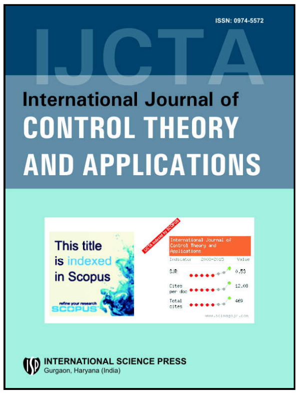

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 18 • 2017

Implementation of Classification Rule Mining to minimize Liver Disorder risks

Sushruta Mishra[1], Hrudaya Kumar Tripathy[2], Brojo Mishra[3] and Soumya Sahoo[4]

[1] School of Computer Engineering KIIT University, Bhubaneswar, Odisha, INDIA, Emails: mishra.sushruta@gmail.com
[2] School of Computer Engineering, KIIT University, Bhubaneswar, Odisha, INDIA, Email: hrudayakumar@gmail.com
[3] Dept of IT C.V. Raman college of Engineering, Odisha, INDIA, Email: brojokishoremishra@gmail.com
[4] Dept of CSE C.V. Raman college of Engineering, Odisha, INDIA, Email: soumya.sahoo685@gmail.com

Abstract: A new arena for generation of data has developed due to massive growth of internet. Handling such massive amount of data is a hot topic of discussion in this new age of technology. Classification Rule Mining is the solution to address this issue. It is a recent technology that integrates conventional data analysis methods with complex algorithms to process massive data in real time scenario. Several successful implementation of Classification Rule Mining includes finance, agriculture, business analytics etc. Healthcare diagnosis is one such critical area where such mining can be successfully applied. It can be used to identify health risk factors thereby reducing diagnosis time and improving diagnosis accuracy rate. Recently in healthcare domain Liver Disorders are very prevalent in human beings. These liver disorder risks are tough to detect at early phases hence sophisticated techniques need to be developed to determine the condition of the Liver disorders. The prime focus of this paper is to study the use of Classification Rule Mining in efficient diagnosis of risk factors of Liver Disorders. Many Classification Mining based algorithms which include Decision Table, OneR and DTNB are used to solve this issue. Some vital performance parameters are used to evaluate the effectiveness of such algorithms. The results are demonstrated in graphical form and it is inferred that DTNB algorithm yields the most optimal performance identifying and reducing liver disorder risks.

Keywords: Classification rule Mining, Liver Disorders, OneR, DTNB, Decision Table, Healthcare Domain.

1. INTRODUCTION

Data mining in healthcare sector has the ability to show hidden patterns in data samples of medical field. [1] Data samples available in medical records are in distributed manner in heterogeneous locations which are complicated to handle. These medical records contain much vital information which cannot be explored in this present scenario. Sophisticated automated tools are required to extract the required knowledge from such hidden data stores. There are many algorithms which were used to extract the vital knowledge from these datasets. Data Mining is an important field of study which can be successfully used in such situations. It is a process where intelligent techniques are used to uncover hidden patterns in large amount of data. It finds huge applications in wide areas like telecommunications, scientific zone, informatics, tourism and many other sectors. Recently data mining is applied in the field of healthcare for identification of disease risks. Classification Rule Mining is an important aspect of Data Mining that is widely used today. Many complex issues can be handled using this Classification Rule Mining [2]. There are several health hazards occurring every now and then in the world. Hazards due to Liver disorders poses a serious threat to human lives now where it is difficult to detect the origin of problems at an early phase. Diseases due to liver disfunctioning are a very life threatening disease internally now which need to be addressed [3]. Beyond one million cases suffering from liver disorder are detected every year on a global level [4]. It is one of the leading causes of death in human race. Due to imprecise knowledge of its symptoms it is very challenging to identify its risks [3]. Several possible risk factors leading to liver damage include smoking, obesity and excess alcoholism [4][5]. These liver disorders are more frequent in case of male community [6]. Thus precise diagnosis of such disorders is the need of the hour which requires efficient execution at root level. Due to resource depletion and imposition of heavy cost an automated intelligent system model is required for effective detection and prevention of such hazards caused due to liver disorders. Thus Classification Rule Mining is a feasible alternative that can be helpful in efficient diagnosis of liver disorder risks that can maximize performance thereby minimizing cost [1].

