

Heart Disease and Data Mining: Collaboration in the Making

Neha Chauhan

*Student, Computer Science and Engineering Department,
AP Goyal Shimla University, India*

Nisha Gautam

*Asistant Professor, Computer Science and Engineering Department,
AP Goyal Shimla University, India*

Abstract: The healthcare sector assemble enormous quantity of healthcare data which cannot extract to uncover secret information for making effectual decision. Although, these data have a plenty of unrevealed information which is untapped and not being used appropriately for predictions. In case of Heart Disease, it becomes more influential which is considered as the predominant reason behind death all around the world. Detection of several disease of heart is an expressive and vapid task in medicine. Heart Disease defines various healthcare conditions that are highly voluminous in nature which are related to heart and has many root causes that effect the entire body. In medical field, Data Mining provide various methods that are useful in resoluting the clinical support systems for predicting and detecting the several kind of diseases. This paper discussed the overall review of predicting heart disease by using various techniques of data mining. These techniques used in heart diseases takes less time and make process fast for the prediction system to predict heart diseases with good validity in order to improve their health.

Keywords: Heart Disease, Data Mining, K-Mean Clustering, SVM, KNN, Artificial Neural Network.

I. INTRODUCTION

In our busy lifestyle, people toil as a machine in order to live a very luxurious life and to earn a lot of money but in this type of life style people forget to take care and even don't give enough rest to themselves. Owing to this, people suffer from multifarious diseases at an extremely little one age which affects our health like: Diabetes, Cancer, Heart diseases and many more. Nowday's, Heart disease (HD) is a major health problem and cause of deaths all over the world. The term Heart Disease defining a numerous health conditions and holds various diseases which affects the heart [1]. Heart is a very significant part of the human body as the entire life is rely on efficient working of heart. At present, the numerous category of heart diseases exist but most well known category of disease is coronary artery disease which causes owing to the narrow or blockage of coronary arteries because of plague that buildup inner side of the arteries which lessen the flow of blood to heart [2]. Plague is made up of calcium, fat, cholesterol and other substances that found in the blood [8]. This condition lead to increases the risk of angina or a heart attack. Heart Attack disease is a serious medical emerging and prevail the chief reason of morbidity all around the world and will prevent the attacks when possible detection at a prior stage [3]. Heart diseases are the primary root of death globally, and in numerous developing countries most of the population die every year suffered from HD than from any other root [5]. World Health Organization reported that every year, an estimated 12 million deaths happens due to the Heart Disease [9]. In the United State, HD kill one person every 32 second and it affects the whole circulatory system which consist of heart and blood vessels [[HYPERLINK \l "VMa13" 1](#)]. In the sector of medicine, current research has been accomplished to determine the reasons which may contribute concerning the expansion of heart disease but in order to use this knowledge in an unconventional way more research is needed to reduce the occurrence of HD [5]. The leading problem of the heart disease detection from diverse symptoms which is not free from faulty presupposition and most of the time clinical decisions which are based upon clinicians suspicion and practice rather than on the knowledge hidden data in the database. This experience results causes disastrous results, undesirable delusion and immoderate medical prices which are therefore inappropriate that influence their quality which has given to patients and the risk of death increases [7]. In order to overcome the numerous deaths due to the cause of heart diseases there have to be an efficient detection technique that could provide useful knowledge by searching the relationships and patterns within the medical data for efficient medical diagnosis [6]. The pateints of HD have several risk factors in familiar which can be used very efficiently for diagnosis [5]. There are some of the major level of risk in HD which incorporates Diabetes, High Cholesterol, Obesity,

Smoking, Eating habits, Alcohol intake, Physical inactivity etc that affects our heart or entire body. In medical field, Data Mining provides various methods for the transformation of data into useful information for the decision making and helps to discovery of relationships in patients data in order to improve their health. Researchers have been applied various techniques of Data Mining such as Association rule mining, Clustering, Classification algorithms like decision tree, neural network, naive bayes, kernel density, support vector machine etc for prediction and recognition of HD [5]. By employing some of the techniques of Data Mining, Heart Disease prediction can be made easy with respect to various attributes to find out whether the person is suffering from HD or not and also takes less time for the prediction and improve the medical recognition of the disease with good precision and reducing the occurrence of heart disease [4].

