ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

Clinical Decision Support System for Diagnosing Heart Disease

Suchithra[1], Dr.P.Uma Maheswari[2]

[1]PG Scholar, Info Institute of Engineering., India.

[2]Professor and Head, Department of CSE., Info Institute of Engineering., India.

ABSTRACT: Coronary artery disease (CAD) is the greatest risk for health. It causes the myocardial infarction or cardiogenic sudden death. Therefore, the diagnosis and prevention of the CAD are very worth to study. The purpose of this project is to develop a cost effective treatment using JADE for facilitating database J48 Decision Tree data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is a complex activity. In this project data mining technologies and agent framework has made an attempt to assist in the diagnosis of the disease in question.

KEYWORDS*:* Coronary Heart Disease, JADE ,Decision tree.

I. INTRODUCTION

The heart is a muscular organ, which receives blood from the coronary arteries. Coronary artery disease generally refers to the build-up of cholesterol in the inside layers of these arteries causing the blood flow to slow or stop. If the myocardium i.e., the fibrous muscle of the heart does not receive enough blood due to narrowing or blockage of any of these coronary arteries, a myocardial infarction (heart attack) can occur. Coronary artery disease, also called coronary heart disease, or simply, heart disease, is the No. 1 killer in America, affecting more than 13 million Americans. Heart disease is a result of plaque build up in your arteries, which blocks blood flow and heightens the risk for heart attack and stroke.

Heart disease is a result of plaque build up in your coronary arteries -- a condition called atherosclerosis -- that leads to blockages. The arteries, which start out smooth and elastic, become narrow and rigid, restricting blood flow to the heart. The heart becomes starved of oxygen and the vital nutrients it needs to pump properly.

From a young age, cholesterol-laden plaque can start to deposit in the blood vessel walls. As you get older, the plaque burden builds up, inflaming the blood vessel walls and raising the risk of blood clots and heart attack. The plaques release chemicals that promote the process of healing but make the inner walls of the blood vessel sticky. Then, other substances, such as inflammatory cells, lipoproteins, and calcium that travel in your bloodstream start sticking to the inside of the vessel walls.

Eventually, a narrowed coronary artery may develop new blood vessels that go around the blockage to get blood to the heart. However, during times of increased exertion or stress, the new arteries may not be able to supply enough oxygen-rich blood to the heart muscle. In some cases, a blood clot may totally block the blood supply to the heart muscle, causing heart attack. If a blood vessel to the brain is blocked, usually from a blood clot, an ischemic stroke can result. If a blood vessel within the brain bursts, most likely as a result of uncontrolled hypertension (high blood pressure), a hemorrhagic stroke can result.

ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

Cardiac ischemia occurs when plaque and fatty matter narrow the inside of an artery to a point where it cannot supply enough oxygen-rich blood to meet your heart's needs. Heart attack can occur - with or without chest pain and other symptoms.

Ischemia is most commonly experienced during:

- Exercise or exertion
- Eating
- Excitement or stress
- Exposure to cold

Coronary artery disease can progress to a point where ischemia occurs even at rest. And ischemia can occur without any warning signs in anyone with heart disease, although it is more common in people with diabetes.

The most common symptom of coronary artery disease is angina, or chest pain. Angina can be described as a heaviness, pressure, aching, burning, numbness, fullness, squeezing or painful feeling. It can be mistaken for indigestion or heartburn. Angina is usually felt in the chest, but may also be felt in the left shoulder, arms, neck, back, or jaw.

Other symptoms that can occur with coronary artery disease include:

- Shortness of breath
- Palpitations (irregular heart beats, skipped beats, or a "flip-flop" feeling in your chest)
- A faster heartbeat
- Weakness or dizziness
- Nausea
- Sweating

Artificial intelligence (AI)[1][2][10] is a branch of computer science that includes study and development of intelligent machines and software. Major AI researchers and textbooks define this field as "the study and design of intelligent agents", where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success.

The simplest AI applications can be divided into two types: classifiers ("if shiny then diamond") and controllers ("if shiny then pick up"). Controllers do however also classify conditions before inferring actions, and therefore classification forms a central part of many AI systems. Classifiers are functions that use pattern matching to determine a closest match. They can be tuned according to examples, making them very attractive for use in AI. These examples are known as observations or patterns. In supervised learning, each pattern belongs to a certain predefined class. A class can be seen as a decision that has to be made. All the observations combined with their class labels are known as a data set. When a new observation is received, that observation is classified based on previous experience.

