patient's age is not tooclose to the average while the standard deviation of Main Pulmonary Artery diameter which is 0.62 suggests every patient is pretty close to average.

6. EXPERIMENTATION

As the goal of this study is to detect heart disease using data mining techniques a classification technique was adopted to develop a predictive model. The models were built with three different supervised machine learning algorithms i.e. Decision Tree Classification Algorithm, Bayesian Classifier and Neural Network using Weka 3.6.4machine learning software.

6.1 Experimental Setup

Four experiments were conducted for this study and for all experiments two scenarios were considered, one containing all the 15 attributes and the other containing 8 selected attributes. With four experiments and eight different scenarios a total of eight modelswere developed.

The experiments were conducted on a full training dataset containing 7,339 instances and 10- Fold Cross Validation was adopted for randomly sampling the training and test sets. While performing the experiments all parameters were set to their default setting for each algorithm except for J48 classifier where the parameter "Unpruned" which had a default value "False" was changed to "True" for the first experiment to observe the performance of J48 unpruned tree.

The performances of the models in this study were evaluated using the standard metrics of accuracy, precision, recall and F-measure which were calculated using the predictive classification table, known as Confusion Matrix. ROC area was also used to compare the performances of the classifiers.

6.1.1 Model Building Using J48 Decision Tree

Two experiments were conducted using J48 decision tree classifier. These twoexperiments were designed to investigate:

The effect of attribute selection on classification accuracy as well as modelcomplexity on both unpruned and pruned J48 Decision Tree Classifiers.

The effect of tree pruning methods on classification accuracy, Decision Tree size andmodel complexity when building a J48 decision tree model

Experiment 1

The first experiment was designed to evaluate the performance of a J48 classifier Unpruned tree in predicting heart disease and to investigate the effect of attribute selection on the performance of the model. In this experiment two scenarios were considered, one containing all 15 attributes and the other containing the selected 8 attributes.

On the first scenario the algorithm was run on a full training set containing 7,339 instances with 15 attributes. It took 0.89 second to build the model and the modelgenerated a tree with a size of 473 and 323 leaves.

On the second scenario the algorithm was run on a full training set containing 7,339 instances with selected 8 attributes. It took 0.36 second to build the model and the model generated smaller and less complex tree with a size of 126 and 71 leaves making it less complex and faster than the experiment conducted on all attributes.

Table 6.1: Confusion Matrixes for Experiment 1

Model	**Confusion Matrix**		
J48 unpruned with all attributes	**Yes (Predicted)**	**No (Predicted)**	**Actual**
	2,841	206	**Yes**
	213	4,079	**No**
J48 unpruned with Selected attributes	**Yes (Predicted)**	**No (Predicted)**	**Actual**
	2,875	172	**Yes**
	157	4,135	**No**

DOI: 10.17148/IJARCCE.2022.11915

Table 6.2 Detailed Performance Measures for Experiment 1

Model	Accuracy	TP Rate	TN Rate	Precision	F-Measure	ROC Area
J48 unpruned with all attributes	94.29 %	0.932	0.95	0.943	0.943	0.942
J48 unpruned with Selected attributes	95.52 %	0.944	0.963	0.955	0.955	0.965

The model built with J48 unpruned tree with all attributes correctly classified (predicted the correct outcome) 6,920 (94.29%) instances while 419 (5.71%) of the instances were classified incorrectly. The overall accuracy rate of the model is highly successful, but it is worth to consider the TP Rate (Sensitivity), that is, patients which have heart disease and that are correctly identified and TN Rate (Specificity), that is, patients which do not model correctly identified 2,841 patients out of 3,047 patients who had heart disease and the remaining 206 wereidentified incorrectly to be free from the disease while they actually had the disease. This result gave the model a TP Rate of 0.932. The model is better in identifying negative cases as the TN Rate of the model is 0.95 by correctly identifying 4,079 patients out of 4,292 patients who didn't had heart disease and the remaining 213 were identified to have the disease while they actually didn't.Regarding to Precision score of the model, 93% of patients labeled as belonging to class Yes does indeed belong to class Yes while 95.2% of patients labeled as belonging to class No does indeed belong to class No. With an average precision of 94.3% it is a very successful model in retrieving relevant values for each class. With F-Measure value of 0.943 it can be concluded that the Precision and the Recall of the model are significantly balanced.

The second model built with J48 unpruned tree with selected 8 attributes correctly classified 7,010 (95.52%) instances while 337 (4.48%) of the instances were classified incorrectly. Like the experiment the overall accuracy rate of the model is highly successful, in fact it is better compared to J48 unpruned tree implemented on all attributes.

The model correctly identified 2,872 patients out of 3047 patients who had heart disease and the remaining 172 were identified incorrectly to be free from the disease while they actually had the disease. This result gave the model a TP Rate of 0.944. The model is better in identifying negative cases as the TN Rate of the model is 0.963 by correctly identifying 4,135 patients out of 4,292 patients who didn't had heart disease and the remaining 157 were identified to have the disease while they actually didn't.

Regarding to Precision score of the model, 94.8% of patients labeled as belonging to class Yes does indeed belong to class Yes while 96% of patients labeled as belonging to class No does indeed belong to class No. With an average precision of 95.5% it is a very successful model in retrieving relevant values for each class. With F-Measure value of 0.955 it can be concluded that the Precision and the Recall of the model are significantly balanced.