The purpose of this paper to provision a survey employing diverse methods of Data Mining for HD prediction. This paper discussed and compared these methods for predicting HD w.r.t precision and by employing this study, this will be helpful for the researchers in predicting and analyzing medical disease.

The following Section of this paper is described as, a survey of literature employed in predicting heart disease is discussed in Section II, the brief study of data mining is presented in Section III, Section IV described diverse methods of data mining and Section V concludes that survey.

II. RELATED WORK

M.A.Nishara Banu et al [15] proposed a system of Disease Forecasting by employing methods of Data Mining. The major objective of this analysis to designed a predictive system that is enable of predicting the heart attack victoriously. The researchers used such algorithms of Data Mining namely, K-Means clustering, MAFIA and C4.5 algorithms for predicting and analyzing the heart diseases from the heart disease database that is simulated in MATLAB tool. An experiment accomplished by researcher on the dataset to identify patterns for predicting HD. Firstly, this database is pre-processed to make the process efficient and results composed by K-Means algorithm and then frequent patterns are extracted efficiently to the HD by using MAFIA. This research compared the prediction accuracy between simple MAFIA and proposed K-means based MAFIA and the outcome appeared that the K-means based MAFIA with C4.5 improved the accuracy of the pateint.

H.W. Khaing [11] proposed a method which is effectual for fragmenting and extracting meaningful patterns from the dataset for the prediction of risk level of heart attack. The heart disease dataset consist of assorted features which hold both numerical and categorical data. By utilizing a variety of data mining techniques in the sector of medical, researcher diagnose the efficacious care for patient. In this, they used K- Means Clustering algorithm to make cluster of given database which extract from the dataset. This method permitted the number of fragments and frequent patterns from the extracted data are assorted that applicable to HD by using MAFIA algorithm. To display the effectual risk level of heart attack utilized ID3 algorithm with the decision tree and outcome appeared that the prediction system proficient with good precision of effectively predict the heart attack.

S.U.Amin et al [5] proposed a method for HD prediction utilizing several crucial risk elements. This method employed two tools which are of data mining namely neural network and genetic algorithms and twelve important risk factors. This hybrid system used the advantages of genetic algorithm for initialization of the weights of neural network so as ANN improved their performance and predict the disease at very eary stage. This system was simulated employing MATLAB which gave the good accuracy to predict the risk of HD.

S. Palaniappan et al [7] proposed an Intelligent Heart Disease System for prediction using some techniques. This research designed an prediction system with the help of various DM methods namely, Decision tree, Neural Network and Naïve Bayes to uncover and extract hidden trends and relationships from HD dataset which correspond to heart disease. Researcher used Lift Chart and matrix classification methods to mined the patterns and outcomes showed that Naïve Bayes appeared most effective model followed by Neural Network and Decission tree to predict the patients of HD. To ease of model interpretation and accuracy, these models could explain multiplex queries accompanied by its own intensity. This research utilized five mining goals and by using these three models, the result showed that Naïve Bayes answered four goals out of five and remaining models had two and appeared the most effective model as to identify all the significant predictors.

M.Anbarasi et al [19] proposed an enhanced prediction of HD with feature subset selection employing Genetic Algorithm. They utilized three classifiers namely, Naïve Bayes, Classification through Clustering and Decision Tree. This paper's objective is to predict the HD presence more accurately with less number of test attributes and this is done with the help of Genetic Algorithm. Basically, 13 attributes were utilized to predict the HD and with the help of Genetic Algorithm, they reduced the attributes from 13 to 6 attributes. This paper concluded that, among these three classifiers, Decision Tree accuracy outperforms better and after applying the FSS with high construction time and in the case of Naïve Bayes before and after minimization of features, it performs the same with same construction time and Classification through Clustering performs poor as differentiated to others.

H.D.Masetha et al [3] proposed a system of HD prediction employing Classification algorithms to predict the heart attack which is simulated on Weka tool. They utilized three algorithms such as J48, Naïve Bayes, RETTREE, CART and Bayes Net on the given parameters to check the presence of HD. The purpose of this paper is to predict the performance of the classifier and estimation of comparison to check which classifier is best. This concluded that J48, Simple CART and RETTREE performs same and proved best and from the others two Bayes Net outperforms as compared to naïve Bayes algorithm.

