International Journal of Energy, Information and Communications
Vol.6, Issue 4 (2015), pp.17-34
http://dx.doi.org/10.14257/ijeic.2015.6.4.02

Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System

Shubpreet Kaur[1] and Dr. R.K.Bawa[2]

[1]*Punjabi University, Patiala, Punjab*
[2]*Punjabi University,Patiala, Punjab*
shubpreetkaur@gmail.com, rajesh.k.bawa@gmail.com

Abstract

The thriving medical applications of data mining in the fields of medicine and public health has led to the popularity of its use in knowledge discovery in databases (KDD). Data mining has revealed novel biomedical and healthcare acquaintances for clinical decision making that has great potential to improve the treatment quality of hospitals and increase the survival rate of patients. Disease diagnosis is one of the applications where data mining tools are establishing the successful results. Data mining intends to endow with a systematic survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research. Discussion is made to enable the disease diagnosis and the breakthrough of hidden healthcare patterns from related databases is offered. Also, the use of data mining to discover such relationships as those between health conditions and a disease is presented. It further discusses about the tools that can be used for the processing and classification of data. This paper summarizes various technical articles on medical diagnosis and prognosis. It has also been focused on current research being carried out using the data mining techniques to enhance the disease(s) forecasting process. This research paper provides future trends of current techniques of KDD, using data mining tools for healthcare. It also confers significant issues and challenges associated with data mining and healthcare in general. The research found a growing number of data mining applications, including analysis of health care centers for better health policy-making, detection of disease outbreaks and preventable hospital deaths. The root causes of all diseases get closer towards drugs i.e. the foremost risk factor of all hilarious diseases. Drug addiction using WEKA has been used that brings into light concerning majority of drug abusers started abusing drugs at age below 20yrs.It is to make aware the druggist about the various diseases that are caused with heavy or long term intake of drugs in their life. So, to make an expert system that will awake the youth about precarious use of drugs and also alert the affected person.

Keywords: *Data Mining Techniques, Data Mining Tools, Data Mining Models, Healthcare, Medical Diseases, Drug Abuse.*

1. Introduction and Motivation

Data mining is the extraction of hidden predictive information and unknown data, patterns, relationships and knowledge by exploring the large data sets which are difficult to find and detect with traditional statistical methods. Data mining it is powerful technology which will discover most important information from the data warehouse of the organizations. It is a very crucial step that collectively examine large amount of routinely data. To find latest patterns in healthcare industry, there exist various interactive and scalable data mining methods. Data mining is a quantitative approach which is user friendly in reading reports and reducing errors and controls the quality more uniformly. Important task of data mining is data pre-processing.

ISSN: 2093-9655 IJEIC

Data mining tools are used for decision making. Prediction and classification techniques are used in which classification technique predicts the unknown values with respect to generated model. An assortment of data mining techniques can be applied to find associations and regularities in data, extract knowledge in the forms of rules and predict the value of the dependent variables. Common data mining techniques which are used in almost all the sectors are listed as: Naive Bayes, Decision Tree, Artificial neural network (ANN), Bagging algorithm, K- nearest neighborhood (KNN), Support vector machine (SVM) etc. Data mining is an important step of knowledge discovery in databases (KDD) which is an iterative process of data cleaning, integration of data, data selection, pattern recognition and data mining knowledge recognition. KDD and data mining are also used interchangeably. Data mining encompasses association, classification, clustering, statistical analysis and prediction.

Data mining has been widely used in areas of communication, credit assessment, stock market prediction, marketing, banking, education, health and medicine, hazard forecasting, knowledge acquisition, scientific discovery, fraud detection, etc but data mining holds significant presence in every field of medical for the diagnosis of several diseases such as diabetes, skin cancer, lung cancer, breast cancer, heart disease, kidney failure, kidney stone, liver disorder, hepatitis etc. Data mining applications include analysis of data for better policy making in health, prevention of various errors in hospitals, detection of fraudulent insurance claims early detection and prevention of various diseases, value for more money, saving costs and saving more lives by reducing death rates.

Drug is any substance that when taken into a living organism may modify one or more of its functions. Drugs can provide temporary relief from unhealthy symptoms and/or permanently supply the body with necessary substances the body can no longer make. Some drugs lead to an unhealthy dependency that has both physiological and behavioral roots. Drug addiction can cause serious, long-term consequences, including problems with physical and mental health, relationships, employment, and the law [3]. Adolescence is typically a period of experimentation, irrespective of parenting skills and influence. However, the more likely threat to any teenager's health is the use of drugs such as alcohol and tobacco.

2. Related Work

Automated medical diagnosis helps the doctors to calculate the correct disease with less time. Table 1 highlights the foremost objectives of the authors working in the field of predicting medical disease(s) using data mining methodology. Knowledge gained by exercising of aim(s) of data mining can be used to make booming decisions that will improve success of healthcare organization and health of the patients.

Table 1. Objective(s) of Related Work Done by Number on Medical Diseases

Author	Year of Publication	Disease Considered	Objectives
Dursun Delen et al [3], Bellaachia et al [4]	2005, 2006	Breast cancer	Analysis of the prediction of breast cancer survivability data mining methods.
Asha Rajkumar et al [9]	2010	Heart Disease	To achieve high accuracy by classifying algorithms
D.Senthil Kumar et al [16]	2011	Diabetes, Heart, Hepatitis	Development and evaluation of a clinical decision support system for the treatment of patients with heart disease, diabetes and hepatitis.
Jyoti Soni [22]	2011	Heart Disease	Predictive data mining for medical diagnosis: An overview of heart disease

			prediction
Akhil jabbar et al [11]	2012	Heart Disease	Proposed a system for heart disease prediction using data mining techniques
DSVGK Kaladhar et al [19]	2012	Kidney Stone	Statistical and data mining aspects on kidney stones: a systematic review and meta analysis
Mai Shouman [12]	2012	Heart Disease	Applying K-nearest neighbor in diagnosing heart disease patients
Abhishek Taneja [13]	2013	Heart Disease	To design a predictive model for heart disease detection to enhance their liability of heart disease diagnosis.
Kawsar Ahmed et al [7],[23]	2013	Lung Cancer, Skin Cancer	Early prevention and detection of skin cancer and lung cancer risk using data mining
Syeda Farha Shazmeen et al [21]	2013	Liver Disorder	Performance evaluation of different data mining classification algorithm and predictive analysis
V. Krishnaiah et al [8]	2013	Lung Cancer	Diagnosis of lung cancer prediction system using data mining classification techniques
Vikram Kumar Gupta et al [18]	2013	Drug Addiction	A study of profile of patients admitted in the drug de-addiction centers in the state of Punjab
K R Lakshmi et al [20]	2014	Kidney dialysis	Performance comparison of three data mining techniques for predicting kidney dialysis survivability

Ample of research is done on drugs and its ill effects. Bounty of surveys is available. But no expert system or such effort is seen to have control over drugs. The objectives of work done on drug abusers in past are displayed in Table 2.