The results of this experiment indicated that a J48 unpruned decision tree algorithm is highly capable in predicting heart disease cases. Furthermore, the results showed the impact of attribute selection on classification accuracy, Decision tree size and model complexity.

Experiment 2

The second experiment was designed to investigate:

- The performance of a J48 classifier pruned tree in predicting heart disease
- The effect of attribute selection on the performance of a J48 classifier pruned treemodel.
- The effect of tree pruning methods when building a J48 decision tree model,

Like Experiment 1 in this experiment two scenarios were considered, one containing all 15 attributes and the other containing the selected 8 attributes. On the first scenario the algorithm was run on a full training set containing 7,339 instances with 15 attributes. It took 1.05 second to build the model and the model generated smaller and less complex tree with a size of 104 and 63 leaves.

On the second scenario the algorithm was run on a full training set containing 7,339 instances with only 8 selected attributes. It took 0.41 second to build the model and the model generated smaller and less complex tree with a size of 93 and 52 leaves making it the least complex model built from J48 classifier.

Table 6.3: Confusion Matrixes for Experiment 2

Model	**Confusion Matrix**		
J48 pruned with all attributes	**Yes (Predicted)**	**No (Predicted)**	**Actual**
	2,872	175	**Yes**
	162	4,130	**No**
J48 pruned with selectedattributes	**Yes (Predicted)**	**No (Predicted)**	**Actual**
	2,892	155	**Yes**
	171	4,121	**No**

Table 6.4 Detailed Performance Measures for Experiment 2

Model	**Accuracy**	**TP Rate**	**TN Rate**	**Precision**	**F-Measure**	**ROC Area**
J48 pruned with all attributes	95.41%	0.943	0.962	0.954	0.954	0.964
J48 pruned with selected attributes	95.56%	0.949	0.96	0.956	0.955	0.965

The predictive model built from a J48 classifier pruned tree with all attributes correctly classified 7,002 (95.41%) instances while 337 (4.59%) of the instances were classified incorrectly. Like the experiment 1 the overall accuracy rate of the model is highly successful, in fact it is better compared to J48 unpruned tree. The model correctly identified 2,872 patients out of 3047 patients who had heart disease and the remaining 175 were identified incorrectly to be free from the disease while they actually had the disease. This result gave the model a TP Rate of 0.943. The model is better in identifying negative cases as the TN Rate of the model is0.962 by correctly identifying 4,130 patientsout of 4,292 patients who didn't had heart disease and the remaining 162 were identified to have the disease while they actually didn't.

Regarding to Precision score of the model, 94.7% of patients labeled as belonging to class Yes does indeed belong to class Yes while 95.9% of patients labeled as belonging to class No does indeed belong to class No. With an average precision of 95.4% it is a very successful model in retrieving relevant values for each class. F-Measure value of 0.954 suggests that the Precision and the Recall of the model are significantly balanced.

The predictive model built from a J48 classifier pruned tree with 8 selected attributes correctly classified 7,013 (95.56%) instances while 326 (4.44%) of the instances were classified incorrectly. The model scored a highly successful overall accuracy rate.

The model correctly identified 2,892 patients out of 3047 patients who had heart disease and the remaining 155 were identified incorrectly to be free from the disease while they actually had the disease. This result gave the model a TP Rate of 0.949. The model is better in identifying negative cases as the TN Rate of the model is 0.96 by correctly identifying 4,121 patients out of 4,292 patients who didn't had heart disease and the remaining 171 were identified to have the disease while they actually didn't.

Regarding to Precision score of the model, 94.4% of patients labeled as belonging to class Yes does indeed belong to class Yes while 96.4% of patients labeled as belonging to class No does indeed belong to class No. With an average precision of 95.6% it is a very successful model in retrieving relevant values for each class. With F-Measure value of 0.956 it can be concluded that the Precision and the Recall of the model are significantly balanced.

6.1.1 Model Building Using Naïve Bayes Classifier

The third experiment was designed to evaluate the performance of Naïve BayesClassifier in predicting heart disease. In this experiment two scenarios were considered, one containing all 15 attributes and the other containing the selected 8 attributes. The intention here is to investigate the effect of attribute selection on the performance of the models.