III. DATA MINING AND THEIR METHODS

Data Mining is the detection of big amount of data and incorporate statistical examination, artificial intelligence and database technology to obtain unrevealed knowledge and formerly hidden trends and relationships which is not incorporate with traditional statistics and are inconvenient to observe [10]. It involves theories, methodologies and computer system for knowledge extraction and search for the relationships and overall patterns that exist in large volume of data repositories [11]. The chief purpose of the procedure of Data mining is to identify viable, new, possibly useful and eventually comprehensible trends in data with the vast use of databases and the volatile progress in their size [12]. Data mining also mentioned to as Knowledge Discovery in Database, but both Data mining and KDD are distinct with each other. The term KDD concern to uncover the convenient knowledge from big amount of databases. On the other side, Data mining concern to the approach of algorithm for mining patterns from data without involving the steps of KDD process [13]. This procedure involves these steps which include Data Selection, Processing, Transformation, Mining, Pattern Evaluation and Representation of Knowledge. Data mining provide different technology that transforms numerous data into convenient results and finding the latest and indefinite patterns in the data for decision making. Data Mining defines two tasks such as Predictive tasks and Descriptive tasks. Predictive Data mining tasks includes prediction of unknown and future values by performing conclusion on the existing data which include Classification, Regression, Time Series analysis, Prediction. On the other side, Descriptive Data mining tasks involves the specific properties of the data in database and locating the sequence and behavior in data which include Clustering, Association Rules, Sequence Discovery [13],[14]. These techniques of Data mining helps to identify the knowledge in the database and obtaining these knowledge like so can be used in these fields like business, education, healthcare sector, scientific, etc [12]. Among all these areas, Data mining plays a pre-eminent part in Healthcare sector for the prediction of Heart disease. Most of the researchers are using various techniques of Data mining for the identification of many diseases for instance: stroke, cancer and heart disease. The data are too complex and voluminous that is generated for prediction of heart disease in order to processed by traditional methods. To get useful information, healthcare industry encompasses the huge amount of data which has to be refined [8]. Orderly uncover hidden trends, these methods of data mining explore, examine and abstract medical data by using multiplex algorithms [5] and these discovered knowledge can assist the healthcare administrators to enhance the operations and quality of levels which is provided to patients.

IV. METHODS OF DATA MINING

Various methods of Data mining that are used in extracting the useful knowledge from the databases as Classification, Regression, Clustering, and Association Rule mining and by employing these techniques, mined the medical data for the identification of disease. Diverse mining algorithms for medical data utilized by most of the researchers to predict the heart disease, some of these algorithms that are commonly used in medical field are explained below:

A. K-means clustering: The k-means is the one of the manageable unsupervised learning algorithm and a repetition procedure to separation a particular set of data into predefined number of k clusters [17], where k is

marked as an input variable to resolve the conventional clustering problem. The k-means algorithm is a technique and frequently employed in medical visualizing and their associated fields [20]. This technique selects points in multidimensional space to symbolize each k cluster called centroids. A centroid is the point whose ordinates are acquired by evaluating the average of each co-ordinates of samples point that allocated the clusters [16]. The chief target of using this kind of clustering is to emphasize the overall intra-cluster deviation or the squared error function.

$$k = \sum_{k=1}^m \sum_{j=1}^{n_k} \|d_{jk} - c_k\|^2$$

Where,

k, the objection function i.e. the overall expand of data points and should be small as feasible,

m, the number of clusters,

n, the number of classes,

d_{jk} , the k^{th} point in the j^{th} cluster and c_k is the set of point that belongs to k^{th} cluster and

$\|d_{jk} - c_k\|$, defines the distance between d_{jk} and c_k .