Table 2. Objective(s) of Related Work Done by Number on Drug Abuse

Author	Year	Objective(s)
Andres et al [8]	2000	Examination of the effects of acculturation and acculturative stress on the intensity of alcohol involvement during middle school among Latino adolescents.
Brian Borsari et al [10]	2008	Aim to increase the knowledge of drinking norms on college campuses.
Isaac C. Rhew et al [20]	2011	Drug use and risk among youth in different rural contexts
Angelina et al [27]	2013	Aim of this study was to describe the prevalence and predictors of alcohol drinking behavior in children 8-12yrs.
Naresh Nebhinani et al [24]	2013	Aim to study the demographic and clinical profile of women seeking de addiction treatment at a tertiary care center in North India.
Tasia Huckle et al [32]	2014	Long-term effect of lowering the minimum purchase age for alcohol from age 20 to age 18 years on alcohol-involved crashes in New Zealand.
Eduardo et al [31]	2014	The objective of the present study was to examine brain activity during execution and inhibition in young binge drinkers in relation to the progression of their drinking habits response over time.

3. Research Objectives and Questions

This study is aimed at uncovering and analyzing a range of data mining tools and techniques for optimally predicting the numerous medical diseases to endow the healthcare section with high competence and more effectiveness. In achieving the research objectives, this study intends to answer the following five questions:

(i) Does data mining really provide an efficient way to extract the required clinical information from voluminous, raw and heterogeneous data?

(ii) Are there any promising techniques to predict and forecast the medical diseases with high accuracy and low cost?

(iii) What are the various issues and challenges in data mining as applied to the medical practice?

(iv) What are the plausible benefits of this research in relation with substance abuse?

(v) How can we simultaneously retrieve the information and minimize the efforts of an expert system for other clinical concerns like drug addiction?

4. Use of Data Mining in Medical

Today is the era of data mining where prediction of variety of disease is enduring procedure. Data mining has proved with flourished results in medical. But such work is seen in direction to control over drugs usage. Data mining has plenty of techniques and tools available.

4.1. Comparison of Distinct Data Mining Techniques

Different types of mining algorithms in the healthcare field have been proposed by different researchers in recent years. A particular algorithm may not be applied to all the applications due to complexity for appropriate data types of the algorithm. Consequently the choice of an acceptable data mining algorithm depends on not only the purpose of an application, but also on the compatibility of the data set. Table 3 presents the comparative analysis of different data mining techniques and algorithms which have been used by most of the researchers in medical data mining.

Table 3. Comparison of Distinct Data Mining Techniques

Author Name	Year	Data Mining Techniques						
		ANN	DTrees	Logistic Regression	KNN	NB	SVM	Other
Dursun Delen et al [3]	2005	√	√	√	×	×	×	-
Bellaachia et al [4]	2006	√	√	×	×	√	×	-
Asha Rajkumar et al [9]	2010	×	√	×	√	√	×	-
D.Senthil Kumar [16]	2011	×	√	×	×	×	×	-
Jyoti Soni [22]	2011	√	√	×	√	√	×	-
Akhil jabbar et al [11]	2012	√	√	×	×	√	×	-
DSVGK Kaladhar et al [19]	2012	×	√	×	√	√	√	Random Forest, Bagging Algorithm

Mai Shouman [12]	2012	×	×	×	√	×	×	-
Abhishek Taneja [13]	2013	√	√	×	×	√	×	-
Kawsar Ahmed et al [23]	2013	×	×	×	×	×	×	Mafia
Kawsar Ahmed et al [7]	2013	×	√	×	×	×	×	Apriori Algorithm
Syeda Farha Shazmeen et al [21]	2013	√	√	×	√	√	√	-
K R Lakshmi et al. [20]	2014	√	√	√	×	×	×	-

4.2. Comparative Analysis of Data Mining Tools

Due to the extensive use and intricacy involved in building data mining applications, a large number of data mining tools have been developed over the decades. Different tools use diverse algorithm base and techniques to carry out data mining tasks. Every tool has its own advantages and disadvantages. Numerous functionalities offered by these tools incorporate:

- characterization and classification of data,
- patterns evaluation,
- associations and correlations,
- prediction over the data

The maturity and relevance of data mining algorithms necessitates the utilization of influential software tools. As the number of accessible tools continues to develop, the preference of the most suitable tool becomes increasingly tricky. Consequently, a number of authors have suggested and/or used the multiplicity of data mining tools as presented in Table 4.

Table 4. Comparison of Different Used Data Mining Tools

Author Name	Year	Tool Used					
		Weka	Tanagra	Orange	R	Rapid Miner	Knime
Bellaachia et al [4]	2006	√	×	×	×	×	×
Asha Rajkumar et al [9]	2010	×	√	×	×	×	×
D. Senthil Kumar [16]	2011	√	×	×	×	×	×
Jyoti Soni [22]	2011	√	×	×	×	×	×
DSVGK Kaladhar et al [19]	2012	√	×	√	×	×	×
Syeda Farha Shazmeen et al [21]	2013	√	×	×	×	√	×
K R Lakshmi et al [20]	2014	×	√	×	×	×	×

Table 4 illustrates that mostly the tool used by various authors is Weka. Data mining tools foretell the future trends and behaviors that permit the businesses to formulate positive and knowledge-driven decisions. Data mining tools can respond to the business issues that were time consuming to resolve traditionally.

An assortment of surveys and research had been conducted over the utilization of the most popular tool among the organizations. Table 5 represents a wide range of parameters based upon which, a variety of renowned data mining tools have been compared.