JADE (Java Agent Development Framework) is a software Framework fully implemented in Java language. It simplifies the implementation of multi-agent systems through a middle-ware that complies with the FIPA specifications and through a set of graphical tools that supports the debugging and deployment phases. The agent platform can be distributed across machines (which not even need to share the same OS) and the configuration can be controlled via a remote GUI. The configuration can be even changed at run-time by moving agents from one machine to another one, as and when required. JADE is completely implemented in Java language and the minimal system requirement is the version 1.4 of JAVA (the run time environment or the JDK).

JADE is free software and is distributed by Telecom Italia, the copyright holder, in open source software under the terms of the LGPL (Lesser General Public License Version 2). Since May 2003, a JADE Board has been created that supervisions the management of the JADE Project. Currently the JADE Board lists 5members: Telecom Italia, Motorola, Whitestein Technologies AG, Profactor GmbH, and France Telecom R&D.

Java Agent Development Framework, or JADE, is a software for the development of agents, implemented in Java. JADE system supports coordination between several agents FIPA and provides a standard implementation of

ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

the communication language FIPA-ACL, which facilitates the communication between agents and allows the services detection of the system. JADE was originally developed by Telecom Italia and is distributed as free software.
JADE is a middleware which facilitates the development of multi-agent systems under the standard FIPA for which purpose it creates multiple containers for agents, each of them can run on one or more systems. Is understood that a set of containers constitutes a platform.
JADE provides:

- An environment where JADE agents are executed.
- Class Libraries to create agents using heritage and redefinition of behaviours.

A graphical toolkit to monitoring and managing the platform of Intelligent Agent agents.

A classifier can be trained in various ways; there are many statistical and machine learning approaches. The most widely used classifiers are the neural network, kernel methods such as the support vector machine, k-nearest neighbor algorithm, Gaussian mixture model, naive Bayes classifier, and decision tree. The performance of these classifiers have been compared over a wide range of tasks. Classifier performance depends greatly on the characteristics of the data to be classified. There is no single classifier that works best on all given problems; this is also referred to as the "no free lunch" theorem. Determining a suitable classifier for a given problem is still more an art than science.

II. LITERATURE SURVEY

CAD is caused by the accumulation of plaques within the walls of the coronary arteries that supply blood to the myocardium. CAD may lead to continued temporary oxygen deprivation that will result in the damage of myocardium. The presence of CAD is considered to exist when the narrowing of at least one of the coronary arteries is more than 50%. Coronary angiogram or cardiac catheterization is considered as “gold standard” method to diagnose the presence of CAD. This method has high accuracy but it is invasive, expensive and not possible as a diagnosis for large population. Many research works have been conducted to diagnose the CAD using less expensive and non-invasive methods such as electrocardiogram (ECG) based analysis, heart sound analysis, medical image analysis, etc [2-5].

Development of computer methods for the diagnosis of heart disease attracts many researchers. At the earlier time, the use of computer is to build knowledge based decision support system which uses knowledge from medical experts and transfers this knowledge into computer algorithms manually. This process is time consuming and really depends on medical expert’s opinion which may be subjective. To handle this problem, machine learning techniques have been developed to gain knowledge automatically from examples or raw data. Detrano, et al, built a new discriminant function model for estimating probabilities of angiographic coronary disease [6].

This discrimination function operates based on logistic regression which is not interpretable easily. Modeling of heart disease using Bayesian network (also called belief network) is proposed by Jayanta and Marco [7][8].
Gamberger, et al, proposed Inductive Learning by Logic Minimization (ILLM). The aim of using machine learning technique is also to find the important and useful information extracted from medical data [9].

Another work is proposed by Yan, et al, by using multi layer perceptron to build decision support system for the diagnosis of five major heart diseases [10].

Research work on Rough Set Theory (RST) to model prognostic power of cardiac tests has been proposed by Komorowski and Ohrn.

The work explores and identifies the need of a scientigraphic scan of a group of patients using rough set approach [11]. A research work on automated diagnosis on CAD based on rule induction and fuzzy modeling is proposed by Tsipouras, et al. The rule induction method that used to extract rules indirectly is C4.5 algorithm [12][13].

ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

A decision support system for the diagnosis of CAD is proposed in this paper. RST based technique is used to discover the knowledge from CAD data sets in the form of decision rules. A hybrid RST and support based rule selection is also proposed.

The data set is taken from Data Mining Repository of University of California, Irvine (UCI)[14]. Finally the system is validated using data sets from Cleveland, Hungarian, Long Beach, Switzerland and from Ipoh Specialist Hospital, Malaysia.

III. MATERIALS AND METHODS

Bayesian classifiers: Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". [19].

Function classifiers: Function classifiers include four classifiers namely multilayer perceptron, logistic regression, support vector machines and regression techniques. Neural Networks are used for both classification and prediction.

Lazy classifiers: Non-Parametric method used for classification and regression. Some Lazy Classifiers are: IBk[15], LWL, KSTAR, J 48, LBR[14] classifiers are used and J48[16] has demonstrated better accuracy for our data set.

Meta classifiers: Vote meta classifier is a classifier combining classifiers using unweighted average of probability estimates or numeric predictions. A method called ―Filtered Classifier‖ allows a filter to be paired up with a classifier. Classification can be made cost-sensitive, or multi-class, or ordinal-class. Parameter values can be selected using cross-validation.

Tree classifiers: Decision Tree Classifier is a simple and widely used classification technique. It applies a straightforward idea to solve the classification problem. Decision Tree Classifier poses a series of carefully crafted questions about the attributes of the test record. Each time time it receives an answer, a follow-up question is asked until a conclusion about the class label of the record is reached.

Rules classifiers: A total of 10 models are available having a dataset consisting in couples x and y, where x is each element of the population and y the class it belongs to, a classification rule can be considered as a function that assigns its class to each element.

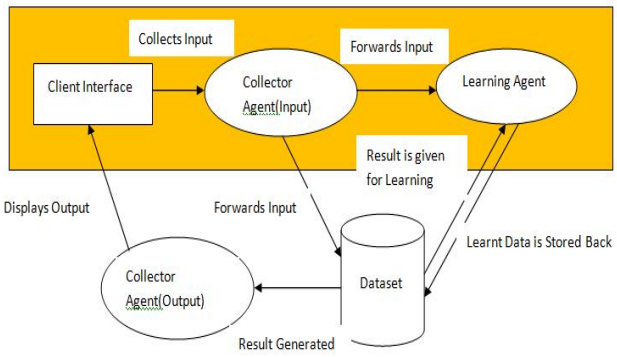


Fig 3.1: System and its Function

Agent Creation:

In this module agent such as Collector Agent and Learning Agent can be created. The input data can be sent to Collector agent. The collector agent passes the input data to Learning agent. Learning agent contains the training set of datas in which it predicts the type of heart disease .Finally the predicted result has been sent back to the collector agent .

ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

The Agent class is a super class which allows the users to create JADE agents. To create an agent one needs to inherit directly from *Agent*. Normally, each agent recorder several services which they should be implemented by one or more behaviours.

This class provides methods to perform the basic tasks of the agents as:

- Pass messages by objects **ACLMessage**, with pattern matching.
- Support the life cycle of an agent.
- Plan and execute multiple activities at the same time.

The cycle of life of a JADE agent follows the cycle proposed by FIPA. These agents go through different states defined as:

1. Initiated: The agent has been created but has not registered yet the **AMS**.
2. Active: The agent has been registered and has a name. In this state it can communicate with other agents.
3. Suspended: The agent is stopped because its thread is suspended.
4. Waiting: The agent is blocked waiting for an event.

Decision tree Induction:

The heart of the input data can be predicted by using the J48 decision algorithm based on the training set in Learning Agent. The prediction of disease can be done by the J48 decision algorithm as follows:

- J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool. C4.5 is a program that creates a decision tree based on a set of labeled input data. This algorithm was developed by Ross Quinlan. The decision trees generated byC4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

1. A tree structure internal node denotes a test on an attribute. Branch represents an outcome of the test. Leaf nodes represent class labels or class distribution

2. Decision tree generation consists of two phases :

Tree construction:

At start, all the training examples is at the root. Partition examples recursively based on selected attributes.

Tree pruning

Identify and remove branches that reflect noise or outliers

3. Use of decision tree: Classifying an unknown sample and Test the attribute values of the sample against the decision tree.

