I.J. Information Engineering and Electronic Business, **2017, 4, 38-46**
Published Online July 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2017.04.06

Data Mining Based Hybrid Intelligent System for Medical Application

Adane Nega
Faculty of Computing, Bahir Dar University, Ethiopia
Email: nega2002@gmail.com

Alemu Kumlachew
Faculty of Computing, Bahir Dar University, Ethiopia
Email: alemupilatose@gmail.com

***Abstract*—**Hybrid intelligent system is a combination of artificial intelligence (AI) techniques that can be applied in healthcare to solve complex medical problems. Case-based reasoning (CBR) and rule based reasoning (RBR) are the two more popular AI techniques which can be easily combined. Both techniques deal with medical data and domain knowledge in diagnosing patient conditions. This paper proposes a hybrid intelligent system that uses data mining technique as a tool for knowledge acquisition process. Data Mining solves the knowledge acquisition problem of rule based reasoning by supplying extracted knowledge to rule based reasoning system. We use WEKA for model construction and evaluation, Java NetBeans for integrating data mining results with rule based reasoning and Prolog for knowledge representation. To select the best model for disease diagnosis, four experiments were carried out using J48, BFTree, JRIP and PART. The PART classification algorithm is selected as best classification algorithm and the rules generated from the PART classifier are used for the development of knowledge base of hybrid intelligent system. In this study, the proposed system measured an accuracy of 87.5% and usability of 89.2%.

***Index Terms*—**Hybrid intelligent system, Data mining, Rule based reasoning; Case based reasoning, knowledge acquisition.

I. Introduction

Computer systems have an increasing impact on the practice of medicine. Artificial intelligence (AI) system is one of the promising and ambitious areas of information technology to improve human health and longevity. AI has a number of important medical applications, such as modeling brain functions, speech analysis and synthesis, patient monitoring, medical diagnostic systems, drug dosage administration and health care services [19]. Apart from medical applications, Ioannis and JIM [6] described that AI can be used in the design of intelligent tutoring system, which uses an expert system to make decisions during the teaching process.

In medical diagnosis, AI is a study realized to emulate human intelligence into computer technology that could assist both doctors and patients by providing a laboratory for examination, representation and cataloguing of medical information as well as by devising novel tools to support decision making and research. The increased integration of intelligent AI techniques in everyday medical applications could improve the efficiency of diagnosis and treatments services by supporting patients and healthcare practitioners [20].There are various ways of solving problems using artificial intelligence, the two more popular techniques are case-based reasoning (CBR) and rule based reasoning (RBR). Both techniques deal with medical data and domain knowledge in diagnosing patient conditions as well as recommending suitable treatments for the particular patients. They serve to improve the quality of medical decision-making, increases patient compliance, minimize complications and medical errors [21].Rule based reasoning and case based reasoning systems can be easily combined to form a hybrid intelligent system. In hybrid intelligent systems, each AI technique has its own strengths and weaknesses.RBR depends on basic rules and regulations of the relevant field and the knowledge from the experts. However, CBR is a way of solving new problems that can adapt to conditions without needing the help of experts [22].

The cost and performance of intelligent system depends directly on the quality of acquired knowledge. The traditional approach to knowledge acquisition is time consuming, costly and error prone as it involves a minimum of two expensive people to communicate i.e. the domain expert and the knowledge engineer. Knowledge acquisition is one of the greatest bottlenecks in the development of hybrid intelligent systems. This is due to that the human expert will usually have insufficient knowledge about intelligent system techniques and the expert will find it difficult to describe his knowledge completely and correctly [24].In order to solve the traditional knowledge acquisition problems and to enrich the knowledge base, data mining techniques, more general, knowledge discovery techniques can be integrated with the hybrid intelligent systems. Data mining (DM) is a subfield of machine learning that enables finding interesting knowledge (patterns, models and relationships) in very large databases [25].

This paper presents a hybrid intelligent system that uses data mining technique as a tool for knowledge acquisition. The study focuses on applying data mining algorithms to a medical database of Tuberculosis (TB) and uses the database resulting of the mining process in a hybrid intelligent system that will help in medical diagnosis and treatment. The proposed system continues our previous work [24] by considering an automatic knowledge acquisition approach (i.e. using data mining) in the development of hybrid intelligent system.