Table 5. Description of Numerous Parameters with Respect to Various Data Mining Tools

Tools →	Weka	Orange	R	Knime	Rapid Miner	Tanagra
Parameters ↓						
Ability to Partition the Datasets	Limited	Limited	Limited	Limited	Limited	Limited
Ability to Cross Validate the Model	Limited	Limited	Full	Limited	Full	No
Analysis and Processing Capability	Yes	Yes	Yes	Yes	Yes	Yes
Association Rule Support	Yes	Yes	Yes	Yes	Yes	Yes
Bayes Network Approach	Yes	Yes	No	Yes	Yes	Yes
Clustering Approach	Yes	Yes	Yes	Yes	Yes	Yes
Database System Support	Yes	No	Yes	Yes	Yes	No
Decision Tree Approach	Yes	Yes	Yes	Yes	Yes	Yes
Descriptor Selection Facility	Yes	No	No	No	Yes	Yes
Descriptor Scaling Facility	No	No	No	Yes	Yes	No
Ease of Learning	Better	Best	Better	Better	Good	Good
Evaluation Ability	Yes	Yes	Yes	Yes	Yes	Yes
Feature Selection	Yes	No	No	Yes	Yes	Yes
Graphical Representation	Limited	Limited	Limited	Full	Limited	Limited
Interface	Graphical/ Command Line	Graphical/ Command Line	Graphical/ Command Line	Graphical	Graphical	Graphical
License	Open	Open	Open	Open	Open	Open
Major Users	Academics/ Industry	Academics/ Industry/ Testing/ Bioinformatics	Academics/ Industry	Academics/ Industry	Marketing/ Sales/ Manufacturing/ Telecom	Research
Neural Network Approach	Yes	No	No	Yes	Yes	Yes

Operating System Support	Linux/ Windows/ Mac	Linux/ Windows/ Mac	Unix/ Windows/ Mac	Linux/ Windows/ Mac	Linux/ Windows/ Mac	Windows
Parameter Optimization	No	No	Yes	No	Yes	No
Programming Language	Java	C++, Python	C, Fortran, R	Java	Java	HTML
SVM Approach	Yes	Yes	Yes	Yes	Yes	Yes
Usability	Better	Best	Best	Better	Best	Good

5. Data Sets used in Medical

Below table 6 depicts various databases used with respect to unsafe diseases. Different authors have used various datasets available online or from some hospitals or respective centers. But for drug abusers, no data sets are available online nor are they available at hospitals. It needs interview each patient individually at drug addiction centers.

Table 6. Databases used with Respect to Treacherous Diseases

Author	Disease Considered	Data Set/Database Used
Dursun Delen et al [3], Bellaachia et al [4]	Breast cancer	SEER cancer incidence database
Jyoti Soni [22], Mai Shouman[12]	Heart	Cleveland heart disease database
Akhil jabbar et al [11]	heart	UCI machine Learning repository
Kawsar Ahmed et al [23][7]	Lung cancer, skin Cancer	Different diagnostic centres
Abhishek Taneja [13]	Heart	PGI, chd
D.Senthil Kumar et al [16]	Heart, Diabetes, Hepaititis	UC-Irvine archieve of machine learning repository
K R Lakshmi et al [20]	Kidney dialysis	Hasheminejad kidney centre of Tehran
Syeda Farha Shazmeen et al [21]	Liver Disorder	Fisher's IRIS dataset
V. Krishnaiah et al.[8]	Lung Cancer	UC1 repository

6. Performance Evaluation Measures

Performance is measured using WEKA; an open source machine learning software which gives the output in terms of TP, TN, FP, and FN. Accuracy is interpreted from the given formula.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

Where TP, TN stands for True Positive and True Negative and FP, FN stands for False Positive and False Negative. TP+TN signify percentage of correctly classified instances and TP+FP+TN+FN signifies total of correctly and incorrectly classified instances.

7. Analysis of Accuracies in Medical

Below are accuracies shown applied on various diseases using different data mining techniques. Accuracy is then computed from the above formula to find the no. of affected cases with respect to values in given parameters. Commonly used techniques are Decision trees (Dtrees), Artificial Neural Network (ANN), Naïve Bayes (NB). The Comparison of these techniques used year wise on different disease is analyzed. Table 7 shows the comparison of various diseases on ANN. No such work is done in drug abusers.

Table 7. Accuracies Applied on Artificial Neural Networks

Disease Considered	Author of Publication	Year of Publication	Accuracy in ANN
Breast Cancer	Dursun Delen et al.	2005	91.21%
Heart	Andreeva, P	2006	82.77%
Breast Cancer	Bellaachia et al	2006	86.50%
Heart	Palaniappan, et al.	2007	93.54%
Heart	De Beule, et al.	2007	82.00%
Heart	Tantimongcolwata,et al.	2008	74.50%
Heart	Hara, et al.	2008	82.30%
Heart	Akhil jabbar et al	2012	82.00%
Heart	Abhishek Taneja	2013	93.83%
Liver	Syeda Farha Shazmeen et al	2013	67.59%
Kidney	K R Lakshmi et al.	2014	93.85%

Table 6 shows maximum accuracy of 93.85% of kidney. More diseases are coming to research to find the chance with high accuracy of prediction of risky diseases. Figure 1 shows the graph of accuracies with range of percentage to diseases.

