Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INTELLIGENT –AGENT BASED FRAMEWORK FOR LIVER DISORDER DIAGNOSIS USING ARTIFICIAL INTELLIGENCE TECHNIQUES

[1]H.RATNAMALA KIRUBA,, [2]Dr.G.THOLKAPPIA ARASU

[1] Assocprof., Department Of Computer Science And Engineering, Avs Engg College, Salem

[2] Principal.,Avs Engg College, Salem

E-mail: [1]rmkavs@gmail.com [2]tholsg@gmail.com

ABSTRACT

Artificial intelligence is an emerging area of modern research that aims at infusing machine intelligence through computational techniques. Data mining (DM) enables efficient knowledge extraction from large datasets, in order to discover hidden or non-obvious patterns in data. Our motivation for using DM was based on the hypothesis that the application of the appropriate DM technique on patient records could form a suitable mechanism for the knowledge extraction representing the correlation between patient symptoms and disease. The extracted knowledge was then used for the provision of personalised recommendations to patients in collaboration with the agent-based framework developed. The agent – based system developed interacts with different modules of the overall integrated system developed to support liver disease diagnostic system. This research work aims at exploring the impact of machine learning techniques in liver disorder detection on two different datasets comprising of more than 900 patient records acquired from the University of California, Irvine, Machine Learning Repository . The findings revealed that C4.5 decision tree algorithm and the Random Tree algorithm produced 100 percent accuracy in classification of the liver disorders and we believe implementation of the proposed intelligent agent-based system will raise a precise and accurate diagnostic system for clinical ailments of diverse kind. To the best of our knowledge, this is the first attempt to explore this large collection of supervised machine learning techniques in the design of intelligent agent-based clinical systems for diagnostic purposes.

Keywords: *Artificial intelligence, Intelligent Agents, Supervised learning, Data mining, Clinical diagnosis, Liver Disorder detection*

1. INTRODUCTION

Applications of Artificial intelligence is an intense area of modern-day research holding sway over diverse application fields that include finance, robotics and medicine, to name a few [1]. Mining clinical and biological data to unearth significant patterns and solutions has led to the emergence of varied spheres in contemporary research [1] [2]. Utilization of computational science and techniques to instill intelligence in technical gadgets has led to profound research and analysis in recent years [3]. Data mining is a related area of research that aims at discovering important patterns from an exhaustive collection of raw data and ascertained facts [4]. Application of artificial intelligence and data mining techniques to medical and biological data has given rise to extensive research in Bioinformatics that involve analysis of diverse variety of data whose scope extends from DNA sequences to amino acid substitutions and micro arrays[5][6]. We restrict our research to the diagnosis of liver disorders in this paper.

Liver cancer is stated to be one of the leading causes of death around the world [7]. According to the latest statistics reported in February 2012, approximately 2.7 to 3.9 million people in the U.S. are chronically infected with the hepatitis C virus [8]. Approximately 12,000 people die of hepatitis C annually in the U.S. and 800,000 to 1.4 million people in the US are infected with the Hepatitis B virus[7][8]. In India statistics based on the liver disease patients reveal the fact that, roughly 40% are suffering from hepatitis B and C, and about 60% are alcohol addicts [9]. Enlarged liver, Hepatitis B infection, Hepatitis C infection and subsequent Liver Cirrhosis are precedents to developing Hepatocellular Carcinoma (HCC/Liver Cancer) in a number of afflicted cases[7][8][9]. It has also been stated from previous studies and surveys that this

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

oncogenic ailment occurs as a result of chronic liver disorders over a prolonged period of time [9]. The vast segment of the population that is succumbing to this oncoailment, supplemented by a sizeable population suffering from liver diseases as a result of their irresponsible lifestyle and harmful addictions has been the rationale for this research.

This article places focus on designing an efficient classifier for diagnosing liver diseases by evaluating twenty two supervised machine learning techniques on the two liver cancer datasets namely the BUPA dataset comprising of 345 patient records and the Indian Liver Patient dataset(ILPD) comprising of 583 patient cases. A novel framework based on data mining techniques is proposed to design an efficient diagnostic system for liver diseases. The performance of the supervised learning techniques is evaluated on the basis of training accuracy and computational complexity.

1.1 Paper Organization

Section 2 presents a brief survey on related research and analysis in the field of artificial intelligence. Section 3 describes the proposed framework for classifier design while Section 4 elaborates on the algorithms for classification. Section 5 discusses the experimental results. Section 6 discusses the pros and cons of the proposed approach along with a few directions for future research while Section 7 concludes the paper.