IV. PROPOSED SYSTEM

In proposed system, Agent based framework is used for predicting the coronary heart disease. This proposed system uses the dataset gathered from domain expert including the symptoms, stages and treatment. The proposed system facilities that provides an efficient and easy method to diagnose if a patient is affected by Heart disease, If so in which stage and the treatment options for the prognosis of heart disease. The proposed system is implemented in JADE, and Decision from the agent can be done by using J48 Decision Tree data mining technologies. JADE platform contains two agents such as Collector agent and Learning Agent. Learning Agent can be comprised on training set whereas Collector agent uses inputdata in turn the disease can be predicted based on J48 Decision tree algorithm.

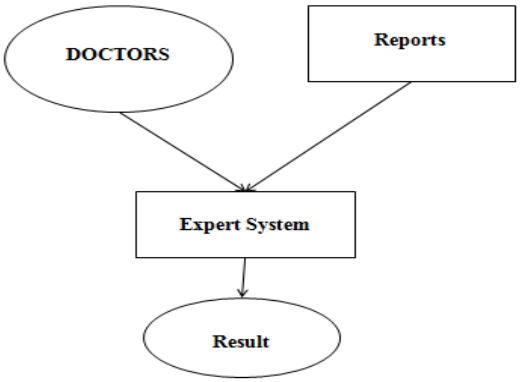
ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

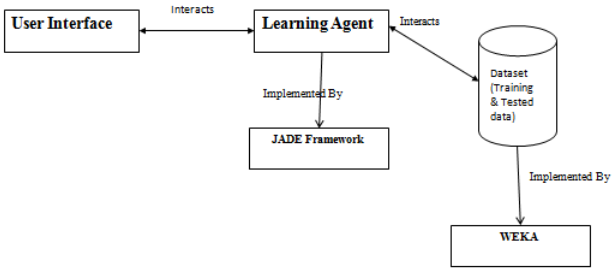
(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

The Overall Architecture of the proposed system is as follows:



The processing view of the system is:



ADVANTAGES

- Prediction performance of coronary heart disease is obtained
- Less complexity is obtained.

V. RESULTS AND DISCUSSIONS

Specificity

Specificity measures the proportion of negatives which are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition). These two measures are closely related to the concepts of type I and type II errors.

Mathematically, this can also be written as:

ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

$$= \frac{\text{number of true negatives}}{\text{total number of well individuals in population}}$$

$$= \text{probability of a negative test given that the patient is well}$$

Sensitivity

Sensitivity measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$= \frac{\text{number of true positives}}{\text{total number of sick individuals in population}}$$

$$= \text{probability of a positive test, given that the patient is ill}$$

Accuracy

The **accuracy** of a measurement system is the degree of closeness of measurements of a quantity to that quantity's actual (true) value. It is a parameter of the test.

$$\text{accuracy} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}}$$

Precision

Precision value is calculated is based on the retrieval of information at true positive prediction, false positive. In healthcare data precision is calculated the percentage of positive results returned that are relevant.

Precision =TP/ (TP+FP)

TP-True positive

FP-true negative

Recall

Recall value is calculated is based on the retrieval of information at true positive prediction, false negative. In healthcare data precision is calculated the percentage of positive results returned that are Recall in this context is also referred to as the True Positive Rate. Recall is the fraction of relevant instances that are retrieved,

Recall =TP/(TP+FN)

FN – false negative

F-measure

The F-Measure computes some average of the information retrieval precision and recall metrics**.**

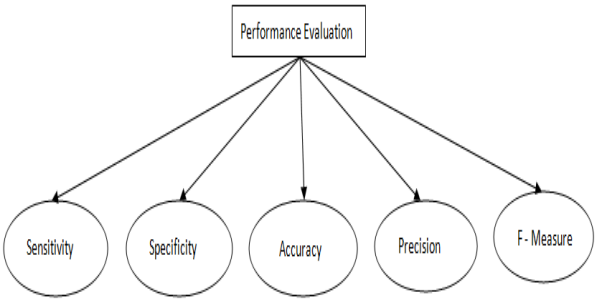
ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

F-Measure= $\frac{2*Recall*Precision}{Precision+Recall}$



VI. CONCLUSIONS

In the present work expert system is created for predicting the coronary heart disease. The proposed system uses JADE tool for creating agent. Two agents such as Collector and Learning agent have been created. The input data can be sent to collector agent and in which it passes the data to learning agent to predict the disease. Learning agent predict disease of input using J48 Decision tree algorithm and the result can be send back to Collector agent. Thus the proposed system provides an efficient prediction of heart disease when compare with existing researches.