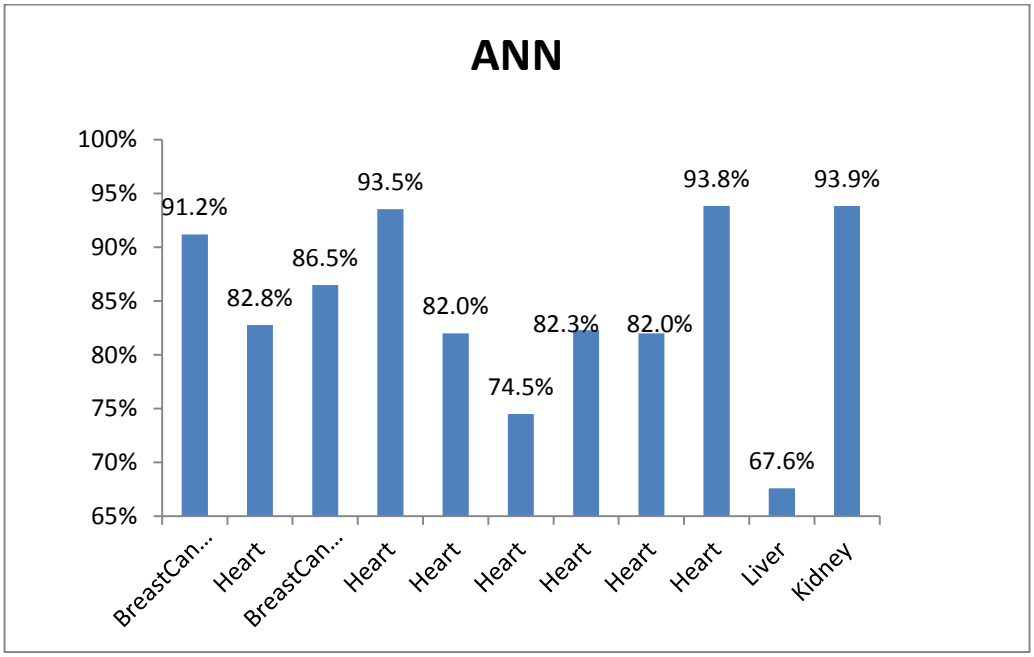

Figure 1. Comparison of Various Diseases on ANN Technique

ANN shows the maximum accuracy of 93.9% and minimum accuracy of 67.6%. Table 7 shows the comparison of various diseases on Dtrees.

Table 7. Accuracy Measure in Decision Trees

Disease Considered	Author	Year of Publication	Accuracies in DTrees
Heart	Cheung,	2001	81.11%
Skin Diseases	Bojarczuk	2001	89.12%
Breast Cancer	Dursun Delen et al.	2005	93.62%
Heart	Andreeva, P	2006	75.73%
Breast Cancer	Bellaachia et al	2006	86.70%
Heart	Palaniappan, et al.	2007	94.93%
Heart	Sitar-Taut et al.	2009	60.40%
Heart	Tu, et al.,	2009	78.90%
Skin Diseases	Polat and Gunes	2009	96.71%
Heart	Asha Rajkumar et al	2010	52.00%
Heart	Jyoti Soni	2011	99.20%
Heart	Akhil jabbar et al	2012	80.00%
Heart	Abhishek Taneja	2013	94.29%
Liver	Syeda Farha Shazmeen et al.	2013	69.58%
Kidney	K R Lakshmi et al.	2014	78.44%

The main objective is to achieve the maximum accuracy among hazardous diseases. Dtrees are most popular data mining technique that is applied and used everywhere and gives maximum optimal results. Figure 2 shows graph of Dtrees used in various medical diseases.

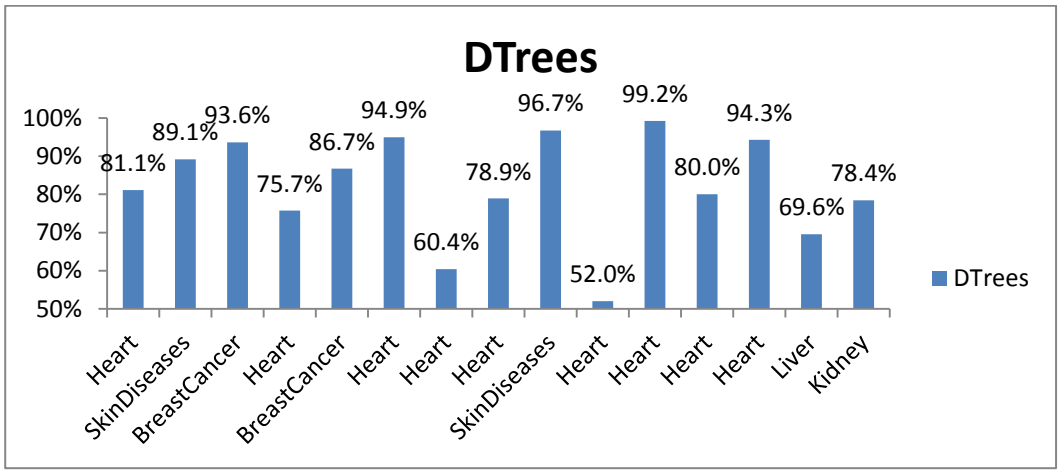

Figure 2. Comparison of various diseases on Dtrees technique

Decision tree is the most popular technique and it has given the maximum accuracy of 99.2%. Table 8 shows the comparison of various diseases on Naïve Bayes.

Table 8. Accuracy Measure in Naïve Bayes

Disease Considered	Author	Year of Publication	Accuracies in NB
Heart	Cheung,	2001	81.48%
Heart	Andreeva, P	2006	78.56%
BC	Bellaachia et al	2006	84.50%
Heart	Palaniappan, et al.	2007	95.00%
Heart	Sitar-Taut et al.	2009	62.03%
Heart	Asha Rajkumar et al	2010	52.33%
Heart	Jyoti Soni	2011	96.50%
Heart	Akhil jabbar et al	2012	76.00%
Heart	Abhishek Taneja	2013	91.96%
Lung Cancer	V. Krishnaiah et al.	2013	84.14%

Navies Bayes is the most common technique that is used in data mining. It gives maximum accuracy of 96.5% in curing heart patients as shown in Figure 3.

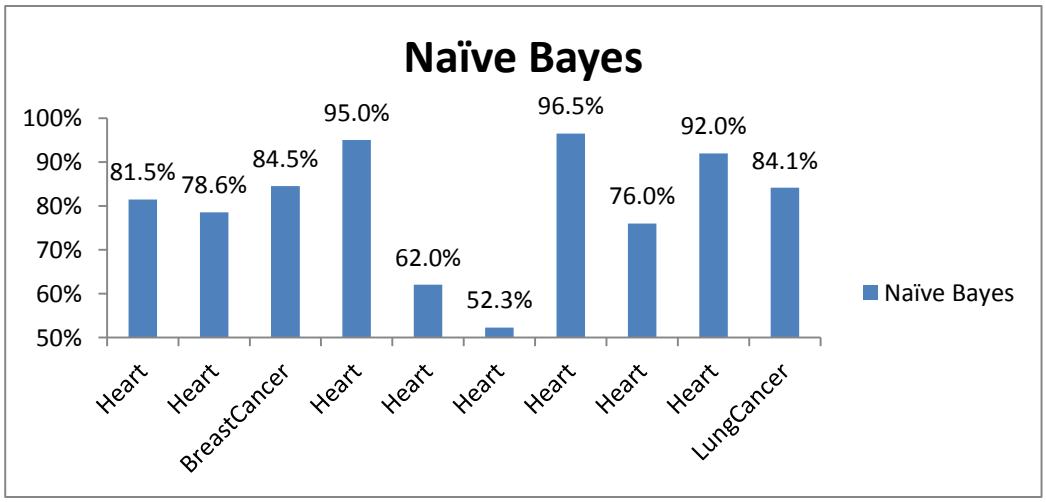

Figure 3. Comparison of Various Diseases on Naïve Bayes Technique

Naïve bayes is widely used technique in prediction of various diseases and has the maximum accuracy of 96.5%.