7. RESULTS AND SNAPSHOTS

7.1 Login Page

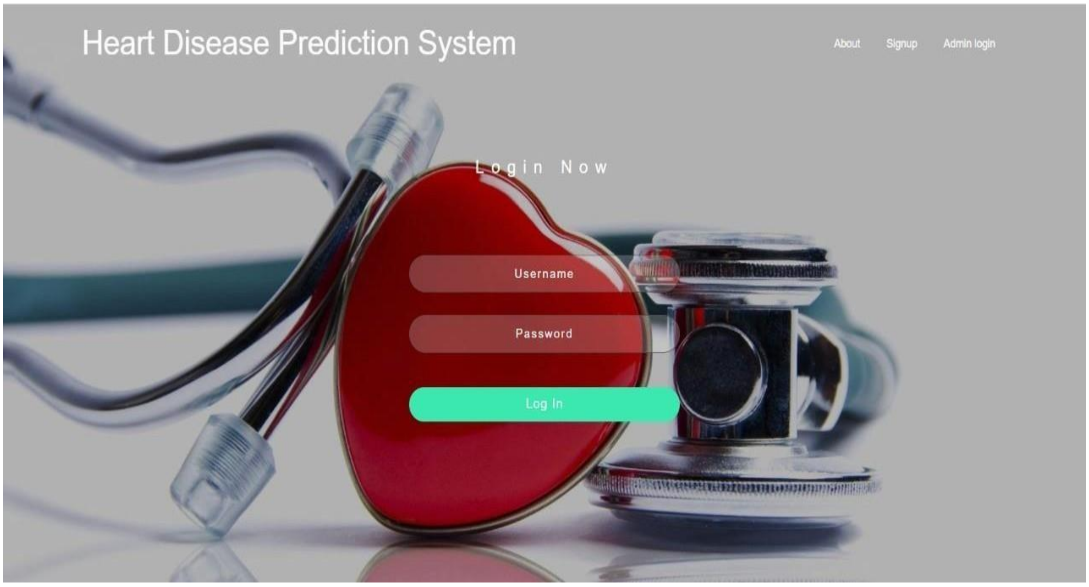
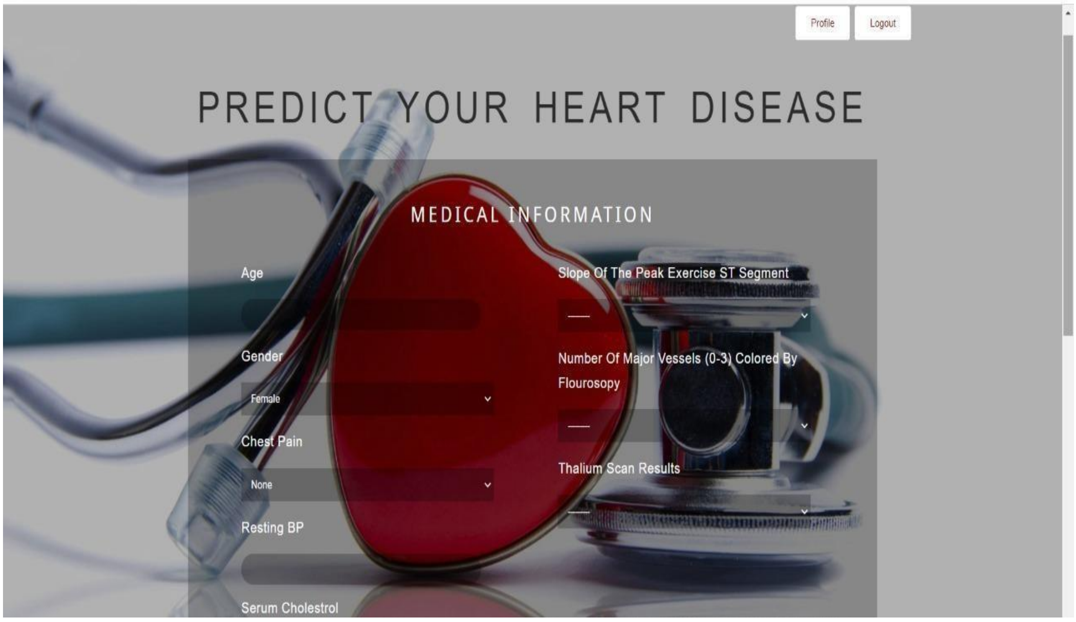


Figure :7.1 Login Page

The above fig:7.1 refers to the Login Page of the Website. It helps the user to login to the site with provided user name and password. This will helps to maintain safety measures and once if the users signup with the website they need not to sign up again and again .

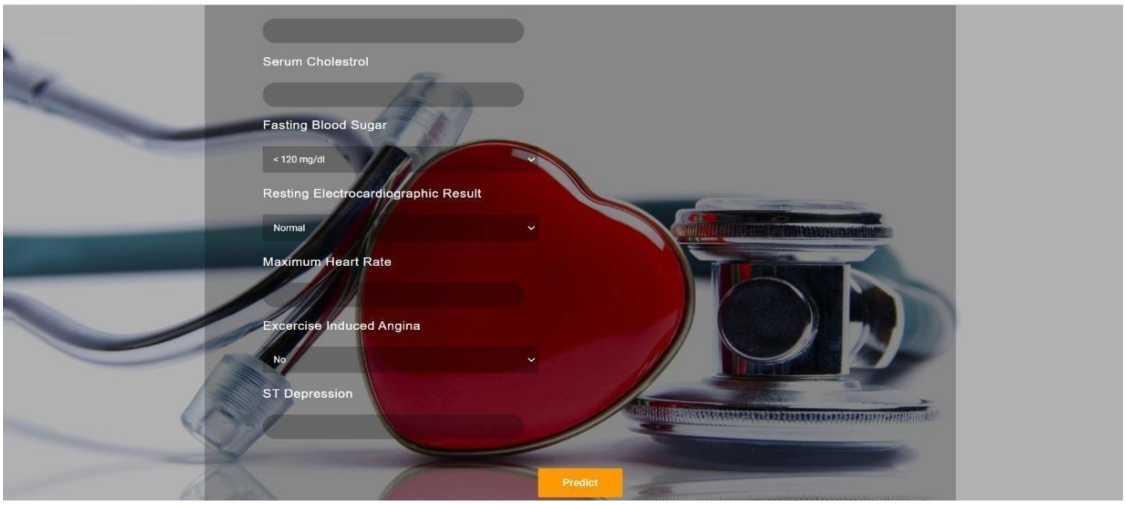
7.2 Medical Information Page


Profile
Logout
PREDICT YOUR HEART DISEASE
MEDICAL INFORMATION
Age
Slope Of The Peak Exercise ST Segment
Number Of Major Vessels (0-3) Colored By Flourosopy
Thalium Scan Results
Resting BP
Serum Cholestrol