2. BACKGROUND OF THE STUDY

Previous research in the field of artificial intelligence techniques in the arena of medicine and biology are narrated below.

Padgham and Winikoff (2004) define an agent as being “a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives”. Because agents are situated in an environment, they are instantiated once and continue to run in memory until stopped. In the last years, some proposals for intelligent and agent-based decision support systems (e.g.Kebair & Serin,2006; Liu, Qian, & Song, 2006; Ossowski et al., 2004; Petrov & Stoyen, 2000; Urbani & Delhom, 2005) have been described. New approaches of researching intelligent decision support system (IDSS) appear following the rapid progress of agent systems and network technology. Thus, a large range of works dedicated to environment and human health implemented as multi-agent systems (MAS), which are in the center of active research for more than ten years and resulted in many successful applications, have emerged. The application of data mining (DM) techniques for environmental monitoring, medicine, social issues is also an integrated topic, the survey of which is described below.

Recent work on liver disease diagnosis was done by Ribeiro et.al, [10] who addressed the issue of identification and diagnosis of various stages of chronic liver disease. The classification results of a support vector machine, a decision tree and a k-nearest neighbour classifier were compared in the work. Ultrasound image intensity and textural features were jointly used with clinical and laboratorial data in the staging process. The training phase of the classifier was done using a population of 97 patients at six different stages of chronic liver disease and a leave-one-out cross-validation strategy. The best results were obtained using the support vector machine with a radial-basis kernel, giving an overall accuracy of 73.2%. The good performance of the method seemed a promising indicator for usage, in a non invasive way, to provide reliable information about the chronic liver disease staging. However this work was limited in the number of training records included for study and the classification algorithms explored.

S.Karthik et.al, [11] reported their findings on application of soft computing techniques for intelligent diagnosis of liver disease. The classification and its type detection were implemented in three phases. In the first phase, ANN classification was applied to classify the liver disease records. In the second phase rough set rule induction using Learnable Evolution Model (LEM) algorithm was executed to generate classification rules. The rule induction approach overcame the drawback of Multi Layer Perceptron (MLP) and raised the classification accuracy by 6%. In the following phase, fuzzy rules were utilized to identify the diverse nature of liver diseases. Using LEM algorithm 6 rules were obtained with an accuracy of 96% in classification. On applying LEM rules, enhanced classification accuracy of 6% was obtained compared to MLP. Four fuzzy rules were formulated to indentify the types of liver diseases. The work reported in our paper affirms higher classification accuracy with supervised machine learning techniques reporting an accuracy of 100% in liver disorder diagnosis. In 2012, Barnaghi et.al, [12] discussed data mining techniques to process a medical dataset and identify the relevance of liver disorder and alcohol

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

consumption by classification of blood test data. They analyzed four classification methods viz, decision tree, Bayesian algorithms (Naive Bayes and Bayesian Networks), Neural Network classification and Rough Sets methods. They utilized the open source data mining suite named Waikato Environment for Knowledge Analysis (WEKA) for their experimental analysis with classification algorithms and Rosetta software for exploring Rough Sets. Their work identified as a performance indicator only the classification accuracy to grade the classification algorithms. The evaluation results revealed that using Neural Networks best classification results were obtained. However the accuracy obtained on different training sets was not as expected ranging from 53% to 77% pertaining to the approaches analyzed. Ramana et.al, [13] previously investigated the Bayesian classification technique with Bagging and Boosting for Liver disease diagnosis. 751 patient details were used in this experimentation. Patients were chosen from Andhra Pradesh state of India. The features considered were Gender, Age, Total bilirubin, Direct bilirubin, Indirect bilirubin, Total proteins, Albumin, Globulin, A/G ratio, SGOT, SGPT and ALP. Samples were labelled by a Gastroenterologist, who classified patients into five groups. Bayesian classifier was able to accurately classify 94% of male patients in the age group of 16-40 and 98% of female patients aged over 40. Moreover they claimed and proved with necessary findings that Bagging and Boosting methods enhanced the classification accuracy of Bayesian Classifiers.