8. Risk Factors Involved in Healthcare Systems

Although scientists do not know the exact causes of some medical problems, they do know some of the risk factors that increase the likelihood of a disease in human beings. Table 9 highlights some of the major risk factors involved in the diagnosis of medical diseases.

Table 9. Risk Factors involved for the Prediction of Various Medical Diseases

S.No.	Disease	Risk factors
1.	Lung cancer	Smoking (beedi, hookah or cigarette) or second hand smoke, High dose of ionizing radiation, Air pollution, Insufficient consumption of fruits & vegetables [8]

2	Skin cancer	UV light, smoking, large no. of moles on the skin, family history of skin cancer, work outdoors [23]
3	Breast cancer	Drinking alcohol, tobacco, passive smoking, obesity and lack of exercise, night shift work.[4]
4	Heart disease	Smoking, Diabetes, high blood pressure, high cholesterol, low cholesterol, not getting enough physical activity and obesity. [14]
5	Kidney failure	Diabetes, High blood pressure, Heart disease, Smoking, Obesity, High cholesterol, Family history of kidney disease [28]
6	Liver Disorder	Excessive alcohol consumption, Obesity, Diabetes, Tobacco use, Cirrhosis, Hereditary, Exposure to aflatoxins, viruses(primarily hepatitis A [HAV], hepatitis B [HBV], or hepatitis C [HCV]) [21]

The risk factors of various diseases conclude towards the common risk factor i.e. intake of drugs that leads to every hazardous diseases like cardiovascular disease, cancer, Lungs and kidney malfunction and many more. Drug abuse may trigger or exacerbate mental disorders, anxiety, stress, depressions and even death.

Effect on health of drug users

A drug is a social problem these days and youth is in the falling in trap of drugs. There is need to aware the youth about the common problems with drug abusers.

Common Problems with drug abusers are:-

- HIV, Hepatitis and Other Infectious Diseases
- Cardiovascular Effects
- Respiratory Effects
- Gastrointestinal Effects
- Musculoskeletal Effects
- Kidney Damage
- Liver Damage
- Neurological Effects
- Mental Health Effects
- Hormonal Effects
- Cancer
- Prenatal Effects
- Other Health Effects
- Mortality

The research work mainly focused on knowing the major motive for commencement and indulging the youth in drugs. The research work is of great help for analyzing various factors for booming situation of drugs. The system is of great relevance to the user in detection of the various factors related to drug addiction which will help in providing the correct medication about him/her and will help in saving his precious human live. No such work is done in this direction.

9. Parameters Selected for Drug Addicted People

21 parameters have been selected that are most important. The author has performed the Experiments on a real data set to study the impact of constraints and the elimination of unreliable rules with validation on the test set. Table 10 specifies the values with respect to parameters.

Table 10. Parameters for Drug Addicted People

Parameters	Values
Sex	Male, Female
Age	upto20,21_40,41_60,above60
Residence	urban, rural
Type of locality	slum, private, govt. approved
Marital Status	Married, Unmarried, Divorced
Type of family	Joint, Nuclear
Education	illiterate, Primary, Less than primary, Secondary, graduation, above graduation
Occupation	unemployed, student, job, business
Age of Initiation	upto20, above20
Duration of substance abuse	1_10 yrs,11_20 yrs ,above 20yrs
Family Income	Rs(upto5000, 6_10000, 11_25000, 26_40000, above40000)
House	Own, Rented
Who prompted you for drugs	internal (personal) factors, external factors
In which company it was taken	internal (house, shop, etc) place, external place (marriage , schools, colleges, etc)
Family history of drugs	Yes, No
Daily expenditure on drugs	Rs (1_200, 201_500, 501_1000, above1000)
In case no money for drugs	Yes(stealing, theft, borrow, etc), No
Did you sell any house hold articles if you don't have money for buying drugs	Yes, No
Any accident while having drugs	Yes (1time or more), No
Conflict with law	Yes (1time or more), No
Anyone helped you in getting rid of this drug taking problem	Yes, No

10. Proposed Methodology using WEKA Tool

To evaluate the performance of our approach, a data set is surveyed with 40 patients admitted at Ludhiana drug de-addiction centre. Then data is loaded into the WEKA tool, after the dataset has been loaded. Naïve Bayes, Decision tress (J48), Multilayer perceptron (MLP), Logistic regression are selected. Data is then cross validated using performance classifier measure, the results and performance of each algorithm is then compared to each other. Figure 4 reveals the working of WEKA tool.