Gender: Female

Chest Pain: None

DOI: 10.17148/IJARCCE.2022.11915



Serum Cholestrol

Fasting Blood Sugar

< 120 mg/dl

Resting Electrocardiographic Result

Normal

Maximum Heart Rate

Excercise Induced Angina

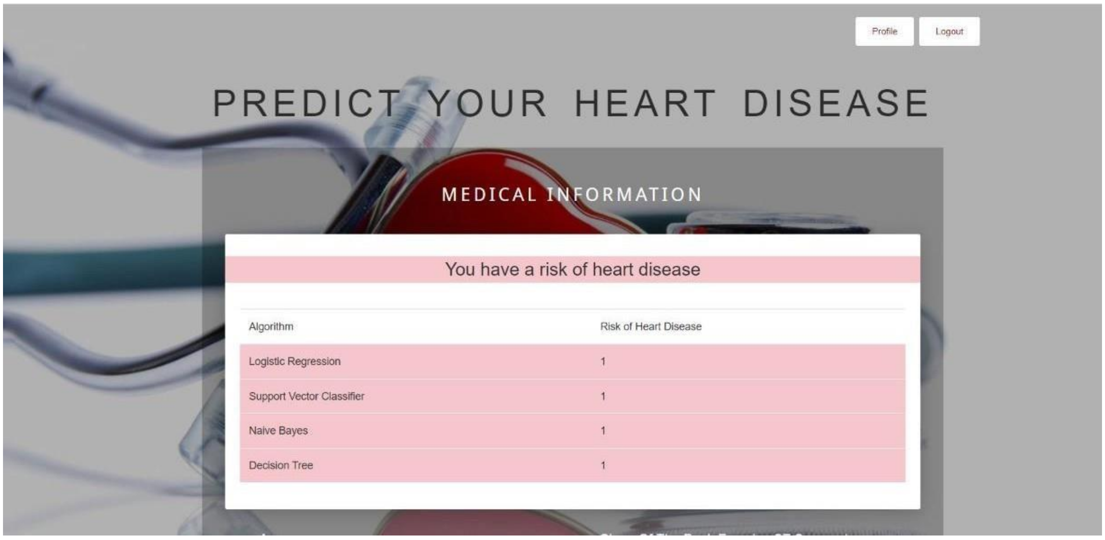
No

ST Depression

Figure :7.2 Medical Information Page

The above fig:7.2 refers to Medical Information Page. By using this page we can add the information like age, gender, and medical information like chest pain, resting BP, serum cholesterol, thalium scan results, fasting blood sugar, maximum heart rate, ST depression.

7.3 Result Page



PREDICT YOUR HEART DISEASE

MEDICAL INFORMATION

You have a risk of heart disease

Algorithm	Risk of Heart Disease
Logistic Regression	1
Support Vector Classifier	1
Naive Bayes	1
Decision Tree	1

Figure:7.3 Result Page

The above fig:7.3 refers to the Result Page. This page shows whether the person have a heart disease or not based on the details we have entered in the medical information page. If all the algorithm shows 1 it indicates the person have heart disease, and all thealgorithm shows 0 this indicates person do not have a heart disease.

7.4 Logistic Regression Graph

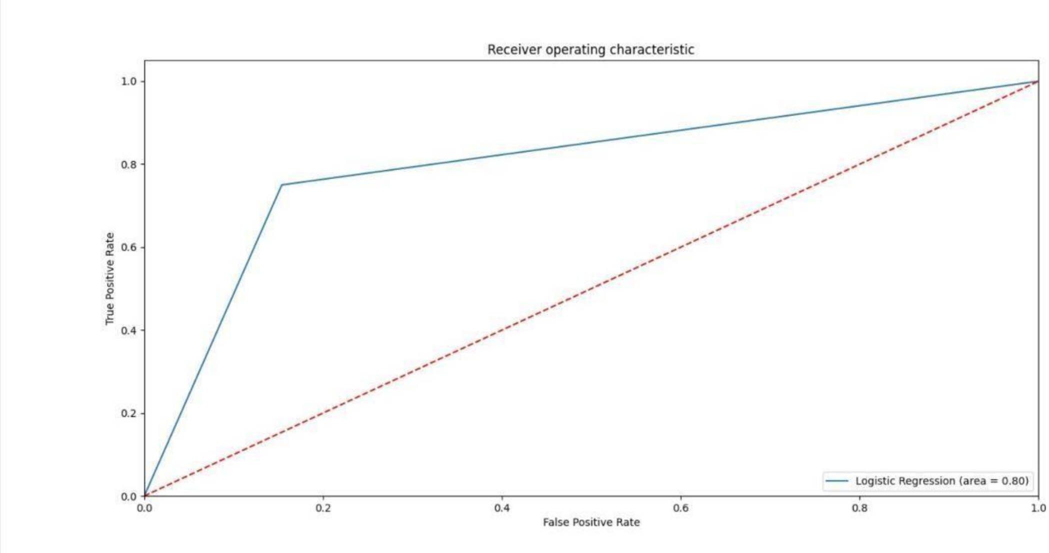


Figure:7.4 Logistric Regression graph

The above fig:7.4 refers to Logistric regression graph. This graph shows the accuracy result of heart disease.