Further Ramana et.al, [14] also proposed a Modified Rotation Forest algorithm for accurate liver classification by analyzing the combination of selected classification algorithms and feature selection techniques. The category of classification algorithms chosen were Tree based, Statistical based, Neural Networks based, Rule based and Lazy learners. Modified Rotation Forest algorithm for UCI liver data set utilized the Multilayer Perceptron (MP)classification algorithm and Random Subset feature selection technique and to classify the INDIA Liver Data Set (ILPD), the Nearest - Neighbor with generalized distance function and correlation based feature selection technique were analyzed. The highest obtained classification accuracy was produced by the Multilayer Perceptron algorithm whose accuracy was stated to be ~75% for the BUPA data and the KStar classification algorithm was reported to classify with ~73% accuracy, the records of ILPD. The results reported and the approaches investigated are reduced in value than the classification accuracy and techniques reported in our work for both the datasets. Tomar et.al, [15] utilized a Case Based Reasoning methodology to develop a clinical decision support system prototype for supporting diagnosis of occupational lung diseases. 127 cases were collected for 14 occupational chronic lung diseases that contained 26 symptoms. After removing the duplicated cases from the database, the system was trained on 47 cases for Indian Lung patients. The retrieval strategy using Nearest-Neighbor approaches was investigated. The results suggested that the Nearest Neighbor approach as a retrieval strategy showed potentially good results. A Consultant Pathologist's interpretation was used to evaluate the system. Results for Sensitivity, Specificity, Positive Prediction Value and the Negative Prediction Value were stated to be 95.3%, 92.7%, 98.6% and 81.2% respectively. A clinical decision support system prototype was developed for supporting diagnosis of occupational lung diseases from their symptoms and signs through employing Microsoft Visual Basic .NET 2005 along with Microsoft SQL server 2005 environment with the advantage of Object Oriented Programming technology.

Braaten et.al, [16] investigated the issue of whether artificial intelligence methods could represent objective methods that were essential in syndrome diagnosis. They applied two basic artificial intelligence methods to a database of machine-generated patients - a 'vector method' and a set method. As reference methods they executed the ID3 algorithm, a cluster analysis and a naïve Bayes' calculation on the same patient series. The overall diagnostic error rate for the vector algorithm was 0.93%, and for the ID3 0.97%. For the clinical signs found by the set method, the predictive values varied between 0.71 and 1.0. The artificial intelligence methods that were used, were stated to be simple, robust and powerful, and represented objective diagnostic methods.

It is evident from the review of past research in this field that the data records considered for analysis and the algorithms investigated for prediction and diagnosis are limited in number. This paper addresses both the issues by using a training set of more than 900 records and exploring over 22 machine learning approaches in a single study. In the following section, we propose a methodology to analyze and classify clinical patient records through artificial intelligence techniques.

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. INTELLIGENT AGENT BASED CLINICAL SYSTEM DESIGN

Classification [17] is the process of assigning the class of a data record whose class label is previously unknown. We have investigated the working of twenty-two supervised machine learning techniques to formulate an efficient classifier for diagnosing liver diseases. The methodology adopted comprises of three phases viz, Training the classifier with the clinical data, evaluating the supervised machine learning techniques for accurate classification and Verification of classification accuracy with a test patient case. The following sections elaborate on the experimental setup for classifier design.

3.1 Materials

We have utilized the Liver Disease datasets from the UCI Irvine Machine Learning Repository to train the classifier system. The description of the datasets is given in Table 1 and Table 2. The former data was obtained from BUPA Medical Research Ltd [18]. The BUPA liver dataset comprises of 345 patient records while the Indian Liver Patient Dataset (ILPD) [18] contains 583 patient cases.

Table 1. Description of the Bupa Liver Dataset

Attributes	Description
MCV	Mean Corpuscular Volume
AAP	Alkhos Alkaline Phosphotase
SGPT	SGPT Alamine Aminotransferase
SGOT	SGOT Aspartate Aminotransferase
GGT	Gamma Glutamyl Transpeptidase
Drinks	Number of half-pint equivalents of alcoholic beverages drunk per day
Class	Class –Target (Liver Patient, Healthy)

The Indian Liver Patient Data (ILPD) contains 416 liver patient records and 167 non liver patient records. The data set was collected from north east of Andhra Pradesh, India. This data set contains 441 male patient records and 142 female patient records.

Table 2. Description of the ILPD

S.No	Attributes	Description
1.	Age	Age of the patient
2.	Gender	Gender of the patient
3.	TB	Total Bilirubin
4.	DB	Direct Bilirubin
5.	AAP	Alkhos Alkaline Phosphotase
6.	SAA1	SGPT Alamine Aminotransferase
7.	SAA2	SGOT Aspartate Aminotransferase
8.	TP	Total proteins
9.	ALB	Albumin
10.	AGR	Albumin Globulin Ratio
11.	Class	Class –Target (Liver Patient, Healthy)

The methodology to explore, execute and analyse the supervised machine learning techniques is detailed in the following sub-sections.