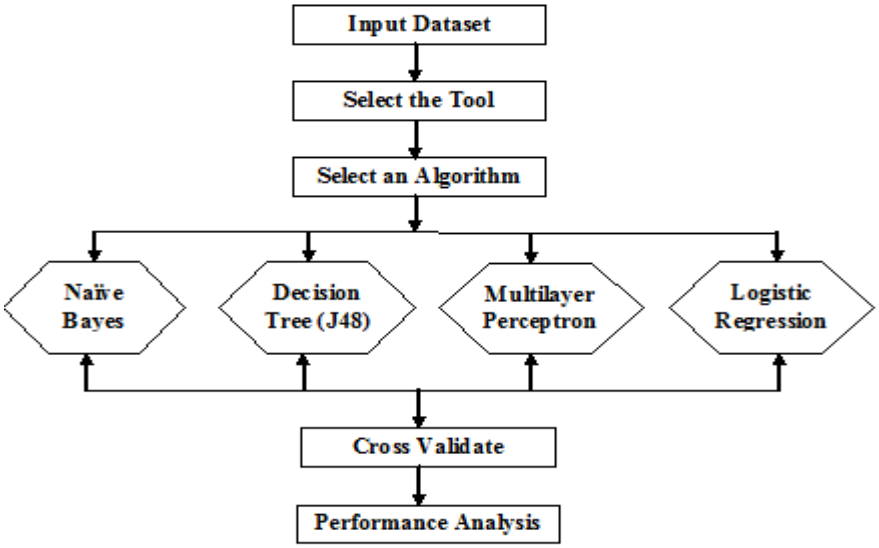

Figure 4. Working of WEKA Tool

11. Results

Accuracy is major constraint in medical field which fails with minor fluctuations. More the accuracy better are the results. Day by day more research work is going to achieve results with high accuracy and less effort to save precious human life before the problem occurs. Accuracy is measured in terms of correctly classified instances. Table 11 exemplifies data mining techniques with accuracy or correctly/ incorrectly classified techniques.

Table 11. Comparison of Various Data Mining Techniques

Algorithm	Correctly Classified Instances	Incorrectly Classified Instances
Navie Bayes	80%	20%
J48 (DTrees)	95%	5%
Multilayer Perceptron(MLP)	90%	10%
Logistic Regression	87.50%	12.50%

Decision trees give the maximum accuracy of 95% which respect to age of initiation parameter *i.e.*, J48 gives maximum accuracy of 95% among the affected patients started using their drug usage at below 20 yrs of age. The Statistics are disclosed in Figure 5 exposing the J48 having 95% of accuracy.

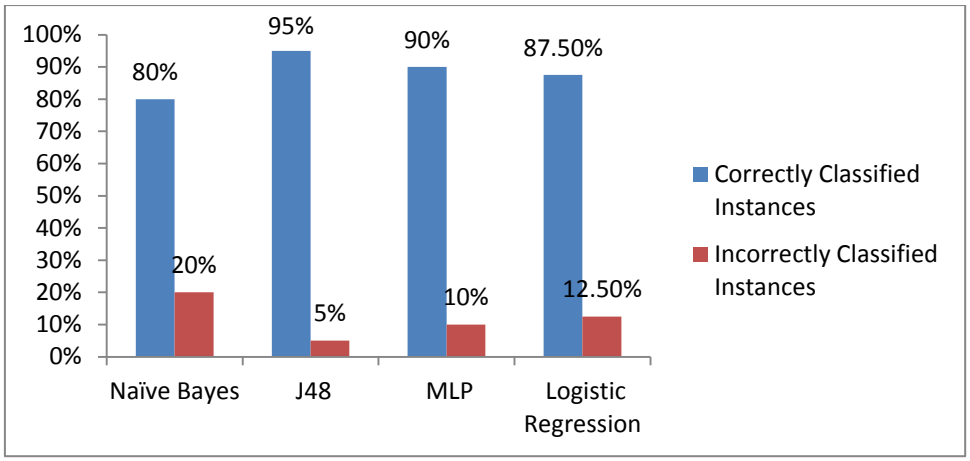

Figure 5. Analysis of Data Mining Techniques

12. Conclusion

Applying data mining in the medical field is an incredibly challenging mission due to the idiosyncrasies of the medical profession. It characterizes widespread process that demands thorough understanding of needs of the healthcare organizations. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve success of healthcare organization and health of the patients. This review of data mining applications in medicine and public health has endowed us with an overview of current practices and challenges. Health care organizations and agencies could come across into these applications to find ideas on how to dig out knowledge from their own database systems. Data mining requires suitable technology and analytical techniques, as well as systems for reporting and tracking which can facilitate measuring of results. The available raw medical data are widely distributed, different and voluminous by nature that must be collected and stored in data warehouses in organized form.

Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires and conditions of the patients and to make adequate and optimal decisions about their treatments. Thus, there exist promising techniques to predict and forecast the medical diseases with high accuracy and low cost With the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in the medical data so as to simultaneously retrieve the information and minimize the efforts of an expert system for various diseases under consideration. Within the issue of knowledge integrity assessment, two biggest challenges are: (1) How to develop efficient algorithms for comparing content of two knowledge versions (before and after). This challenge demands development of efficient algorithms and data structures for evaluation of knowledge integrity in the data set; and (2) How to develop algorithms for evaluating the influence of particular data modifications on statistical importance of individual patterns that are collected with the help of common classes of data mining algorithm.

Algorithms that measure the influence that modifications of data values have on discovered statistical importance of patterns are being developed, although it would be impossible to develop a universal measure for all data mining algorithms. Even if data mining results are credible, convincing the health practitioners to change their habits based on evidence may be a bigger problem. Data mining in healthcare can be limited in data access, since the raw inputs for data mining frequently exist in different settings and systems, like administrations, clinics, laboratories etc. Therefore, data must be collected and integrated before data mining can take place. Building of data warehouse before data mining begins can be a very expensive and time consuming process. Data mining project

can fail from numerous reasons, like lack of managerial support, inadequate data mining expertise etc. Healthcare organizations that develop data mining must use big investment resources, especially time, effort and money.

Survey of data mining in medical field outputs towards the common risk factor i.e. drug addiction. Many people do not know when they get addicted to drugs. Drug is a brain eater. Everyday a new stories comes disastrous effects of drugs into the human life. Long term drug usage affects the brain and other important organs of the body. So, data mining is preferred for prediction that gives results in terms of accuracy which is the major concern in medical field.

Yes Data mining really provide an efficient way to extract the required clinical information from voluminous, raw and heterogeneous data. Decision trees, Neural Networks, Logistic regression, Naïve Bayes etc are the promising techniques that predict and forecast the medical diseases with high accuracy and low cost. Various issues and challenges in data mining applied to medical field are focusing on survival of the unaware affected person by predicting the cause of disease at earlier stage. Drug addiction using WEKA has been shown and it brings into light that majority of drug abusers started abusing drugs at age below 20yrs.It is to make aware the druggist about the various diseases that are caused with heavy or long term intake of drugs in their life. So, to make an expert system that will awake the youth about precarious use of drugs and also alert the affected person.