B. Decision Tree: This is one of the most familiar supervised learning method and applied for classification which shows their structure like a tree in a flow- chart manner, where every single node on a predictive test represents a test and every single branch symbolizes an attribute value and their leaf nodes represents class distribution or predicted classes [21]. Their structure provides “if-then” rules that makes the results easy for explanation or understanding and can be used visually and unambiguously to express decisions. This is a predictive model that split-up the input gape into cells where everyone belongs to one class and splitting or separating represents as a series of tests [14]. This technique is commonly used in various operation researches, particularly in making decisions and to identify an approach to accomplish a goal. The numerous algorithms of Decision Tree exist which include: Iterative Dichotomises 3, C4.5 (successor of ID3), Classification and Regression Tree, Logitboost Alternating Decision Tree, CHI-squared Automatic Interaction Detector and many more.

C. MAFIA (Maximal Frequent Pattern Algorithm): Mining of recurrent item set is most fundamental Data Mining problems that aims to discover group of items which coincide with frequently in a dataset [11]. The main aim is to find out the fascinating patterns from data warehouse in a diversity of data mining methods such as clusters, classifiers, association rules, sequences and many more. The novel approach MAFIA (Maximal Frequent Pattern Algorithm) exploit an effective algorithm which combines the ideas of old and new algorithms to configure a realistic algorithm [11]. This is used to extract maximum frequent item set for searching with effectual diminishing mechanism [15].

D. K-Nearest Neighbour: KNN is most simplest technique or method which performs only classification to classify the objects when the tests attributes objects match absolutely one of the training examples and it also retain the entire training data [17]. The predominant objective of using KNN classification is to find the set of k objects present in the training set that are nearest to the test objects and assign a label on a specific class in their neighbourhood [17]. KNN is a supervised learning algorithm and very effectual deduction method for multiplex target functions and noisy training data. To resolve the closeness of occurrences, the measurement of distance is required. It classifies an occurrence to choose the class among the neighbours by finding its nearest neighbour [22]. Closeness is express by using Euclidean distance to compute the distance between two points i.e.

$$R(r_1, r_2, \dots, r_n)$$

$$S(s_1, s_2, \dots, s_n)$$

And the following equation is defined as,

$$p(R, S) = \sum_{j=1}^n (R_j - S_j)^2$$

This KNN algorithm works based on the minimum distance from the new data to the K- Nearest Neighbour [23].

E. Naïve Bayes: Naïve Bayes is a classification technique and especially preferred when the higher input dimensionality is present. This technique is utilized to generate models with prognostic abilities and premises for several methods of machine learning [12]. This method provides an innovative way of investigating and perceiving the data. The term Naïve Bayes classifiers concerns with the probabilistic classification which is very simple and based on Bayes Theorem [18].

$$P(Y/X) = \frac{P(X/Y)}{P(X)}$$

This theorem calculates the probability of Y given X, where Y represents the dependent event and X represents the initial event, this algorithm computes the cases where X and Y occur simultaneously and divides it by the number of cases where only X occur.

The Naïve Bayes classifiers presume a class conditional independence which appears that the values effect of variable on a specific class which is individualistic of the values of alternative variable [24]. Naïve Bayes Classifier accomplish well in multiplex real world situations. In healthcare sector, Naïve Bayes deduce the attributes of patients of heart disease and display every input attribute probably for the predictable state [12].

F. Support Vector Machine: SVM is an artificial learning concept which is used to examine the data and identify the patterns in classification and regression analysis and also it has multilayer perceptrons and spiral basis for networks. Support Vector machine used for both type of classification data linear and non-linear and converts the original data into higher dimension using non-linear mapping. SVM maps the training vectors with higher spatial dimensionality and designates every vector through its class. It produces a model utilized for predicting the objective values of the test data which is based on given training data. The main purpose of this method is to locate the best hyper plane by disjoining the data from two classes as much as feasible with a suitable non-linear mapping. Support Vector Machine requires the solution to perform classification of the optimization problem [26].

$$\min_{a,c,b} \frac{1}{2} M^K M + P \sum_{j=1}^{\epsilon} \epsilon_j$$

Subject to,

$$x_j (M^K \varphi(y_j) + C) \geq 1 - \epsilon_j$$

Where, $\epsilon_j \geq 0$, a slack variable which enable the delusion in the classification,

y_j is the training vectors, $y_i \in R^n$

φ is a function mapping y_j into higher dimension space,

P is a penalty parameter of the error term ($C > 0$),

x_j , a class label, $x_j \in \{1, -1\}^1$

Support Vector Machine is the method of an accomplishment of structural uncertainty minimization and has appeared good extensive performance on the problems of pattern classification and in various number of application areas [25].