7.5 Bar graph

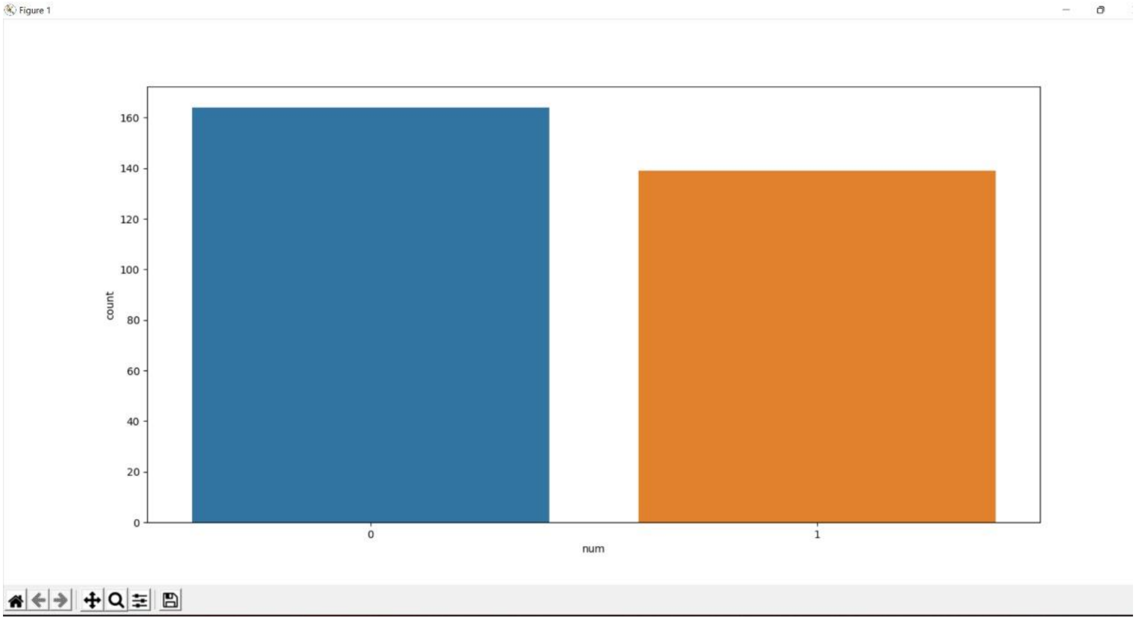


Figure:7.5 Bar graph

The above fig:7.5 refers to Bar graph. In this graph 0 represent the person not having heart disease and 1 represent the person having heart disease. This indicates 160 count of persons not having heart disease and 135 count of persons having heart disease.

CONCLUSION

In this study, the aim was to design a predictive model for heart disease detection using data mining techniques from Transthoracic Echocardiography Report dataset that is capable of enhancing the reliability of heart disease diagnosis using echocardiography.

DOI: 10.17148/IJARCCE.2022.11915

Data collected by International Cardiovascular Hospital from the year 2008 to 2011 containing 7,339 instances was selected and preprocessed for this study. The models werebuilt on the preprocessed Transthoracic Echocardiography dataset with three different supervised machine learning algorithms i.e. J48 Classifier, Naïve Bayes and Multilayer Perception using Weka machine learning software.

The performances of the models were evaluated using the standard metrics of accuracy, precision, recall and F-measure. 10-Fold Cross Validation was adopted for randomly sampling the training and test data samples. All eight models performed well in predicting heart disease cases. The most effective model to predict patients with heart disease appears to be a J48 classifier implemented on selected attributes with a classification accuracy of 95.56%.

Three data mining goals were defined based on the medical problems. The goals were evaluated against the selected model and the selected model built with J48 Decision Tree Algorithm successfully met all the three data mining goals. Significant rules that are useful for predicting the presence of heart disease were extractedfrom the dataset. The domain expert confirmed that most of the rules generated are important in interpretation of echocardiography examinations.

From a total of 15 attributes that were available, 8 attributes that are highly relevant in predicting heart disease from Transthoracic Echocardiography dataset were selected For predicting heart disease could not exceed a classification accuracy of 95.56% and still much remains to fill the gap of 4.44% misclassified cases.

This study showed that data mining techniques can be used efficiently to model and predict heart disease cases. The outcome of this study can be used as an assistant tool by cardiologists to help them to make more consistent diagnosis of heart disease. Furthermore, the resulting model has a high specificity rate which makes it a handy tool for junior cardiologists to screen out patients who have a high probability of having the disease and transfer those patients to senior cardiologists for further analysis.