Data mining is the extraction of interesting and previously unknown information or patterns from data sources [26]. It can applicable in diverse areas such as biological data analysis [36], financial data analysis [37], and weather forecasting [38] and so on. Data mining is the central point of knowledge discovery in databases (KDD) process and it correspond to the modeling step in the knowledge discovery in databases process. It involves the application of intelligent methods in order to discover new and useful patterns from large volumes of data. Several models for KDD process have been proposed, but the most known is the industrial model - CRISP-DM. Accordingly to this model, KDD is an iterative and interactive process consisting of six steps: business understanding, data understanding, data preparation, modeling, evaluation of the model and deployment [23].

Data mining applications may solve two kinds of problems: prediction and knowledge discovery [11] [25].For each of these problems it is indicated to use some associated methods. For prediction, we may use classification or regression, while for knowledge discovery we may use clustering, association rules, database segmentation, sequence analysis or visualization. A classification rule attempts to predict the value of a discrete dependent variable from various known attributes. One of the most frequently used methods is classification based on decision tree. The decision tree can predict a new data instance, by following a path that starts from the root to a leaf node. One of the advantages of decision trees lies in the fact that they can easily translate into a set of ‘IF –THEN’ rules, easier to understand. Clustering, often referred to as unsupervised learning, involve a process that discovers structures in data without any supervision. As the name clustering implies, unsupervised algorithm is able to discover structures on its own, by exploiting similarities or differences between individual data points on a data set. Association rules mining is an important data mining method that aims to find interesting dependencies in large sets of data items. Interesting associations between data items can lead to information used for decision making.

There are three basic categories of approaches for integrating rule-based reasoning (RBR) with case-based reasoning (CBR). The categorization is based on the importance of each of the two component schemes in the inference process [5].

Rule-Dominant Approach: This approach focuses on the rule-based component and invokes the case-based component only when rules are unable to deal with specialized situations. This augmentation is done by taking the rules as a starting point of problem-solving and then invoking case-based reasoning to handle exceptions to the rules.

Case-dominant approach: Here, the CBR module comes first followed by the RBR module. In this paradigm, the rules play a supportive role to case-based reasoning, useful for instance when the case library contains a limited number of cases.

Balanced approach: Balanced approaches follow a ‘mixed’ paradigm, where the invocation order of the integrated components is not preset and usually during inference one component dynamically calls the other and vice versa.

For this study, the authors use rule-dominant approach. The main reasons for adopting rule-dominant approach as demonstrated by [2] are acceptable accuracy of the inference process, good explanatory ability and the convenient knowledge acquisition process. Moreover, this approach allows the cases and rules to be stored separately in the knowledge base which makes the system a lot easier to be maintained and modified whenever it is needed.

II. Related Works

There are a many researches that have been done in the area of data mining, knowledge based systems and hybrid intelligence systems for improving healthcare services.

Sellppan and Rafiah [30] have developed a prototype intelligent heart disease prediction system using data mining techniques, namely, decision tree, naïve bayes and neural network. The system uses medical profile such as age, sex, blood pressure and blood sugar to predict the likelihood of patients getting a heart disease. Kapil and Durga [12] presented a general hybrid framework that can support design, planning and management of long term medical conditions. They propose a combination of model-based, case based and rule-based reasoning. The model based reasoning is used as separate reasoning method only if the combination of case-based and rule-based methods is unable to suggest a solution.

I.G.L. da Silva et al. [27] proposed an integration of data mining and hybrid expert system. The main goals of this work was to apply knowledge discovery in databases (KDD) to a medical database of breast cancer, so that detection and prediction patterns are discovered, and use the database resulting of the mining process in a hybrid expert system that help medical diagnosis. Tesfamariam and Tibebe [33] have done a research on integrating data mining results with the knowledge based system for diagnosis and treatment of visceral leishmaniasis. The general objective of the study is to investigate the construction of visceral leishmaniasis knowledge based system through integrated knowledge acquisition technique. The authors used the different techniques and tools such as rule based knowledge representation approach, the different classification algorithms, SWI-Prolog 6.4.0 with UTF-8 and Java NetBeans IDE 7.3 with JDK to achieve the objective of the study. Finally, they found that the proposed system could be operational and acceptable if it could be implemented properly. Other similar research works are also presented in [16][28][29][34] that introduces the use of data mining results with the knowledge based system (expert system).