G. Artificial Neural Network: In machine learning, an Artificial Neural Network or Neural Network is a rigorous model to make predictions and reveal patterns. NN is a class of artificial neurons which are highly interconnected with each other and operates information to computation by using the approach of connectionist and this process is inspired through biotic neural networks [8]. ANN have become a very influential tool in various tasks such as decision problem, data classification, pattern recognition or prediction application through learning process and a novel signal processing technology. ANN can be supervised or unsupervised learning and used in a variety of applications such as fraud detection, pattern recognition, customer response prediction and many more [14]. ANN modifies its structure because of its versatile system throughout the learning process and connecting of inputs and outputs of data it uses multiplex relationships. ANN defines parameters i.e. i) inter-connection sequence between divergent layers of neurons, ii) procedure of learning for modernizing the weights of the interconnections, iii) an activation function which transforms the weight of a neurons input to its output activation [8]. A typical relation of input and output can be expressed as,

$$net_k = \sum_{i=1}^n w_{jk} y_i + b_k$$

where y_j is the input to j^{th} node in input, w_{jk} defines the weight between j^{th} input node and k^{th} hidden node, b stands for bias at k^{th} node, net defines the adder and a is th activation function.

COMPARE DIVERSE METHODS IN TERMS OF THEIR PRECISION FOR HEART DISEASE PREDICTION

Contribution	No. of Attributes	Tool used	Accuracy
Using K-means, MAFIA and C4.5 designed Disease Prediction System [15]	13	MATLAB	89%
Using MAFIA, ID3 and K-means designed an approach for fragmentation and Prediction of Medical Data [11]	13	WEKA	85%
Using Neural Networks and Genetic Algorithms designed a Prediction of Heart Disease [5]	12	MATLAB	89%
Using Decision tree, Naïve Bayes and Neural Network designed an Intelligent Heart Disease Prediction System [7]	15	.NET	86.53%
Using Feature Subset Selection, Genetic Algorithm, Decision tree, Naïve Bayes and Classification via Clustering [19]	13	WEKA	99.2%
Using J48, Naïve Bayes, REPTREE, CART and Bayes Net designed Prediction of Heart Disease [3]	11	WEKA	99%

Table I. Comparison table of heart disease prediction using data mining techniques.

V. CONCLUSION

In healthcare environment, awareness is the most informative strength of any organization. Healthcare environment is rich of information and still a severe challenge of generating knowledge out of this information. Healthcare data are voluminous in nature which collect from various resources but their quality or formation not absolutely appropriate. This paper presented a survey by employing several techniques of Data Mining for predicting heart disease. Data Mining plays an influential task in predicting HD when analogized with different methods such as clustering, classification, association rule and regression. In this paper, we discussed a

comparative chart of heart disease using diverse techniques in terms of their attributes and their precision. This survey has been used various techniques for HD and every technique has its own strength and limitations. By employing this study, we concluded that Feature Selection technique has been worked well and gave the highest accuracy among of these techniques for predicting HD. This comparative study will make things easy for the various learners to find out the best techniques for their research according to their tools. Heart Disease prediction with Data Mining techniques will became most victorious with less number of attributes and enhance the excellence of clinical decisions.