REFERENCES

[1] Braunwald, E., Douglas, P. Zipes, Peter, L., Robert, B. (1988). Braunwald's Heart Disease*:* A Textbook of Cardiovascular Medicine. Third Edition, Harcourt Brace Jovanovich Inc.

[2] Connolly, T., Begg, C. and Strachan, A. (1999). Database Systems*:* A Practical Approach to Design, Implementation and Management.

[3] Berlin Heidelberg Dunham, M.H. (2003). Data Mining Introductory and Advanced Topics. Pearson Education, Inc., Upper Saddle River, New Jersey.

[4] Berry J.A. Michael and Linoff S. Gordon (2004). Data Mining Techniques for Marketing, Sales, and Customer Relationship Management. Second Edition.

[5] Chakrabarti, S., Earl, C., Frank, E., Ralf Hartmut, G., Han, J., Xia Jiang, Kamber, M., Sam S. Lightstone and others (2009).Data Mining Know It All. Morgan Kaufmann Publishers, Elsevier Inc., Burlington Chen, M.S., Han, J., and Yu, P. S. (1997).

[6] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3033–3049, 2015.

[7] M. Chen, P. Zhou, G. Fortino, "Emotion Communication System," IEEE Access, DOI: 10.1109/ACCESS.2016.2641480, 2016.

[8] M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, "Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring," ACM/Springer Mobile Networks and Applications, Vol. 21, No. 5, pp. 825C845, 2016.

[9] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, "Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System," IEEE Communications, Vol. 55, No. 1, pp. 54–61, Jan. 2017.

[10] J. Wang, M. Qiu, and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform," Journal of Systems Architecture, vol. 72, pp. 69–79, 2017.

[11] What is Telemedicine. Accessed April,22,2022. [online]. Available: https://evisit.com/resources/what-is-telemedecine/.

[12] R.S. Khandpur (2019). Handbook of Biomedical Instrumentation, 3 / Ed. McGraw Hill Education (India) Private Limited. ISBN-13: 978-93-392-0543-0.

[13] Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. In Journal of Applied Biomedicine (Vol. 11, Issue 2, pp. 47–58). University of South Bohemia. https://doi.org/10.2478/v10136-012-0031-x.

DOI: 10.17148/IJARCCE.2022.11915

[14] Anwer, R. M., Torgersson, O., & Falkman, G. (2008). Data Mining in Oral Medicine Using Decision Trees Digital seniors View project Definitional Programming View project. https://www.researchgate.net/publication/242568733.

[15] Chakarverti, M., Sharma, N., & Divivedi, R. R. (n.d.). Prediction Analysis Techniques of Data Mining: A Review. https://ssrn.com/abstract=3350303.

[16] de Macedo, D. D. J., Perantunes, H. W. G., Andrade, R., von Wangenheim, A., & Dantas, M. A. R. (2008). Asynchronous data replication: A national integration strategy for databases on telemedicine network. Proceedings - IEEE Symposium on Computer-Based Medical Systems, 638–643. https://doi.org/10.1109/CBMS.2008.76.

[17] GHEORGHE, M., & PETRE, R. (2014). Integrating Data Mining Techniques into Telemedicine Systems. Informatica Economica, 18(1/2014), 120–130. https://doi.org/10.12948/issn14531305/18.1.2014.11.

[18] Han, J., Kamber, M., & Pei, J. (2011). Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems).

[19] Haraty, R. A., Dimishkieh, M., & Masud, M. (2015). An enhanced k-means clustering algorithm for pattern discovery in healthcare data. International Journal of Distributed Sensor Networks, 2015. https://doi.org/10.1155/2015/615740.

[20] Milovic, B., & Milovic, M. (2012). PREDICTION AND DECISION MAKING IN HEALTH CARE USING DATA MINING. In Arabian Journal of Business and Management Review (Vol. 1, Issue 12).

[21] Ou-Yang, C., Wulandari, C. P., Hariadi, R. A. R., Wang, H. C., & Chen, C. (2018). Applying sequential pattern mining to investigate cerebrovascular health outpatients' re-visit patterns. PeerJ, 2018(7). https://doi.org/10.7717/peerj.5183.

[22] Peral, J., Ferrandez, A., Gil, D., Munoz-Terol, R., & Mora, H. (2018). An ontology-oriented architecture for dealing with heterogeneous data applied to telemedicine systems. IEEE Access, 6, 41118–41138. https://doi.org/10.1109/ACCESS.2018.2857499.

[23] Rahman, M. F., Wen, Y., Xu, H., Tseng, T. L. B., & Akundi, S. (2020). Data mining in telemedicine. In Advances in Telemedicine for Health Monitoring (pp. 103–132). Institution of Engineering and Technology. https://doi.org/10.1049/PBHE023E_ch6.

[24] Rani MCA MPhil, Du., & Professor, A. (n.d.). A Survey on Data Mining Tools and Techniques in Medical Field. In International Journal of Advanced Networking & Applications.

[25] Razzak, M. I., Naz, S., & Zaib, A. (n.d.). Deep Learning for Medical Image Processing: Overview, Challenges and Future.

[26] Shaikh, A., Memon, M., Memon, N., & Misbahuddin, M. (2009). The role of service oriented architecture in telemedicine healthcare system. Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, CISIS 2009, 208–214. https://doi.org/10.1109/CISIS.2009.181.