2. RELATED WORK

Experienced and proficient medical diagnosis is needed proficiency to deal with uncertainty. Although there is an expansion of medical science these days,some sort of limitations can be beaten [7] by making available a framework to create the model using fuzzy theory towards medical diagnosis improvement. There was a drafting of a fuzzy system for learning, analysis and investigation of liver disorder. The system becomes fast, cheap, and also more amenable and accurate than other conventional diagnostic system. This system is a combination of an expert and a fuzzy system and they are known as hybrid system. The verification or accuracy of proposed system is 91%. The intensity of diseases can be found out by the system properly. ANN learning capacity and accuracy has been evolved by the application of new meta-heuristic algorithm, centripetal accelerated particle swarm optimization (CAPSO) is applied[8]. The hybrid learning of CAPSO and multi-layer perceptron (MLP) network are used to perform classification of the data. The efficiency of the method is figured out based on mean square error, accuracy, sensitivity, specificity, and area under the receiver operating characteristics (ROC) curve. Better performance is achieved by this method in compared to other methods in terms of testing data and data sets with more number of missing values. [9]Another method of unsupervised feature learning has been applied for cancer detection and cancer type analysis from gene expression data. This method is used for detection and classification of cancer types. This method shows better performance than the conventional methods due to application of data from different types of cancer automatically generates feature to uplift the detection and diagnosis of a particular type. There was an improvement of accuracy in the classification problem and it also gives more general and expansible approach to deal with gene expression data over various cancer types. The machine learning techniques are applicable to evolve various classifiers for detecting and performingdiagnosis of disease but in clinical approved examining technique so far restricted and the method is prone to be over fitting. So for overcoming the problem, Kenneth R Foster, Robert Koprowski and Joseph D Skufca [10] focused by using support vector machine, a computation oriented thorough going statistical method.They had used leave-one-out cross validation method here [11] .There was a development of afuzzy MLP (Multilayer Perceptron) model to manage the ambigious or impreciseness in input and output. They had used this model as a connection basedexpert system for diagnosis hepatobiliary disorders. In the real world, there is a high uncertainity and noisy data. To deal with highly uncertain and noisy data e.g. biochemical laboratory examinations a classifier is required. So, I-Jen Chiang, Ming-Jium Shiem, Jane Yung-Jen Hsu, Jau-Min [12] proposed a fuzzy classifier to perform classification on the noisy and uncertain data. Rather performing determination of a single class for a given instance, fuzzy classification gives prediction about the degree of possibility for every class. Varun Kumar, Luxmi Verma [13] used the binary classification tasks to diagnose the clinical test to know if a patient is suffering from certain disease or not. They used the various classification algorithms namely ID3, K-NN, C4.5 and SVM on the breast cancer database and the performance of K-NN gives very good classification result with accuracy rate and robustness. So to handle non-linear and dependent data Rong-Ho Lin, Chun-Ling Chuang [14] developed an smart liver Diagnosis model (ILDM). This model incorporates artificial neural networks (ANN), analytic hierarchy process (AHP) and case-based reasoning (CBR) methods. These methods find out whether a patientis suffering from liver disease and if yes it determines the type of the liver disease.Many research groups have developed different approaches for liver cancer. Sammouda.et.al uses histogram, physically, neighbourhood based segmentation. It uses HNN artificial neural networks for classification [15]. Upadhyay & Wasson uses genetic algorithm, region growing, threshold based, level set method, statistical model, and histogram based approach for liver cancer detection [16]. Selvaraj &S.based on particle swarm optimization on liver cancer diagnosis [17]. Massieh.et.al uses 3-D consistency check based on knowledge based constraints [18]. Kundra & Pandey uses j48 algorithm for the classification of diseases.

3. PROBLEM DESCRIPTION

Various categories of liver disorders are as follows:

3.1. Fatty Liver Diseases (FLD)

In this disease excess fat is developed in the liver cells. It is caused by excess drug use, obesity and alcohol consumption. Related symptoms of FLD include AM, WTL, FG, PRA, LOC and WK.

3.2. Wilson Diseases (WD)

It is an inherited hazard that occurs in an age group of 12-23 people. Its symptoms are ML, FV, BC, DP, FG and SS.

3.3. Inherited Liver disease (IHD)

It is very common type of disorder in liver which have a pathological significance on liver cells. The signs of IHD are FG, JP, DI, JA, CC, SAF and WTL.

3.4. Autoimmune Liver Diseases (AID)

This is a liver disorder where the liver cells are damaged by the body immune system. The symptoms are DU, ABVS, FG, JP, VM and EL.

3.5. Cholestasis Liver Diseases (CLD)

Basically there are two types of Cholestasis which are Primary Sclerosing Cholangitis (PSC) and Primary Biliary Cirrhosis (PBC). Some vital symptoms include FG, SC, FV, IS, RAP and JA.