3.2 Experimental Design Framework

The training data needs to be pre-processed prior to execution on the open source data mining suite. The attributes are taken to be the column headers and the corresponding values of each attribute are imported into the Excel spreadsheet delimited by comma. The agents are designed using the JADE environment wherein three types of agents are required for the design of the clinical system.

- DSA – Diagnostic support agents aid in the process of cleaning and processing data thereby making the data suitable for classification and diagnosis.
- DA – Decision Agents assist in choosing the best classifier for building the clinical classifier for liver diagnosis.
- LA – Learning Agent forms the basis of the intelligent system whereby it learns the classification rules acquired from the training data to build the model.

The design framework for the classifier system design is portrayed in Fig 1. The dependent variables include the target class while the independent variables refer to the specific attributes/features that control the target class. There as such exist no control variables for the given experiment design.

The pre-processed data is then loaded onto the data mining suite and the predictor and target features are defined. Each of the datasets is analyzed individually and the supervised machine learning techniques are executed on the datasets and their training accuracy are recorded for further evaluation.

3.2 Evaluation of Learning Algorithms

The classification algorithms are evaluated based on the training accuracy. The decision tree size is recorded for each algorithm that predicts the class as a classification tree. The size of the decision tree is specified by the number of nodes. The algorithm which predicts the precise class for all the training records for both the datasets is taken to be the most efficient classification algorithm. The performance

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parameters considered for evaluation include the classification accuracy, decision Tree size, and the computation time taken for execution.

3.3 Verification of Classification Accuracy

The prediction rules generated by the classification algorithm are stored as a Knowledge Base. A new, previously unseen liver patient record is input to the Knowledge base and the class of the record is determined accurately.

4. SUPERVISED MACHINE LEARNING ALGORITHMS

Supervised learning [17] is the machine learning task of deducing a function from supervised (labelled) training data. The training data consists of a set of training instances and each instance is a pair consisting of an input value and a desired output value [3]. A supervised learning algorithm analyzes the training data and generates a conclusion. This is called a classifier [4]. The inferred function should predict the correct output value for any valid input record [17]. This requires the learning algorithm to generalize from the training data to previously unknown situations [5]. The Quinlan's C4.5 decision tree algorithm and the Random Tree algorithm classify the training records with 100 % accuracy with a reasonable and acceptable classification tree size and computation time. We discuss the operation of the two classification algorithms in the following sub-sections.

4.1 Quinlan's C4.5 Decision Tree Algorithm

C4.5 builds decision trees from a set of training data according to the algorithm depicted in Fig.2, using the concept of information entropy [20]. The training data is a set S=S1,S2….. of samples whose class label is known. Each sample pertains to the attributes or features of the sample. The training data is augmented with a vector C=C1, C2 … each of which relate to the class categorizing each sample [21]. The algorithm [22] is presented below.

Input: Collection of training instances, Category, Predictor attributes (Predictors)
Output: Decision Tree, Target Class of test record
Algorithm:
C4.5 (Instances, Category, Predictors_list)
Step 1: Generate a root node
Step 2: If all instances are Liver patients,
Step 2.1: Return the tree with the root node as Liver Patient
Step 3: Else If all instances are Healthy,
Step 3.1: Return the tree with the root node as Healthy
Step 4: Else If 'Predictors_list' is Null
Step 4.1: Return the root node with the most common value of the Category in the training instances
Step 5: Else for each predictor
Step 5.1: Calculate Information Entropy

$$E(S) = -\sum_{j=1}^{n} P_s(V) log_2 P_s(V)$$

/ n-Number of values that the predictor holds in Set 'S'*
*P_s – Proportion of value 'V' in 'S' */*

Step 5.2: Determine Information Gain

$$G(S,A) = E(S) - \sum_{i=1}^{m} P_s(A_i)\, E(S_{A_i})$$

/ G(S,A) – Gain of set S when split over predictor A*
'm'-Number of different values of predictor 'A' in S
Ps(Ai)is the proportion of items possessing Ai a value for A in S
S_{A_i} is a subset of 'S' containing all items where the value of A is Ai
**/*
Step 5.3: Choose Attribute that renders higher Information Gain
Step 5.4: Set predictor for Root = A
Step 5.5: For each existing value Val_i
Step 5.5.1: Add a new tree branch below Root, corresponding to the test A = Val_i
Step 5.5.2: Let Instances (Val_i) be the subset of records that have the value 'Val_i' for A
Step 5.5.3: If Instances (Val_i) is empty
Step 5.5.3.1: Add a leaf node with label = most common target value in the instances
Step 5.5.4: Else call C4.5 (Instances (Val_i), Category, Predictors_list –{A}
End For
End For

Step 6: Return Root

At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets contained in one class or the other[21][22]. Its criterion is the normalized information gain that decides on the predictor feature for partitioning the subsets. The attribute with the highest normalized information gain is chosen to make the decision[23]. The C4.5 algorithm then iterates on the smaller subsets.