It can be seen from those mentioned literatures that most of the works have been designed up to the stage of knowledge mining and representation without further discussion on the implementation of automatic knowledge based system and integration of other reasoning systems. Our work differs from those mentioned in the literature that this study suggests a model of TB diagnosis and treatment system integrating data mining , rule based reasoning and case based reasoning . The induced rules and the collected cases are independent and their integration results in improved accuracy. Moreover, most of the previous literatures didn't propose a treatment method for clinical activity and also didn't demonstrate user acceptance testing for their proposed integrated system. Thus, in this study an attempt is made to integrate data mining results with the hybrid intelligent system for TB diagnosis and treatment and the user acceptance testing of the final prototype system is done to ensure that whether the proposed system satisfies the requirements of its end-users.

III. Knowledge Acquisition

Knowledge acquisition is one of the major bottlenecks in the stage of knowledge based system development. Two primary approaches to knowledge acquisition are elicitation of knowledge from experts (traditional knowledge acquisition) and Data mining.

In this research, the traditional knowledge acquisition methods and the data mining classification techniques are used for developing the hybrid intelligent system. For the case based reasoning (CBR) module, previously solved TB patient cases are used to build the case base and for the rule based reasoning (RBR) module, the knowledge base is built by applying data mining classification algorithms from large dataset. In addition, the knowledge acquired from existing documents and interviews are used as supplementary knowledge for building the hybrid system.

A) Knowledge Acquisition through Traditional Methods

In this work, the traditional methods (such as interview and document analysis) are used primarily for understanding the basic concepts related to diagnosis, treatment and prognosis of TB disease. More specifically, interviews and document analysis are used to access the general and domain-specific knowledge and to obtain comprehensive example sets. Data mining approach is particularly fruitful in automating the knowledge acquisition task of rule based reasoning system. However, it is a mistake to believe that one can do data mining process without a domain expert. Because at the very least the researchers need an expert to select the training examples and to explain the domain terminology as well as to identify the features of the examples which are likely to be relevant. Therefore, the researchers used the traditional methods to supplement the automatic knowledge acquisition of the integrated (hybrid) system development.

B) Knowledge Acquisition through Data Mining

The development of an efficient intelligent system or knowledge based system involves the development of an efficient knowledge base that has to be complete, coherent and non-redundant[25].Knowledge acquisition is considered to be the most difficult and error- prone task in the development of knowledge-based system due to that knowledge acquisition involves communications between people with completely different backgrounds, human experts and knowledge engineers, who must formulate the concepts, relations and control mechanisms needed for the knowledge based system [24]. Moreover, tacit knowledge is difficult to transfer to another person by means of writing it down. In this case, the knowledge acquisition problem can be addressed by data mining techniques. Knowledge acquisition using data mining technique eliminates or reduces the difficulty caused by the 'knowledge acquisition bottleneck' of rule based reasoning systems and automates knowledge acquisition by obtaining low-cost and high-quality knowledge base.

Data mining has been defined as the non trivial extraction of implicit, previously unknown and potentially useful information from data [18].It uses machine learning, statistical and visualization techniques to discover and present knowledge in a form, which is easily comprehensible to humans. As noted in [26] data mining is an iterative and interactive process consisting of several steps. To acquire knowledge using data mining, the following steps are included in this research.

i. Data Selection:

The first step in data mining is to select the types of data to be used. A total of 6330 TB datasets were collected from two different places. 5125 datasets were collected from FelegeHiwot hospital and 1205 datasets from Bahir Dar health Center. Each of those records consists of 20 different variables (attributes). For validation purpose, the full dataset is split into 80% (5064) as training dataset and 20 %(1266) as testing dataset. 80% of the original data was selected for training purpose since the classifier learns more from large amount of data and increases its performance. The test data was selected from the original data using simple random sampling technique. The dataset contains the following attributes: age, weight, Haemoglobin, chroniccough,fever, chest pain, bloody sputum, headache, loss of appetite, night sweating, weight loss, exhaustion, HIV test, shortness of breath, sputum test, X-ray test, FNAC test, CSF test, lymph node swelling and TB type.