13. Future Work

Our future work will involve the amalgamation of the various specified algorithms to augment the accuracy so that the diagnosis can develop into more accurate in case of imperceptibly identified data sets. The research work mainly focused on knowing the major motive for commencement and indulging the youth in drugs. Ongoing efforts are geared towards increasing the size of data set. The research work is of great help for analyzing various factors for booming situation of drugs. The system is of great relevance to the user in detection of the various factors related to drug addiction which will help in providing the correct medication about him/her and will help in saving his precious human live.

References

[1] Joseph D. Bronzino, Ralph A. Morelli, John W. Goethe," Design of an expert system for monitoring drug treatment in a psychiatric hospital", Computer-Based Medical Systems (CBMS), Fourth Annual IEEE Symposium CBMS pp. 219-225, 1991.

[2] Neat, Gregory W. ; Rensselaer Polytech. Inst., Troy, NY, USA ; Kaufman, H. ; Roy, Rob J. ," Expert adaptive control for drug delivery systems", Control Systems Magazine, IEEE , Vol. 9 , No. 4, pp. 20 – 24, June 1989.

[3] Gil, A.G., Wagner, E.F., Vega, W.A. Acculturation, familism, and alcohol use among Latino adolescent males (Longitudinal relations) . J Community Psychol (john wiley & sons) Vol. 28, pp. 443–458, 2000.

[4] Borsari, B., & Carey, K. (2003). Descriptive and injunctive norms in college drinking: A meta-analytic integration. Journal of Studies on Alcohol, Vol. 64, pp. 331–341, 2003.

[5] D. Delen, G. Walker, and A. Kadam, Predicting breast cancer survivability: a comparison of three data mining methods, Artificial Intelligence in Medicine (Elsevier), vol. 34, no. 2, pp. 113–127, 2005.

[6] A. Bellaachia and E. Guven," Predicting breast cancer survivability using data mining techniques", In Proceedings of Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining(SDM 2006), 2006.

[7] Rhew, I.C. Drug use and risk among youth in different rural contexts. Health Place (Elsevier). Vol. 17, pp. 775–783, 2011.

[8] Shweta Kharya," Using data mining techniques for diagnosis and prognosis of cancer disease", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.

[9] Angelina Pilatti , Juan Carlos Godoy , Silvina Brussino , Ricardo Marcos Pautassi," Underage drinking: Prevalence and risk factors associated with drinking experiences among Argentinean children", Alcohol(Elsevier), Vol. 47, pp. 323-331, 2013.

[10] Naresh Nebhinani**,** Shubh M Singh**,** Gourav Gupta," Demographic and clinical profile of substance abusing women seeking treatment at a de-addiction center in north India," Industrial Psychiatry journal, Vol. 22, 2013.
[11] Eduardo López-Caneda, Socorro Rodríguez Holguín, Montserrat Corral, Sonia Doallo, Fernando Cadaveira, Evolution of the binge drinking pattern in college students: Neurophysiological correlates, Alcohol (Elsevier) Vol. 48 , 2014.
[12] Taisia Huckle, and Karl Parker, Long-Term Impact on Alcohol-Involved Crashes of Lowering the Minimum Purchase Age in New Zealand, American Journal of Public Health, Vol. 104, pp.1087-1091,June 2014.
[13] G. Ravi Kumar, Dr. G. A. Ramachandra, K.Nagamani," An Efficient Prediction of Breast Cancer Data using Data Mining Techniques", International Journal of Innovations in Engineering and Technology (IJIET),Vol. 2 ,No. 4, pp139-144, August 2013.
[14] Kawsar Ahmed, Abdullah Al Emran, Tasnuba Jesmin, Roushney Fatima Mukti,Md Zamilur Rahman, Farzana Ahmed," Early Detection of Lung Cancer Risk Using Data Mining", Asian Pacific Journal of Cancer Prevention, Vol. 14,pp. 595-598,2013.
[15] V.Krishnaiah "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 4, No. 1 , pp. 39 –45, 2013.
[16] Asha Rajkumar and B. Sophia Reena, "Diagnosis Of Heart Disease Using Data mining Algorithm" , Global Journal of Computer Science and Technology, Vol. 10, No. 10, pp. 38 - 43, 2010
[17] M.Anbarasi, E. Anupriya, N.CH.S.N.Iyengar, Enhanced Preditction of Heart Disease with Feature Subset Selection using Genetic Algorithm. International Journal of Engineering Science and Technology,Vol. 2, No.10,pp. 5370-5376, 2010.
[18] M. Akhil Jabbar, Bulusu Lakshmana Deekshatulu, Priti Chandra," Heart Disease Prediction System using Associative Classification and Genetic Algorithm", International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies(ICECIT), 2012 .
[19] Mai Shouman, Tim Turner, Rob Stocker, "Using Data Mining Techniques In Heart Disease Diagnosis And Treatment ",Proceedings in Japan-Egypt Conference on Electronics, Communications and Computers,IEEE,Vol.2 pp.174-177,2012.
[20] Abhishek taneja," Heart Disease Prediction System Using Data Mining Techniques", Oriental journal of Computer science & technology, Vol. 6, No. 4: pp. 457-466, December 2013.
[21] V. Chauraisa and S. Pal, "Data Mining Approach to Detect Heart Diseases", International Journal of Advanced Computer Science and Information Technology (IJACSIT), Vol. 2, No. 4, pp 56-66, 2013.
[22] S.J Gnanasoundhari, G.Visalatchi, Dr.M.Balamurugan," A Survey on Heart Disease Prediction System Using Data Mining Techniques", International Journal of Computer Science and Mobile Applications, Vol. 2 No. 2, pp. 72-77, February- 2014.
[23] D.S. Kumar, G. Sathyadevi, S. Sivanesh,Decision support system for medical diagnosis using data mining, Journal of Computer Science, 8 (3) (2011), pp. 147–153
[24] Rajesh K & Sangeetha V," Application of data mining methods and techniques for diabetes diagnosis", International journal of engineering and innovative technology, Vol. 2. Issue 3, 2012.
[25] Vikram Kumar Gupta , Paramjeet Kaur, Gurmeet Singh, Amanpreet Kaur, B. S. Sidhu," A study of profile of patients admitted in the drug de-addiction centers in the state of Punjab", International Journal of Research in Health Sciences, Vol. 1, Issue-2, 2014.
[26] DD Kaladhar, KA Rayavarapu, and Varahalarao Vadlapudi," Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis Journal of Biometrics and Biostatistics", Vol. 1, No. 12, pp. 1-5, 2012.
[27] K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna, "Performance comparison of three data mining techniques for predicting kidney disease survivability", International Journal of Advances in Engineering & Technology, Vol. 7, Issue 1, pp. 242-254 , March 2014.
[28] S.F. Shazmeen, M.M.A. Baig, and M.R. Pawar, "Performance Evaluation of Different Data Mining Classification Algorithm and Predictive Analysis," Journal of Computer Engineering, Vol. 10, No. 6, pp. 01-06, 2013.
[29] 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 (0975 – 8887), Vol. 17, No. 8, March 2011.
[30] Kawsar Ahmed,Tasnuba Jesmin,Md. Zamilur Rahman:"Early Prevention and Detection of Skin Cancer Risk using Data Mining", .International Journal of Computer ,Vol. 62 ,No. 4, pp. 1-6, 2013
[31] A. Laribi, S. A. Laribi," An Intelligent System to Facilitate the Diagnosis of Adverse Drug Reactions", IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp. 661-666, 1994
[32] Andreeva, P., "Data Modeling and Specific Rule Generation via Data Mining Techniques", International Conference on Computer Systems and Technologies - CompSysTech, 2006.
[33] Sellappan Palaniappan, Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques" International Conference on Computer Systems and Applications, IEEE/ACS, pp.108-115, AICCSA April 2008.