REFERENCES

- [1] V. Manikantan & S. Latha, "Predicting the analysis of Heart Disease Symptoms using Medicinal Data Mining Methods", *International Journal of Advances Computer Theory and Engineering*, Vol. No.2, Issue No. 2, pp. 5-10, 2013.
- [2] Divya Kundra, Navpreet kaur, "Review on Prediction System for Heart Diagnosis Using Data Mining Techniques", *International Journal of Latest research in Engineering and Technology (IJLRET)*, Vol. No.1, Issue No. 5, pp. 09- 14, 2015.
- [3] Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using Classification Algorithms", *World Congress on Engineering and Computer Science*, Vol. No.2, 2014.
- [4] Beant Kaur, Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. No.2, Issue No.10, 2014.
- [5] Syed Umar Amin, Kavita Agarwal, Rizwam Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", *IEEE Conference on Information and Communication Technologies (ICT)*, pp. 1227-1231, 2013.
- [6] M.Akhil jabbar, Priti Chandra, B.L Deekshatulu, "Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection", *IEEE International Conference on Intelligent Systems Design and Applications*, pp. 628-634, 2012.
- [7] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", *IEEE* (2008).
- [8] Nitin Kumari, Sunita, Smita, "Comparison of ANNs, Fuzzy Logic and Neuro-Fuzzy Integrated Approach for Diagnosis of Coronary Heart Disease: A Survey", *International Journal of Computer Science and Mobile Computing*, Vol. No.2, Issue No.6, pp. 216-224, 2013.
- [9] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", *International Journal of Computer Applications*, Vol. No.17, Issue No.8, pp. 43-47, 2011.
- [10] Sivagowry.S, Durairaj.M, Persia.A, "An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart Disease", *IEEE*.
- [11] Hnin Wint Khaing, "Data Mining based Fragmentation and Prediction of Medical Data", *IEEE* (2011).
- [12] G. Subbalakshmi, K. Ramesh, M. Chinna Rao, "Decision Support in Heart Disease Prediction System using Naïve Bayes", *Indian Journal of Computer Science and Engineering (IJCSSE)*, Vol. No.2, Issue No.2, pp. 170- 176, 2011.
- [13] Neha Chauhan, Nisha Gautam, "Parametric Comparison of Data Mining Tools", *International Journal of Advanced Technology in Engineering and Science*, Vol. No.3, Issue No.11, pp. 291-298, 2015.
- [14] T. Silwattananusam, K. Tuamsuk, "Data mining and its application for Knowledge management: A literature review from 2007 to 2012", *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, Vol. No.2, Issue No.5, 2012.
- [15] M.A. Nishara Banu, B. Gomathy, "Disease Forecasting System Using Data Mining Methods", *IEEE*, pp.130- 133, 2014.
- [16] Soumi Ghosh, Sanjay Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms", *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. No.4, Issue No.4, pp.35- 39, 2013.
- [17] XindongWu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg, "Top 10 Algorithms in Data Mining", © Springer-Verlag London Limited 2007.
- [18] Shadab Adam Pattekari and Asma Parveen, "Prediction System for Heart Disease Using naïve Bayes", *International Journal of Advanced Computer and Mathematical Sciences*, Vol. No.3, Issue No.3, pp. 290- 294, 2012.
- [19] M. Anbarasi, E. Anupriya, N.CH.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", *International Journal of Engineering Science and Technology*, Vol. No.2, Issue No.10, pp. 5370-5376, 2010.
- [20] G.Komarasamy and Amitabh Wahi, "Improving the Cluster Performance by Combining PSO And K-Means Algorithm", pp. 206- 208.
- [21] Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal, Ashish Kumar Sen, "A Heart Disease Prediction Model using Decision Tree", *IOSR Journal of Computer Engineering (IOSR-JCE)*, Vol. No.12, Issue No.6, pp. 83-86, 2013.
- [22] Nurhayati, Arif Nur Rahman, "Implementation of Naive Bayes and K-Nearest Neighbour Algorithm for Diagnosis of Diabetes Mellitus", pp. 117- 120.
- [23] Krati Saxena, Zubair Khan, Shefali Singh, "Diagnosis of Diabetes Mellitus using K Nearest Neighbour Algorithm", *International Journal of Computer Science Trends and Technology (IJCSST)*, Vol. No.2, Issue No.4, pp. 36-43, 2014.
- [24] Manjusha K. K, K. Sankaranarayanan, Seena P, "Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. No.4, Issue No.1, 2014.
- [25] Gur Emre Guraksin And Harum Ugaz, "Classification Of Heart Sounds Based On The Least Squares Support Vector Machine", *International Journal of Innovative Computing, Information and Control ICIC International*, Vol. No.7, Issue No. 12, pp. 7131-7144, 2011.
- [26] G.Parthiban, S.K.Srivatsa, "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients", *International Journal of Applied Information Systems (IJAIS)*, Vol. No.3, Issue No.7, pp. 25- 30, 2012.