ii. Data Cleaning:

Refers to detecting and correcting (or removing) incomplete, incorrect or irrelevant parts of the data and then replacing, modifying, or deleting this dirty data. It is the process of ensuring that all values in a dataset are consistent and correctly recorded. Since the original TB dataset contains dirty data, data cleaning tasks were performed. For example, two attributes namely age and weights contain invalid values. To handle those incorrect values, attribute mean technique was used to replace the missed values for age attribute and manual technique used to replace missing values of weight attribute.

iii. Data Integration:

Data integration combines data from multiple sources into a coherent data store. Since the original dataset collected from different sources and the dataset has different attributes for different years, entity identification problem and redundancy issues are solved by taking the common attributes for all years.

iv. Attribute Selection:

In the original data set, there may be some attributes which are not related or not important for analysis. The removal of inappropriate and unnecessary attributes from the dataset is applied on this step of classification. Therefore, attributes such as 'address of patient, 'ART started', and 'medical record number' don't have any significant values for mining purpose and are removed.

v. Data Mining and Model Selection:

Data mining refers to the application of algorithms for extracting patterns from data. To build the predictive model for TB diagnosis, four classification algorithms namely J48, BFTree, JRIP and PART are constructed. J48 and BFTree are tree based classifiers in WEKA whereas JRip and PART are rule based classifiers. In supervised learning, the training data are accompanied by class labels indicating the class of the observations. Class label is the dependent attribute and the rest are independent and value of the class label attribute is predicted. In this study, the dataset has five classes namely TB Suspect, Smear Positive PTB, Smear Negative PTB, Extra PTB and TB Negative.

a) Experiment 1-J48 pruned tree

Decision trees are data-mining methodologies applied in many real-world applications as a powerful solution to classification problems. In decision tree experiment, the performance of J48 classifier in predicting TB status of patients was evaluated. The experiment was conducted with the default parameters of WEKA. From the total dataset of 6330 records, 5926 were correctly classified and the remaining 404 instances were incorrectly classified.

b) Experiment 2- BFTree classifier

BFTree is also a machine learning method for classification. This experiment is based on best first decision tree classfier. The experiment has been conducted with the default parameters of WEKA with respective values and 10-fold cross-validation test mode. Then, BFTree has correctly classified 5970 instances out of 6330 and it has incorrectly classified 360 instances taking 0.57 seconds to build the model.

c) Experiment 3-JRip classifier

JRip is a rule based classifier that extracts rules from a large dataset. With JRip, IF-THEN rules are generated from the experimental TB dataset with the default parameters of WEKA and 10-fold cross-validation test mode. JRip correctly classified 5792 instances from 6330 and the numbers of incorrectly classified instances are 538.

d) Experiment 4-PART classifier

PART is also a rule-based classifier that uses a set of IF-THEN rules for classification. In this experiment PART rule induction algorithm is employed and has generated 15 rules with 10-fold cross-validation test option. This experiment is also conducted with default parameters of WEKA and the algorithm classified 5971instances correctly and 359 instances incorrectly from the total number of 6330 instances.

i. Model Evaluation

All the selected algorithms allow generating rules from the data set. The results of the algorithms are evaluated based on prediction accuracy in classifying the instances of the dataset into TB suspect, smear positive pulmonary TB (PTB+), smear negative pulmonary TB (PTB-), extra pulmonary TB (extra PTB) and TB negative. The performance of classifier algorithms is compared and the one which performed better is selected as prime choice for the knowledge acquisition step.

The accuracy, precision, recall and f-measure of each of the mentioned classifiers which are obtained from the experiment are shown in Table 1.