4.2 Random Tree Algorithm

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The Random Tree algorithm was first introduced by Leo Brieman and Adele Cuttler [24]. The algorithm is presented for classification of the BUPA and ILPD records addressed in this paper. The algorithm in the data mining suite works for categorical and continuous –valued attributes.

We present the algorithm [25] for liver disorder diagnosis as follows.

Input: LD, the training Liver Data set, A, the set of attributes.
Output: A random decision tree RLDT
Algorithm:

// Main program
RLDT = Craft RLDT(A)
ReviseParameters(RLDT, LD)
Remove subtrees classifiying 0 records
return RLDT
//Subroutines
Subroutine Craft RLDT(A)
if A = 'Null' then
return a leaf node
else
Randomly select an attribute T as test attribute
Form an internal node 'N' with T as the attribute
Assume T has 'v' valid values
For i = 1 to v Do
$Tree_i$ = Craft RLDT(A − {T})
Append $Tree_i$ as a child of N
End for
end if
return N

Subroutine ReviseParameters(N, LD)
for each entry 'e ' in LD Do
AppendInstance(N,e)
end for

Subroutine AppendInstance(N,e)
if N is not a leaf node then
Let T be the attribute in N
Let ch represent the chilld of N that corresponds to the value of T in e
AppendInstance(ch, e)
else
/* N is the leaf node */
Let lab be the label of e
Let α[lab] = # of lab-labeled rows that reach N
α[lab] ← α[lab] + 1
end if

In this methodology, the attribute for test at each iteration is chosen randomly. The algorithm recursively creates the structure of the tree using the Craft RLDT procedure, and then revises the parameters on each iteration using the method Revise Parameters(RLDT,LD) at the leaves by eliminating each training instance through the tree. Each leaf node of the tree holds 'Number' counters, α[1], . . . , α[Number], where Number is the number of possible labels for training instances. Once all the instances are incorporated into the tree, the algorithm prunes away all internal and leaf nodes that did not contribute to classification [24][25]. The running time of the algorithm is linear in the size of the database.

5. PERFORMANCE EVALUATION

The performance parameters chosen to rank the classification algorithms in this paper include the classification accuracy [17] [23], the number of nodes required for classification and the computation time. Section 5.1 briefs about the performance indicators while Section 5.2 presents the obtained results.

5.1 Performance Parameters

Accuracy [17] is defined as the number of input records that are correctly assigned to the target class/ category. This is obtained from the misclassification rate as Acc = [1- Misclassification Rate]. The decision tree size [3] [19] refers to the total number of nodes in the generated tree that needs to be traversed in order to detect the class of a new, previously unknown data. Computation time [19] refers to the time taken to build the classification model for liver disorder diagnosis. Foremost importance is assigned to the classification accuracy following which the latter performance parameters are computed for the classification algorithms that reveal highest accuracy.

5.2 Experimental Results

The accuracy of the classification algorithms for the BUPA and the ILPD is depicted in a sorted manner in tabular form in Table 3. The algorithms showing 100% accurate classification are recorded as the C4.5 Decision tree algorithm and the Random Tree algorithm.

Table 3. Classification Accuracy of Supervised Machine Learning Techniques

Classification Algorithm	Acronym	Accuracy %	
		BUPA	ILPD
C4.5 Decision Tree Algorithm	**C4.5**	**100**	**100**
Random Tree	**RT**	**100**	**100**
K-Nearest Neighbour	K-NN	80.87	79.42