[34] Srinivas, Rao and Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", In IEEE Proceedings of the 5th International Conference on Computer Science and Education, Hefei, China, pp. 1344-1349, 2010.
[35] Andrew Kusiak, Bradley Dixon, Shital C. Shah," Predicting survival time for kidney dialysis patients: a data mining approach.",Comp. in Bio. and Med. 35(4): pp.311-327, 2005.
[36] Annis, H. M., & Davis, C. S," Relapse prevention. In R. K. Hester & W. R. Miller (Eds.)", Handbook of alcoholism treatment approaches: Alternative approaches pp. 170 - 182. New York: Pergamon Press, 1989.
[37] Naresh Nebhinani**,** Shubh M Singh**,** Gourav Gupta," Demographic and clinical profile of substance abusing women seeking treatment at a de‐addiction center in north India," Industrial Psychiatry journal, Vol. 22, Issue 1, 2013.
[38] Indiver Kalra, Pir Dutt Bansal," Sociodemographic Profile and Pattern of Drug abuse among Patients Presenting to a Deaddiction Centre in rural area of Punjab", delhi psychiatry journal, Vol. 15, No.2, 2012.
[39] Angelina Pilatti , Juan Carlos Godoy , Silvina Brussino , Ricardo Marcos Pautassi," Underage drinking: Prevalence and risk factors associated with drinking experiences among Argentinean children", Alcohol(Elsevier),Vol. 47, Issue. 4, pp. 323-331,2013.
[40] C. Duff,"Party drugs and party people: Examining the "normalisation" of recreational drug use in Melbourne, Australia", International Journal of Drug Policy (Elsevier), Vol. 16, Issue. 3, pp. 161–170, 2005.
[41] G.Sathyadevi, "Application of CART Algorithm in Hepatitis Disease Diagnosis", Recent Trends in Information Technology (ICRTIT), International Conference IEEE, pp. 1283-1287, 2011.
[42] Yan, H., et al., Development of a decision support system for heart disease diagnosis using multilayer perceptron. Proceedings of the 2003 International Symposium on, 2003. vol.5: p. pp. 709-712.
[43] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.
[44] M. De Beule, E. Maes, O. De Winter, W. Vanlaere, R. Van Impe, Artificial neural networks and risk stratification: A promising combination Mathematical and Computer Modelling (Elsevier), 46 (2007), pp. 88–94
[45] C. L. Chang and C. H. Chen, "Applying decision tree and neural network to increase quality of dermatologic diagnosis", Expert Systems with Applications, Elsevier, vol. 36, (2009), pp. 4035-4041.
[46] Tu, M.C., D. Shin, and D. Shin, Effective Diagnosis of Heart Disease through Bagging Approach. Biomedical Engineering and Informatics, IEEE, 2009.
[47] Tanawut Tantimongcolwat , Thanakorn Naenna , Chartchalerm Isarankura-Na-Ayudhya , Mark J. Embrechts , Virapong Prachayasittikul, Identification of ischemic heart disease via machine learning analysis on magnetocardiograms, Computers in Biology and Medicine (Elsevier), v.38 n.7, p.817-825, July, 2008
[48] A. Hara and T. Ichimura, "Data Mining by Soft Computing Methods for the Coronary Heart Disease Database", Fourth International Workshop on Computational Intelligence & Application, IEEE SMC Hiroshima Chapter, Hiroshima University, Japan, **(2008)** December 10-11.
[49] H.A. Guvenir and N. Emeksiz, "An expert system for the differential diagnosis of erythemato-squamous diseases," Expert Systems with Applications (Elsevier), Vol. 18, Issue. 1, pp. 43–49, 2000.
[50] C. C., Lopes, H. S., Freitas, A. A., Bojarczuk, "Data Mining with Constrained-Syntax Genetic Programming: Applications in Medical Data Set," in Data Analysis in Medicine and Pharmacology (IDAMAP-2001), a Workshop at Medinfo-2001, London, UK, 2001.
[51] K. Polat and S. Gunes, "A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems," Expert Systems with Applications (Elsevier), vol. 36, no. 2, pp. 1587-1592, 2009.
[52] E. D. Ubeyli and E. Dogdu, "Automatic Detection of Erythemato-Squamous Diseases Using k-Means Clustering," Journal of Medical Systems (Springer), Vol. 34, pp. 179-184, 2010.
[53] J. Xie and Ch. Wang, "Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases," Expert Systems with Applications, vol. 38, no. 5, pp. 5809–5815, 2011.
[54] Kumar, D. S., Sathyadevi, G. and Sivanesh, S.," Decision Support System for Medical Diagnosis Using Data Mining", IJCSI International Journal of Computer Science Issues, Vol. 8, No. 1, 2011.