Table 1. Performance of Classifiers

CR	Correctly classified instances		Incorrectly classified instances		Prec	Recall	F
	Tot	%	Tot	%			
J48 pruned	5926	93.6%	404	6.38%	0.937	0.936	0.936
BFTree	5970	94.32%	360	5.67%	0.944	0.943	0.943
PART	5971	94.3%	359	5.66%	0.946	0.943	0.943
JRip	5792	91.5%	538	8.51%	0.919	0.915	0.916

As shown in Table 1, four experiments were carried out using decision tree classifiers (i.e. J48 pruned tree and BFTree) and rule based classifiers (i.e. PART and JRrip). From this experiment one can observe that the PART classifier achieves best accuracy by classifying 5971 instances out of 6330 correctly comparing with J48, BFTree and JRip. Results of JRip, J48 and BFTree show that nearly equal number of incorrectly classified instances. The highest incorrect classification is registered by JRip algorithm. Table 2 depicts the confusion matrix of the best performing classifier (i.e. PART classifier).

Table 2. Confusion matrix of PART classifier

PTB-**	*TB* negative	*TB* suspect	*PTB+****	Extra *PTB**	Classified as
1362	0	0	50	25	*PTB-*
0	1482	89	0	0	*TB negative*
0	45	1122	0	45	*TB suspect*
14	20	0	808	101	*PTB+*
0	0	45	0	1122	*Extra PTB*

TB denotes tuberculosis
PTB* *denotes pulmonary tuberculosis*
PTB-** *denotes smear negative pulmonary tuberculosis*
PTB+ *** *denotes smear positive pulmonary tuberculosis*

e) Rules generated by PART classifier

In this study, PART classifier has achieved relatively the highest in most of performance evaluation criteria compared to J48, BFTree and JRIP algorithms. The PART classifier has generated 15 rules as shown in Table 3. The rules acquired from the PART classifier algorithm are used for constructing knowledge base. Hence the automatic knowledge acquisition task uses the rules generated from PART classifier for the integration of data mining with hybrid intelligence system.

Table 3. Rules generated by PART Classification algorithm

S.N	Rule
1	Chronic cough=no AND Chest pain = No AND Bloody sputum = No AND Weight loss=No : TB negative
2	Fever =no AND Chronic cough=no AND Bloody sputum = no: TB negative
3	fever=yes and chronic cough=yes AND weight loss= yes: TB suspect
4	sputum_test = negative AND chronic cough = yes: TB suspect
5	headache=yes AND fever=yes and chronic cough=yes AND weight loss= yes AND exhaustion =yes AND sputum_test=positive: Smear positive PTB

6	headache=yes AND weight loss= yes AND sputum_test=negative AND
	X-ray_test=abnormal: Smear negative PTB

7	fever=yes and chronic cough=yes AND CSF=positive: Extra PTB
8	fever=yes and chronic cough=yes AND sputum_test=negative AND CSF=positive AND FNAC=positive: Extra PTB
9	sputum_test = positive AND age = 25-49 AND X-ray_test = normal AND lymph_node _swelling = no :Smear positive PTB
10	Headach=no AND fever=yes and chronic cough=yes AND sputum_test=negative AND FNAC=positive: Extra PTB
11	Chronic cough =yes AND X-ray_test = abnormal: Smear negative PTB
12	Sputum_test=positive AND FNAC =positive: extra PTB

13	Shortness_of_breath = No AND Chest pain = Yes AND Cough = Yes AND Night
	sweats = No AND HIV_test_result = Reactive AND Weight loss = No: TB Negative

14	Shortness_of_breath = Yes AND Chronic Cough = Yes AND Chest pain = Yes AND Loss_of_appetite = Yes: Smear positive PTB
15	: TB suspect

In consultation with tuberculosis (TB) experts, the rules are evaluated to make sure that whether or not they are capable of discovering patterns for predicting TB status of patients. Based on the evaluation, the rules are capable of detecting TB status of the patient even though the generated rules are too small to provide a complete diagnosis for TB disease.

f) Knowledge Representation

The knowledge acquired from data mining classification technique is in the form of production rules. Production rules are the most popular forms of knowledge representation methods for rule based system development. They are constructed in the form of 'if-then' format. For the case based reasoning part of the hybrid intelligent system, the acquired cases are represented using one of the different case representation methods that are appropriate for this research. Among the different case representation methods, feature-value case representation method is used. The reason for representing cases using feature-value representation is that this approach supports nearest neighbor retrieval algorithm and it represents cases in an easy way [8].Cases are selected and retrieved in a ranked order based on their similarity for the given new case query. For this research, nearest neighbor retrieval algorithm is used to measure the similarity of input case with cases in the case base.