[27] Singh Chaudhary Devi, V., Midha, N., & Singh, V. (2015). A Survey on Classification Techniques in Data Mining. In International Journal of Computer Science & Management Studies) (Vol. 16). Issue 01 Publishing Month. https://www.researchgate.net/publication/280737179.

[28] Taha Khan, M., Qamar, S., & Ranjan Sinha, R. (2017). Discussing the Role of Data Mining in Development of Evidence Based Decision Support System for e-Healthcare. In International Journal of Applied Engineering Research (Vol. 12). http://www.ripublication.com

[29] Chethan Chandra s basavaraddi, "Performance Evaluation Of Mesh And Position Based Hybrid Routing In MANETs", Irnet International Conference On Computer Science and Engineering(ICCSE)-February 3rd , 2012-Nagpur, ISBN-978-93-81693-17-9.

[30] Chethan Chandra s basavaraddi, "Current Project Work on Routing Protocols

For MANET: A Literature Survey", Irnet International Conference On Computer Scienceand Informatics(ICCSI)-March 9th , 2012- Hyderabad, ISBN-978-93-81693-25-4.

[31] Chethan Chandra S Basavaraddi," A New Routing Algorithm in MANETS: LocationAided Hybrid Routing", Chethan Chandra S Basavaraddi et al ,Int.J.Computer Technology & Applications,Vol 3 (2), 760-765 760 ISSN:2229-6093.

[32] Chethan Chandra S Basavaraddi, "Performance Analysis of Mesh and Position

- Based Hybrid Routing In MANETS: A Comprehensive Study", Chethan Chandra S Basavaraddi et al ,Int.J.Computer Technology & Applications,Vol 3 (2), 804-812 804 ISSN:2229-6093.

[33] Chethan Chandra S Basavaraddi, "A Comparative Analysis Of Two Position

- Based Hybrid Routing Algorithms Over MANETs", /International Journal Of Computational Engineering Research / ISSN: 2250–3005 IJCER | Mar-Apr 2012 | Vol. 2 | IssueNo.2 |540-546 Page 540.

[34] Chethan Chandra S Basavaraddi, "Current Project Work On Routing Protocols for MANET: A Literature Survey", International Journal of Scientific and Engineering Research (IJSER) - Volume 3, Issue 5, May 2012 (ISSN 2229-5518).

[35] Chethan Chandra S Basavaraddi, "Performance Evaluation Of Mesh And Position Based Hybrid Routing In MANETs", International Journal of Scientific and Engineering Research (IJSER) - Volume 3, Issue 5, May 2012 (ISSN 2229-5518).

[36] Chethan Chandra S Basavaraddi,." A Comparative Performance Analysis Of Two Position Based Hybrid Routing Algorithms Under Mobility Speed Over Manets", "International Conference On Recent Trends In Computer Science And Engg.(Icrtcse 2012) held at May 3rd & 4th 2012. Apollo Engineering College Sriperumbudur, Kanchipuram – 602105. Tamil Nadu, South India.

[37] Chethan Chandra S Basavaraddi," A Stable Route Selection in PBHRA for MANETs", National conference Advances in Electronics & communication Technology(NCAECT 2012) May 18th, 2012. Dept of Studies and Research inElectronics Kuvempu University, Shankaraghatta-577451 Shimoga Dist, Karnataka.

[38] Chethan Chandra S Basavaraddi," A PBHRA IN MANETs", National conference on Emerging Mobile Technologies And Policies (NCEMTP-2012) 28th May 2012 to 30th May 2012. Organized by Department of Telecommunication Engineering, M.S. RAMAIAH INSTITUTE OF TECNOLOGY, Bangalore-560054.

[39] Chethan Chandra S Basavaraddi, "A Comparative Analysis Of Two Position Based Hybrid Routing Algorithms Under Mobility Speed Over MANETs", International Journal of Research and Innovation in Computer Engineering, ISSN 2249-6580, Vol 2, Issue 3, June 2012, (285-291).

[40] Chethan Chandra S Basavaraddi, "MANETs Application on Environment", UGC sponsored National conference on Perspectives of Physics in ReducingEnvironmental Pollution, Kalpataru First Grade Science College, Feb 2014, Tiptur-572002.

[41] Chethan Chandra S Basavaraddi, "How hard is English – Kannada Machine Translation", International seminar on Computational linguistics on Indian Languages, held by CDAC,IIIT-Trivandrum & Kerala university, Thrivandrum, feb-2014.

[42] Chethan Chandra S Basavaraddi, "A Typical Machine Translation System for English to Kannada", International Journal of Scientific & Engineering Research, Volume 5, Issue 4, April-2014, ISSN 2229-5518.

[43] Chethan Chandra S Basavaraddi, "Current Project Work on English toKannada Machine Translation System: a Literature Survey on NLP",Int.J.Computer Technology & Applications,Vol 5 (3),1254-1275,2014.

[44] Chethan Chandra S Basavaraddi, "Simultaneous Prediction of Stock Market Investments by Analyzing Sentiments: A Supervised Joint Aspect Model", NCETSE2018.