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Classification with Least Misclassification	CS-MC4	77.39	71.7
Multilayer Perceptron	MP	76.23	72.21
Core Vector Machine VM	CVM	73.62	71.36
Ball Vector Machine	BVM	72.46	71.7
Logistic Regression	LR	70.72	73.24
Binary Logistic Regression	BLR	70.43	73.58
Multinomial Logistic Regression	MLR	70.43	73.58
Linear Discriminant Analysis	LDA	70.43	71.18
Support Vector Machine for Classification	C-SVC	67.54	71.36
Classification-Tree	C-RT	67.25	71.36
Cost-Sensitive Classification Tree	CS-CRT	67.25	71.36
Prototype-Nearest Neighbor	P-NN	66.96	67.75
Partial Least Squares – Discriminant Analysis	PLS-DA	65.22	71.53
Partial Least Squares-Linear Discriminant Analysis	PLS-LDA	64.93	71.53
Partial Least Squares for classification	C-PLS	63.77	59.69
Iterative Dichomotiser 3	ID3	63.19	71.36
Naïve Bayes Classification	NBC	62.9	69.13
Radial Basis Function	RBF	51.3	71.36
Support Vector Machine	SVM	58.26	71.36

The parameters indicating the number of nodes in the classification tree generated by the accurately classifying algorithms is tabulated in Table. 4

Table 4. Comparison of Decision Tree Size

Classification Algorithms	Acronym	Decision Tree Size (No. of nodes)	
		BUPA	ILPD
C4.5	C4.5	173	299
Random Tree	RT	173	299

The computational complexity of any algorithm lies in the amount of space needed for execution and the time elapsed for result detection. The computation time taken to classify the training data and formulate the classification model for clinical record categorization is portrayed in Table 5.

Table 5. Computation Time Comparison for the Accurate Classification Techniques

Classification Algorithms	Acronym	Computation time (in ms)	
		BUPA	ILPD
C4.5 Decision Tree Algorithm	C4.5	16	62
Random Tree Algorithm	RT	32	63

The graphical representation of the comparative performance analysis of classification algorithms is given in Fig.3.

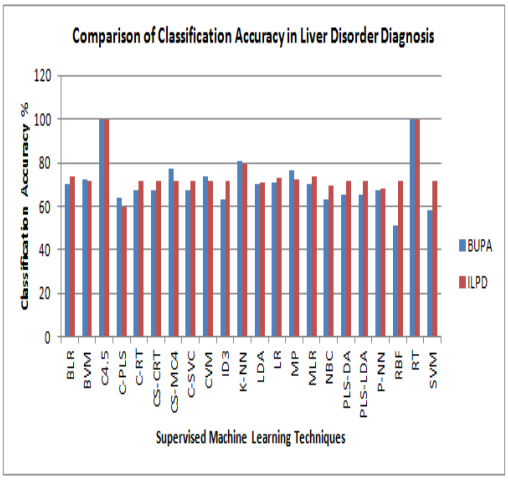

Fig. 3 Graphical Representation of Classification Accuracy

The K-Nearest Neighbor algorithm follows the C4.5 and Random Tree algorithm in ranking, yielding a classification accuracy of ~80% on the BUPA data and ~79% on the ILPD. Further investigation would require implementation of a classier system to diagnose liver disorders utilizing the framework with a knowledge base trained on the generated classification rules and training clinical records.

6. DISCUSSSION

It is evident from the experimental results that the proposed agent-based framework is successful in predicting the liver disorders from the given set of independent variables. However, the cons of the proposed approach include the fact that only patients who have already acquired the disease can be identified. It is not possible to predict the possibility of occurrence or the stages of disease progression. Moreover the proposed approach has not revealed the important features that could be further explored to analyze their contribution to the cause of the liver disease. Hence an investigation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

on available feature selection methods would be a promising area of future research.

7. CONCLUSION

The impact of computational science and techniques in medical investigations has been the theme of intense research in recent years. This paper places emphasis on the manner in which artificial intelligence infused in a machine through machine learning and data mining techniques can result in minimizing human labor involved in clinical data management, analysis and prediction. We have proposed a novel framework based on Artificial intelligence and data mining techniques to design an efficient classifier for liver diseases. Clinical data and facts have to be analyzed with utmost caution as lapses; great or small could be fatal. We report and affirm with necessary findings the performance of machine learning and data mining techniques that give acceptable results in data analysis and class prediction as a case study with the liver disease datasets. We believe that similar performance will be exhibited by the computational techniques on clinical facts of diverse nature. Moreover an investigation into the possibility of applying agent based artificial intelligence techniques to detection stages of disease progression and significant symptoms of disease occurrence is an important area of future research.

REFRENCES:

[1] Buchanan, Bruce G. "A (Very) Brief History of Artificial Intelligence" (PDF). AI Magazine: 53–60, 2005.