Table 1 shows various liver based diseases. In this table diseases are denoted in rows while their corresponding metrics are shown in columns. If a particular symptom is present in a row then its sub-columns including psychiatric, physical, cognitive, neurology and pathology contains “Y” else if it is absent then it contains a “N”. Hence the Problem Description Table is developed with a total of five rows corresponding to diseases and its various metrics constituting 24 columns.As it can be seen it has two critical metrics which include Neurological Psychiatric and Pathological-Physical Cognition. These two parameters are sub-divided into its constituents as follows.

Pathological-Physical Cognition has three subparts.

- Pathological (PATH) metric (Joint Pain (JP), Diabetes (DI), Chronic Cough (CC), Swelling of the Ankles and Feet (SAF)
- Physical (PHY) metric (Fatigue(FG),Weakness(WK), Pain in the Right Abdomen(PRA), Weight loss(WTL), Nausea(NA), Loss of Appetite(LOA), Vomiting(VM), Jaundice(JD)
- Cognition metric (Enlarged Liver(EL), Dark Urine(DU), Abnormal Blood Vessels on Skin(ABVS), Abdominal Distention(AD))

Neurological Psychiatric has two subparts.

- Neurological metric (Slurred Speech (SS), Failing Voice (FV)).
- Psychiatric metric (Behavioural changes (BC), Memory Loss (MC), Depression(DP)).

Table 1
Problem Description Table of liver disorders

DISEASES	NEUROLOGICAL PSYCHIATIC					PATHOPHYSICAL-COGNITIVE																	
	NEUROLOGY (NEU)		PSYCH IATIC (PSY)			PATHOLOGICAL (PATH)				PHYSICAL (PHY)										COGNITIVE (COG)			
	SS	FV	DP	BC	MC	J P	DI	CC	SAF	FG	WK	PRA	WL	NA	LOA	VM	JA	SS	SC	EL	DU	AVS	AD
FLD	N	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
WILSON	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	N	N	N	N	N	Y	N	N	N	N	N	N
CHOLESTATIC	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N	N	N	Y	N	N	N	N	N	N
INHERITED	N	N	N	N	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	N
ALD	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	Y	N	N	Y	Y	Y	Y

4. PROPOSED WORK

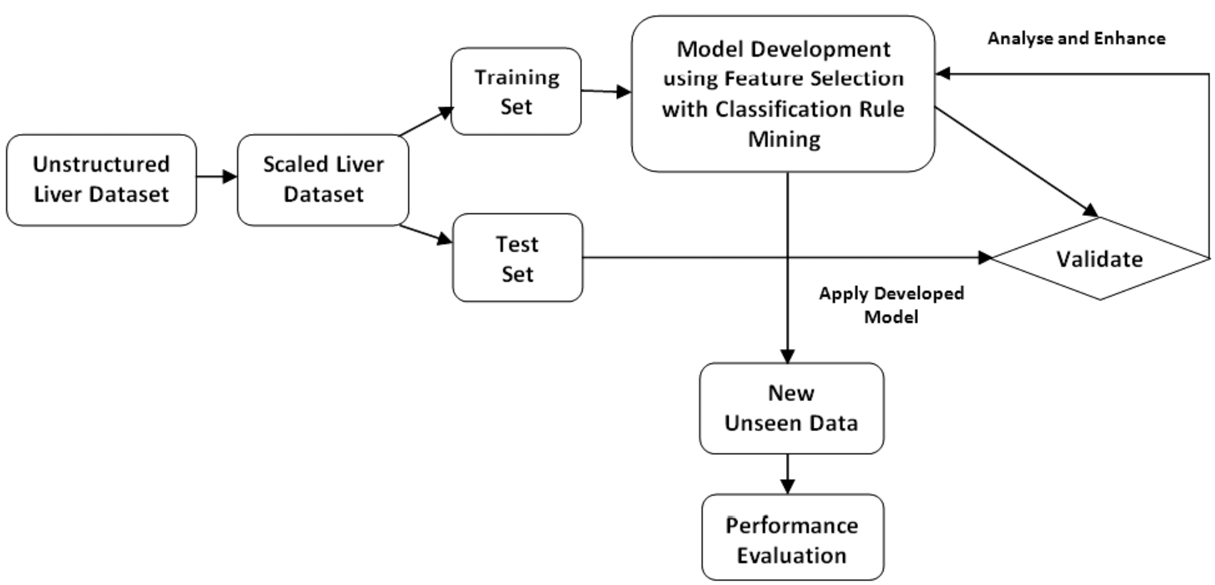

Figure 1: Proposed Classification Rule Mining Model for Liver disorder Risks Diagnosis