IV. Integrating Discovered Rules with Hybrid Intelligent System

As explained in [24], the hybrid intelligent (reasoning) system combines rule based reasoning (RBR) and case based reasoning (CBR). The integration of the data mining results (PART rules) with hybrid reasoning system is done through the RBR sub system. The rules generated by the PART classifier are used to construct the knowledge base of the rule based reasoning module. In order to join the WEKA classifier results with rule based reasoning part of the hybrid intelligent system, Java NetBeans IDE 7.3 with JDK 6 is used.

For representing rules in the knowledge base and constructing hybrid intelligent system, SWI-Prolog 6.4.2 with UTF-8 is used. Figure 1 shows the overall system architecture of integrating data mining results with the hybrid intelligent system.

The system is capable of building domain knowledge from the existing datasets and applies this knowledge to solve critical problems. As data mining extracts domain specific knowledge from large databases, it also enhances the knowledge acquisition process of rule based reasoning module. The solid lines with arrow heads in Figure 1 symbolize the data flows between components. The data mining module consists of the processed data that are mined. The RBR module starts with extraction of domain knowledge using knowledge acquisition methods. The RBR system interacts with users through user interface in a form of menus, questions and answer. Next, the inference engine searches for the appropriate goal that matches with the facts and rules based on the answer of the user. Finally, if the inference engine succeeds, it replies the appropriate result to the user through the user interface; otherwise the user is redirected to the CBR module. The integration module maps the knowledge acquired from data mining classifier into the knowledge base of rule based reasoning. The decision module combines the RBR and CBR modules. It allows the RBR module to display solutions to the user or it calls the CBR module to solve the problem when the RBR module is unable to solve the problem. The CBR module involves collecting and selecting patient cases, building case base, representing cases, measuring similarity for case retrieval and retaining the learned case in the case base. As the new case is entered, the CBR part of the hybrid system matches the new case with the existing case in the case base of CBR by using similarity measurement. If relevant cases are found within the case base, then the CBR module ranks the relevant retrieved cases based on their similarity and proposes a solution.

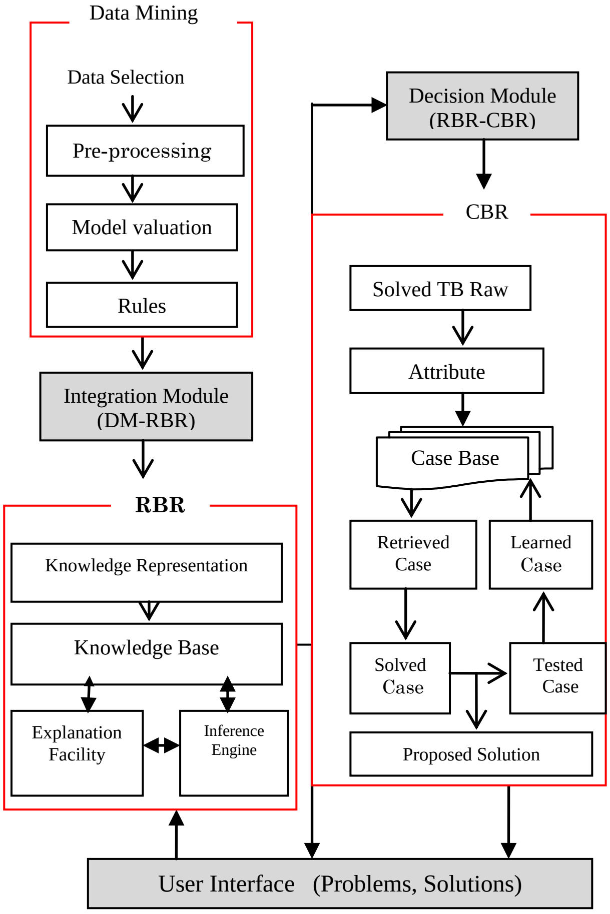

Fig.1. Architecture of the Proposed Hybrid Intelligent System

As shown in Figure 1, the user interface allows the user to enter the symptoms of the disease and the laboratory results into the RBR portion of the hybrid system. The facts and rules of the knowledge base in the RBR portion are accessed to make a diagnosis and provide recommended treatment. If RBR system gives solution to the problem, the result of diagnosis is displayed to the user. If the rule-based module does not give any solution to the given query or problem, the CBR module is consulted to make a diagnosis and treatment. Based on the cases recorded beforehand, a diagnosis is made and the result is presented to the user through the user interface.