[45] Chethan Chandra S Basavaraddi, "Privacy policy controlling for OSN users" ISSN (Online): 2347-2820, Volume -4, Issue-8, 2016, International Journal of Electrical, Electronics and Computer Systems (IJEECS).

[46] Chethan Chandra S Basavaraddi, "Single Hop Cryptographic Server Based Data Sharing in Cloud" ISSN (Online): 2347-2820, Volume -4, Issue-8, 2016, International Journal of Electrical, Electronics and Computer Systems (IJEECS).

[47] Chethan Chandra S Basavaraddi, "Hybrid Neuro Fuzzy Network Applied to Face Recognition from Ocluded Images", International Archive of Applied Sciences and Technology Int. Arch. App. Sci. Technol; Vol 10 [2] June 2019 : 222-235 © 2019 Society of Education, India [ISO9001: 2008 Certified Organization], www.soeagra.com/iaast.html, DOI: .10.15515/iaast.0976-4828.10.2.222235.

[48] Chethan Chandra S Basavaraddi, "**Object Tracking Using Hybrid Neuro Fuzzy Network Applied to Face Recognition with Image Samples**" International Journal of New Innovations in Engineering and Technology, Volume 11 Issue 4 September 2019, ISSN: 2319-6319.

[49] Chethan Chandra S Basavaraddi, "Face Recognition Using Hybrid Neuro Fuzzy Network for Occluded Images", International Journal of Science and Research (IJSR),2020, ISSN: 2319-7064, ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583.

[50] Chethan Chandra S Basavaraddi, "Face Recognition from Feed Forward Neural Network for Occluded Images Using Hybrid Neuro Fuzzy Network", International conference on Recent Advancements in Wireless Communications, Signal and Image Processin (ICWCSIP 2020), Organized by Chenni Institute of Technology, from 29th-30th June, 2020.

[51] Chethan Chandra S Basavaraddi, "Face Recognition From Feed Forward Neural Network Using Occluded Images For Automating The Surveillance Using Hybrid Neuro Fuzzy Network", International Journal of Engineering Applied Sciences and Technology, 2020 Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 508-519 Published Online June 2020 in IJEAST (http://www.ijeast.com).

[52] Chethan Chandra S Basavaraddi, "Multiple Object Tracking Using Hybrid Neuro Fuzzy Network Applied to Face Recognition from Feed Forward Neural Network", International Journal of Advanced Research in Computer and Communication Engineering Vol. 9, Issue 7, July 2020, DOI 10.17148/IJARCCE.2020.9707, ISSN (Online) 2278-1021 ISSN (Print) 2319-5940.

[53] Chethan Chandra S Basavaraddi, "Deep Affinity to Multiple Object Tracking Using Hybrid Neuro Fuzzy Network

DOI: 10.17148/IJARCCE.2022.11915

Applied to Face Recognition", Journal of Seybold Report, VOLUME 15 ISSUE 8 2020 ,ISSN NO: 1533-9211.

[54] Chethan Chandra S Basavaraddi, "Deep Learning Based Multiple Object Tracking for Facial Images Using Hybrid Neuro Fuzzy Network", International Journal of Scientific & Engineering Research Volume 11, Issue 8, August-2020 1096 ISSN 2229-5518.

[55] Chethan Chandra S Basavaraddi, "Machine learning based recommendation system on movie reviews using KNN classifiers", ICACSE 2020 Journal of Physics: Conference Series 1964 (2021) 042081 IOP Publishing doi:10.1088/1742-6596/1964/4/042081.

[56] Chethan Chandra S Basavaraddi, "Implementation of Client-Side Deduplication of Encrypted Data with Public Auditing in Cloud Storage", http://www.ijniet.org/issues/volume-17-issue-2-july-2021, ISSN: 2319-6319. UGC Approved Journal-47645 Impact factor - 4.012.

[57] Chethan Chandra S Basavaraddi, "Applying Artificial Intelligence to Studies on Water Quality and Phytoplankton Diversity of Eachnur Tank, Tiptur, Tumkur District, Karnataka, India", International Journal of New Innovations in Engineering and Technology, Volume 17 Issue 3 August 2021, ISSN: 2319-6319.

[58] Chethan Chandra S Basavaraddi, **"**Using Machine Learning Techniques Studies on Water Quality Index and Phytoplankton Diversity of Tiptur Lake, Tiptur, Tumkur-District, Karnataka, India", Volume 10, Issue 9, September 2021 DOI 10.17148/IJARCCE.2021.10902, Certificate#:IJARCCE/2021/91, ISSN (Online) 2278–1021 ISSN (Print) 2319–5940.

[59] Chethan Chandra S Basavaraddi, "E-Health and Telemedicine in Today's World",
International Journal of Advanced Research in Computer and Communication Engineering Impact Factor 7.39□
Vol. 11, Issue 5, May 2022 DOI: 10.17148/IJARCCE.2022.11521, ISSN (O) 2278-1021, ISSN (P) 2319-5940.

[60] (2021-May-9th)http://en.wikipedia.org/wiki/Smartphone.

[61] (2021-May-9th)http://en.wikipedia.org/wiki/ICT.

[62] (2021-May-9th)http://en.wikipedia.org/wiki/Responsive_web_design.

[63] (2021-May-9th)http://en.wikipedia.org/wiki/EHealth.