Our proposed work is illustrated in figure 1. The liver dataset is collected from different sources with several attributes. This data is in unstructured form so it has to be transformed into structured form. After the data is converted into a format that can be subjected to classification, it is partitioned into training data and testing data. The training set is applied to fit the models while the validation set is used to determine classification error for the system model chosen and subsequently the test data set is used to assess the generalization error of the final selected classification model. While performing Classification rule mining the original dataset is generally segregated into three forms which include Training set, Validation set and Test set.

Training set: a set of examples used for learning: to fit the parameters of the classifier In the MLP case, we would use the training set to find the "optimal" weights with the back-prop rule

Validation set: a set of examples used to tune the parameters of a classifier In the MLP case, we would use the validation set to find the "optimal" number of hidden units or determine a stopping point for the back-propagation algorithm

Test set: a set of examples used only to assess the performance of a fully-trained classifier In the MLP case, we would use the test to estimate the error rate after we have chosen the final model (MLP size and actual weights)

Once the model is trained it can be subjected to new unseen data for classification. At every round there is scope of improvement of the developed model. It can be further boosted up and enhanced so that it can predict the liver disorder risks accurately. Finally the developed model can be evaluated using various performance indices as shown in the result section. These performance indicators determine the efficiency of the model. Our system model comprises of various Classification rule Mining algorithms like Decision Table, OneR and DTNB while Genetic Search is the feature selection method applied to the datasets. The main objective of the research is to first understand or deep studies of the Classification rule Mining algorithms and analyse the data from the surveys and to find which technique gives the better performance than other methods. The comparison of the techniques or methods is done on the basis of various metrics like accuracy, error rate, latency and Kappa metric.

5. RESULTS AND DISCUSSION

Casualties due to liver disorders are a big threat to life. There was a detail survey by medical practitioners from Fortis hospital. They observed that beyond 2 lakh patients die with liver based diseases while around 20,000 patients require liver transplantation every year. The figure 2 depicts the various categories of liver disorder diseases. As it can be seen out of different types of liver diseases maximum type is Wilson disease with 33% while minimum is Inherited Liver Disease with nearly 7%.

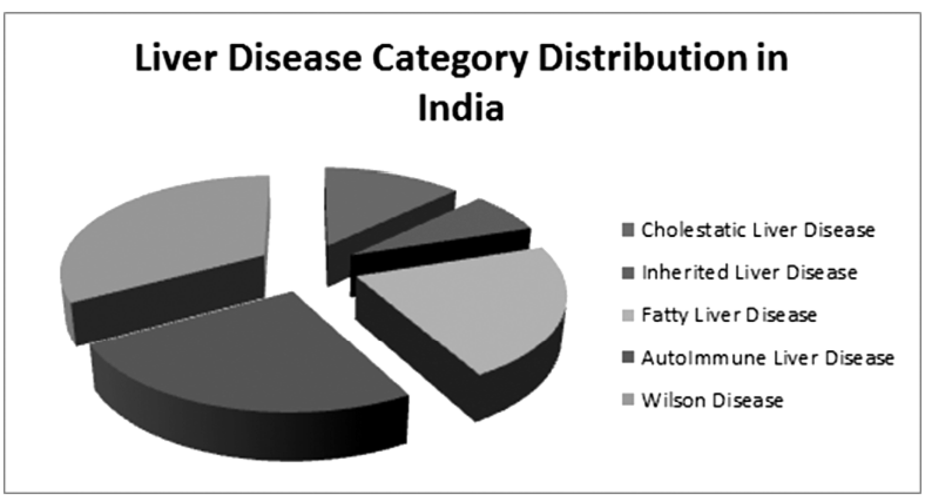

Figure 2: Distribution of Liver Based Diseases in India

Our proposed work constitutes implementation of Classification Rule Mining algorithms to in optimizing various risks of liver disorders. The dataset detail is presented in table 2. Three different algorithms are sued in our study which include Decision table, OneR and DTNB. Various performances parameters are used to evaluate the effectiveness of these algorithms. It is observed that DTNB algorithm produces the maximum classification accuracy of 86.48% in precisely predicting liver disorder risks. Our proposed model efficiency is measured in terms of latency. Classification Rule Mining with DTNB algorithm produces the least delay in predicting the disease risks with a meagre 1.12 sec. Error rate is a vital indicator in evaluating the classification model preciseness. The minimum error rate is associated with DTNB with a negligible 0.2315. the Kappa Coefficient metric is evaluated with all three algorithms and it is seen that DTNB has the maximum value of 0.865. the results analysis is demonstrated in figure 3 to figure 6.