REFERENCES

[1] "Coronary heart disease statistics fact sheets 2008/2009," British Heart Foundation, London 2008.
[2] B. L. Zaret, M. Moser, and L. S. Cohen, *Yale university school of medicine heart book*. New York: Hearst Books, 1992.
[3] B. Phibbs, *The human heart: a basic guide to heart disease*. Philadelphia: Lippincott Williams & Wilkins, 2007.
[4] A. Selzer, *Understanding heart disease*. Berkeley: University of California Press, 1992.
[5] O. S. Randall and D. S. Romaine, *The encyclopedia of the heart and heart disease*. New York, NY: Facts on File, 2005.
[6] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American Journal of Cardiology*, vol. 64, pp. 304-310, 1989.
[7] K. G. Jayanta and V. Marco, "Building a Bayesian network model of heart disease," in *Proceedings of the 38th annual on Southeast regional conference*. Clemson, South Carolina: ACM, 2000.
[8] K. G. Jayanta and V. Marco, "Probabilistic model building of heart disease," Department of Computer Science, University of South Carolina 1999.
[9] D. Gamberger, G. Krstačić, and T. Šmuc, "Medical Expert Evaluation of Machine Learning Results for a Coronary Heart Disease Database," in *Medical Data Analysis*, 2000, pp. 119.
[10] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron-based medical decision support system for heart disease diagnosis," *Expert Systems with Applications*, vol. 30, pp. 272, 2006.
[11] J. Komorowski and A. Ohrn, "Modelling prognostic power of cardiac tests using rough sets," *Artificial Intelligence in Medicine*, vol. 15, pp. 167, 1999.
[12] M. G. Tsipouras, T. P. Exarchos, D. I. Fotiadis, A. Kotsia, A. Naka, and L. K. Michalis, "A decision support system for the diagnosis of coronary artery disease," presented at 19th IEEE Symposium on Computer-Based Medical Systems, 2006.
[13] M. G. Tsipouras, T. P. Exarchos, D. I. Fotiadis, A. P. Kotsia, K. V. Vakalis, K. K. Naka, and L. K. Michalis, "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 12, pp. 447, 2008.
[14] D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz, "UCI Repository of machine learning databases," University California Irvine, Department of Information and Computer Science, 1998.
[15] Dr.Vaidyanathan and K.Rajeswari, - Artificial Intelligence techniques applied to the development of a clinical decision support system for diagnosing Ischemic Heart Disease., International Journal of Medical Informatics Vol.70.,2012.
[16] KS Reddy and committee - International Heart Protection Summit September 2011.
[17] Dr.P.Amirtaraj and K.Rajeswari - Classification of Risk Level for Ischemic Heart Disease in India using Artificial Intelligence., Artificial Intelligence in Medicine., Vol.56.,2011.
[18] Dr.P.Amirtharaj and Dr.Vaidyanathan - Prediction of Risk Score for Heart Disease in India using Machine Intelligence.,. ICMLC 3rd International Conference on MachineLearning and Computing vol7 2011.
[19] Cardiovascular Diseases & its Impact – WHO.,2010.

ISSN(Online):2320-9801
ISSN(Print): 2320- 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Special Issue 3, July 2014

[20] By Rajeev Gupta & his associates - Burden of Cardiovascular Diseases in India.,2007.
[21] T. Gurumoorthy and K.Rajeswari Modeling Effective Diagnosis of Risk Complications in Type 2 Diabetes – A Predictive model for Indian Situation., Indian J Med Res 132.,2012
[22] Dr.V.Vaithiyanathan and Dr.P.Amirtharaj - A Novel Risk Level Classification of Ischemic Heart Disease using Artificial Neural Network Technique – An Indian Case Study., Artificial Intelligence in Medicine Vol.50.,2012.
[23] K. Rajeswari& V. Vaithiyanathan - Heart disease diagnosis: an efficient decision support system based on fuzzy logic and genetic algorithm., Application to Cardiovascular Diseases_, IEEE Transactions on BioMedical Engineering, Vol. 54.,2010.
[24] Preethi S.J. and K.Rajeswari - Image Enhancement Techniques for Improving the Quality of Colour and Gray scale Medical Images., Artificial Intelligence in Medicine Vol.50.,2009.