[2] Nilsson, Nils, "Artificial Intelligence Prepares for 2001", AI Magazine 1 (1), http://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag04-04-002.pdf, Presidential Address to the Association for the Advancement of Artificial Intelligence, 1983.

[3] Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell, Machine Learning: An Artificial Intelligence Approach, Tioga Publishing Company, ISBN 0-935382-05-4, 1983.

[4] Jacob S.G, Dr. Ramani, R.G,"Discovery of Knowledge Patterns in Clinical Data through Data Mining Algorithms: Multi-class Categorization of Breast Tissue Data", International Journal of Computer Applications (IJCA), 32(7): 46-53, 2011, DOI: 10.5120/3920-5521. Published by Foundation of Computer Science, New York, USA.

[5] Jacob S.G, Dr. Ramani.R.G, Nancy .P, "Efficient Classifier for Classification of Hepatitis C Virus Clinical Data through Data Mining Algorithms and Techniques", Proceedings of the International Conference on Computer Applications, Pondicherry, India, January 27-31, 2012,Techno Forum Group, 2012, India. ISBN: 978-81-920575-8-3: DOI: 10.73445/ISBN_0768, ACM#.dber.imera.10.73445.

[6] Ralf Mikut; Markus Reischl, "Data Mining Tools". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 (5): 431–445, 2011.

[7]http://info.cancerresearchuk.org/cancerstats/types/liver/uk-liver-cancer-statistics

[8]http://www.cancer.net/patient/Cancer+Types/Liver+Cancer?sectionTitle=Statistics

[9] http://www.cancerfoundationofindia.org/

[10] Ribeiro, R., Marinho, R., Velosa, J., Ramalho, F., Sanches, J.M.: Chronic liver disease staging classification based on ultrasound, clinical and laboratorial data. ;In ISBI(2011)707-710

[11] S. Karthik, A. Priyadarishini, J. Anuradha and B. K. Tripathy, Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types, Advances in Applied Science Research, 2011, 2 (3): 334-345

[12] Peiman Mamani Barnaghi, Vahid Alizadeh Sahzabi and Azuraliza Abu Bakar, A Comparative Study for Various Methods of Classification, ICICN 2012, pp 62-67.

[13] Bendi. Venkata Ramana et al. , A Critical Evaluation Of Bayesian Classifier For Liver Diagnosis Using Bagging And Boosting Methods, International Journal of Engineering Science and Technology (IJEST), ISSN: 0975-5462 Vol. 3 No. 4 April 2011 3422-3426.

[14] Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu, Liver Classification Using Modified Rotation Forest, International Journal of Engineering Research and Development, ISSN: 2278-067X, Volume 1, Issue 6, June 2012, PP.17-24

[15] Prem Pal Singh Tomar, Dr. Ranjit Singh, Dr. P K Saxena & Jeetu Sharma, Case Based Medical Diagnosis of Occupational Chronic Lung Diseases From Their Symptoms and Signs", International Journal of Biometrics and Bioinformatics (IJBB), Volume (5) : Issue (4) : 2011.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[16] Øivind Braaten* and Johannes Friestad, Syndrome Diagnosis: Human Intuition or Machine Intelligence? Open Med Inform J. 2008; 2: 149–159. Published online 2008 November 19. doi: 10.2174/1874431100802010149

[17] J. Han and M. Kamber , Data Mining; Concepts and Techniques, Morgan Kaufmann Publishers, 2000.

[18] UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA Datasets from UCI.

[19] Tanagra Data Mining tutorials, http://data-mining-tutorials.blogspot.com/

[20] Ron Kohavi and Ross Quinlan, Decision Tree Discovery, 2009.

[21] Thales Sehn Korting, C4.5 algorithm and Multivariate Decision Trees, Image Processing Division, National Institute for Space Research – INPES˜ao José dos Campos – SP, Brazil, 2006.

[22] S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica 31,2007, 249-268.

[23] J. R. Quinlan. Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Research, 4:77-90, 1996.

[24] Leo Breiman, Adele Cuttler, Random Trees, http://www.stat.berkeley.edu/users/breiman/RandomForests/

[25] Geetha Jagannathan, Krishnan Pillaipakkamnatt, Rebecca N. Wright, A Practical Differentially Private Random Decision Tree Classifier, Proceedings of the ICDM International Workshop on the Privacy Aspects of Data Mining, 2009.