Table 2
Liver Disorders Dataset Details

Features	*Datatype*
Neurological	Categorical(high, medium, low, medium/low)
Psychiatric	Categorical(high, medium, low, medium/low)
Pathological	Categorical(high, medium, low, medium/low)
Physical	Categorical(high, medium, low, medium/low)
Cognitive	Categorical(high, medium, low, medium/low)

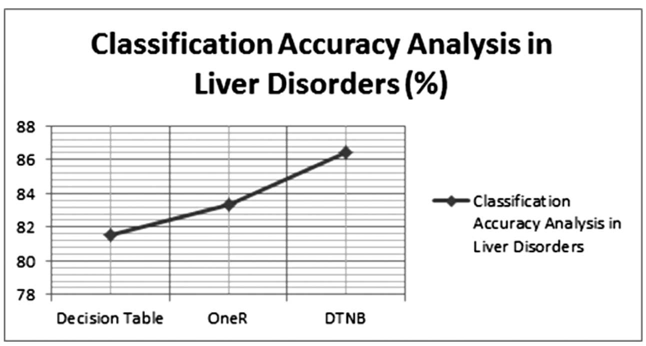

Figure 3: Classification accuracy Analysis in Liver Disorders using Classification Rule Mining Algorithms

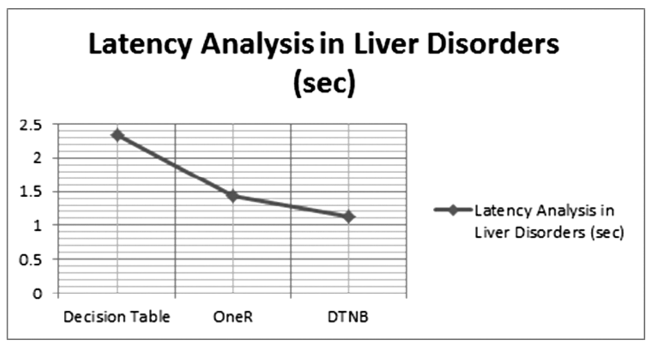

Figure 4: Latency Analysis in Liver Disorders using Classification Rule Mining Algorithms

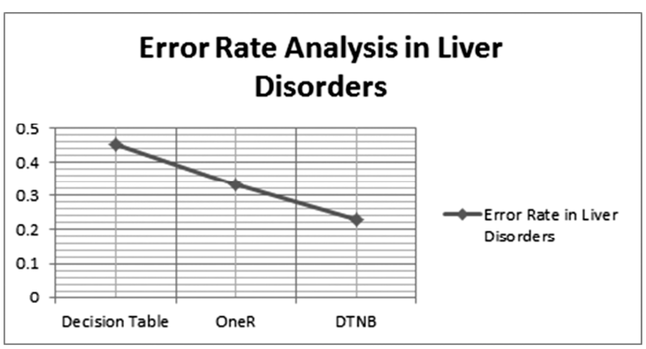

Figure 5: Error Rate Analysis in Liver Disorders using Classification Rule Mining Algorithms

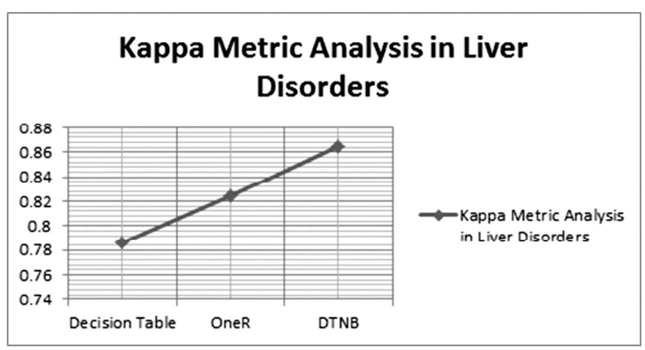

Figure 6: Kappa Metric Analysis in Liver Disorders using Classification Rule Mining Algorithms

6. CONCLUSION

This paper provides an in-depth study of various Classification rule mining algorithms on liver disorders datasets and explores risk detection and effectiveness of such algorithms that offer uncover pattern hidden in data that can help in decision making process. As per the observation it is seen that DTNB algorithm gives the most optimal performance and hence is the most suited algorithm in handling liver hazards. Thus it may be inferred that implementations of DTNB technique are highly acceptable and can help in the clinical domain and also help in decision making process at early diagnosis of any liver disorder.

REFERENCES

[1] JyotiSoni, Ujma Ansari, Dijesh Sharma, SunitaSoni, "Predictive Data Mining for Medical Diagnosis: Overview of Heart Diseases Prediction," IJCST (0975-8887). Vol. 17, No. 8, pp, 43-48, March 2011.

[2] Murat Karabatak, M. CevdatInce, "An expert system for detection of breast cancer based on association rules and neural networks," Elsevier Science Expert System with Application 36, pp. 3465-3469, 2009.

[3] SharifahHafizahSy Ahmad Ubaidillah, RoselinaSallehuddin, NoorfaHaszlinnaMustaffa, "Classification of Liver Cancer Using Artificial Neural network and Support Vector Machine," ElsevierScience Proc. Of Int. Conf on Advance in Communication Network, and Computing, CNC, pp. 488-493, 2014.

[4] Lam, Yee Hong Brian, "Proteomic Classification of Liver Cancer using Artificial Neural Network," May, 2005.

[5] Jung Hun Oh and Jean Gao, "Fast Kernel Discriminant Analysis for Classification of Liver Cancer Mass Spectra," IEEE/ ACM Transactions on Computational Biology and Bioinformatics, Vol. 8. NO. 6. pp. 1522-1534, Nov/Dec 2011.

[6] P. Thangaraju, R. Mehala, "Novel Classification Based approaches over Cancer Diseases," IJARCCE, Vol. 4, Issue 3, pp. 294-297, March 2015.

[7] M. Neshat, M. Yaghobi, M.B. Naghibi, A. Esmaelzadeh, "Fuzzy Expert System Design for Diagnosis of Liver Disorders," IEEE International Symposium on Knowledge Acquisition and Modeling, pp. 252-256, 2008.

[8] Zahra Beheshti. SitiMariyamHj. Shamsuddin. EbrahimBeheshti. SitiSophiayatiYuhaniz, "Enhancement of artificial neural network learning using centripetal accelerated particle swarm optimization for medical diseases diagnosis," Springer Science, Vol. 18, pp. 2253-2270, 15 December 2013.

[9] RasoolFakoor, Faisal Ladhak, Azade Nazi, Manfred Huber, "Using deep learning to enhance cancer diagnosis and classification," Proceeding of the 30thInternational Conference on Machine Learning, Atlanta, Georgia USA, Vol. 28, 2013.

[10] Kenneth R Foster, Robert Koprowski and Joseph D Skufca, "Machine learning, Medical diagnosis, and biomedical engineering research-commentary," Biomedical engineering online 2014.

[11] SushmitaMitra, "Fuzzy MLP based expert system for medical diagnosis," Elsevier Science Fuzzy Sets and System 65, pp. 285-296, 1994.

[12] I-Jen Chiang, Ming-Jium Shiem, Jane Yung-Jen Hsu, Jau-Min Wong, "Building a Medical Decision Support System for Screening by Using Fuzzy Classification Trees," Springer Science Applied Intelligence 22, pp. 61-75, 2005.

[13] Varun Kumar, Luxmi Verma, "Binary Classifier for Heath Care Databases: A Comparative study of Data mining Classification Algorithms in the Diagnosis of Breast Cancer," IJCST Vol. 1. Issue 2, pp. 124-129, December 2010.

[14] Rong-Ho Lin, Chun-Ling Chuang, "A hybrid diagnosis model for determining the types of the Liver Diseases," Elsevier Science Computers in Biology and Medicine 40, pp, 665-670, 2010.

[15] Sammouda.et.al (1999) "Segmentation and Analysis of Liver Cancer Pathological Color Images based on Artificial Neural Networks", IEEE.

[16] Upadhyay & Wasson (2014) "Application of Genetic Algorithm for Liver Cancer Detection", International Journal of Research in Engineering & Technology, volume 03, Issue 05.

[17] Selvaraj & S., (2013) "Improved Feature Selection Based on Particle Swarm Optimization for Liver Disease Diagnosis", Springer International Publishing Switzerland, Part II, pp 214-225.

[18] Massieh.et.al, (2010) "A Novel Fully Automatic Technique for Liver Tumor Segmentation from CT Scans with Knowledge-based constraints", IEEE.