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANNEXURE : FIGURES

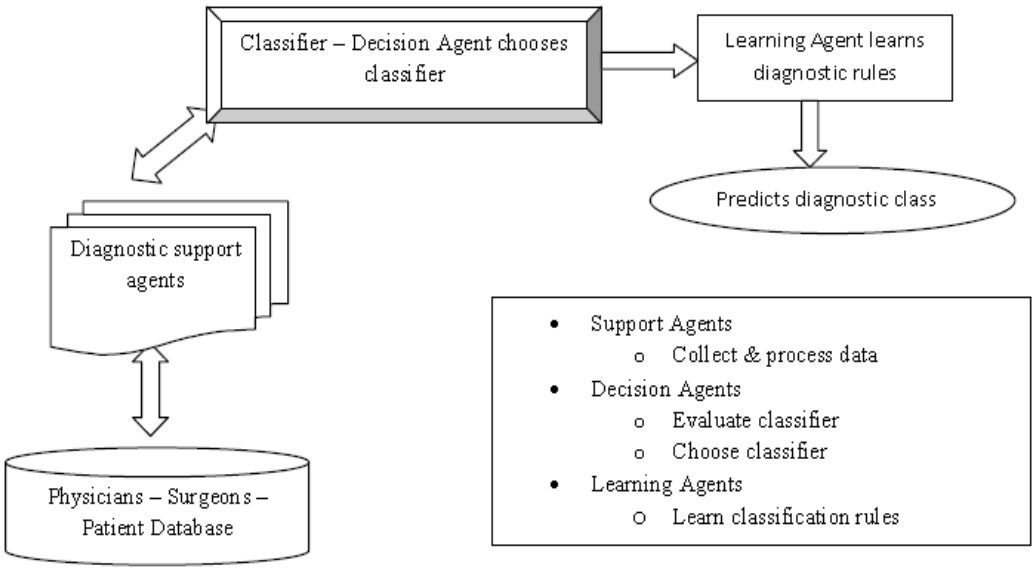

Fig 1. Role of Agents in Design of Clinical Classifier

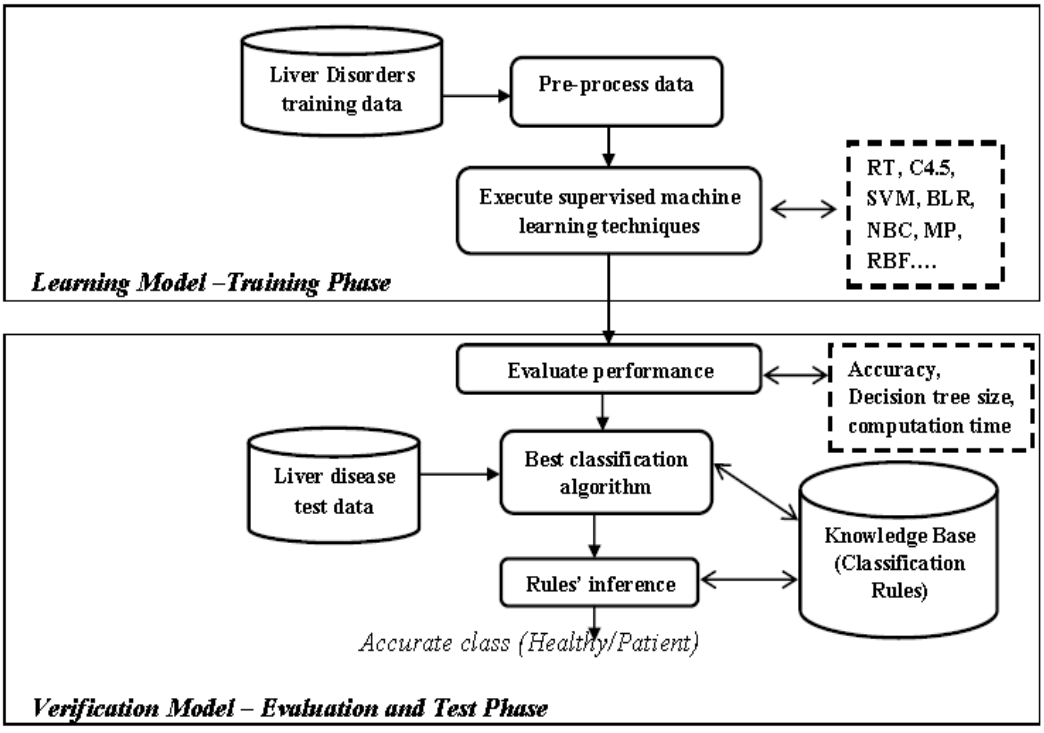

Fig.2. Formulation of the Liver Disorder Classifier System