V. Testing and Evaluation

Evaluation is an important issue for every intelligent system. The purpose of the evaluation process is to get the end user‘s views on the significance or usefulness of the system. The evaluation and testing issue of the system answers the question “To what extent the hybrid intelligent system give acceptable and accurate diagnosis and treatment service to users?” To answer this question, system performance testing and user acceptance testing methods are used. User acceptance testing is the process of ensuring that whether the system satisfies the requirements of its end-users. System performance testing is applied to evaluate the performance of the hybrid intelligent system which helps to compare and contrast the domain expert’s judgment and the proposed system’s response.

The testing of hybrid intelligent system involves testing the rule based reasoning (RBR) and case based reasoning (CBR) modules. Since the representation and implementation of RBR and CBR modules are independent in the hybrid intelligent system, the process of testing procedure for each reasoning modules is different. As a result, the performance of rule based reasoning and case based reasoning modules are tested independently and the average performance of the two modules is calculated that represents the performance of the integrated system.

For the purpose of testing the rule based system (RBR), six domain experts are selected from FelegeHiwot Referral Hospital. The selected experts are professionals who work in the TB related disease and participated in knowledge acquisition phase as well as in the visual interaction of the system. From the total of 47 TB cases collected from FelegeHiwot Hospital, 42 patient’s cases are diagnosed correctly and 5 patient cases are diagnosed incorrectly by the RBR prototype system of the hybrid intelligent system. TB patient test cases were distributed equally for each expert evaluator. Evaluators start evaluation by testing all possible inputs and validate their subsequent outputs. The expert evaluators identify correctly and incorrectly diagnosed cases by comparing the outputs made by the RBR system with that of the experts’ decision on the same patient cases. The accuracy is calculated by using the formula as follows.

$$Diagnosing\ Accuracy\ = \frac{TC}{TT}\ x100$$

Where TC is the total number of test cases diagnosed correctly and TT is the total number of the test cases. Based on the above formula, the accuracy of the RBR system is approximately 89.4%.

Unlike the RBR engine, CBR engine has a built-in set of test cases in their case library. Effective use of this feature can facilitate the validation process by minimizing the involvement of domain experts in the process.

Retrieval of previously stored cases to solve new problems is the first step in any CBR application. Retrieval of similar cases to the new case from previously solved cases is followed by the reuse of similar solutions.

The CBR retrieval test is designed to evaluate the correctness of the retrieval function. To conduct the retrieval testing, for each test case the relevant TB patient's cases from the case base should be identified. For identification of relevant cases, test cases are given to the domain expert in order to assign possible relevant cases from the case base to each of the test cases. The domain expert uses the value of diagnosis and solution attributes of the TB cases as the main concept to assign the relevant case to the test cases. After the identification of the relevant cases to the test cases by the domain expert, precision and recall are calculated. The retrieval evaluation uses 50 TB patient cases that have been collected from FelegeHiwot referral Hospital.

Table 4. Relevant Cases from Sample Test cases

Test Cases	Relevant cases from the case base
Case 2	Case 1, case22,case 30,case20, case30, case45,case42
Case 3	case 13, case2, case 22,case25,case35,case40,case48
Case 4	case 24, case 10, case 13, case 14,case 15, case 16, case 26, case 28
Case 13	case 23, case 4, case 40, case 16, case 26, case 28,case 43
Case 17	case 7,case 19, case 20, case 21, case 22, case 23, case 29,case 44
Case 19	Case1,case 9, case 20, case 21, case 22, case 23, case 29
Case 22	case 19, case 20, case 21, case 2, case 23, case 29
Case 26	case 13, case 16, case 28
Case 29	case 24, case 19, case 12, case 23, case 20, case 19,case 33

After finishing for identifying and assigning relevant cases to the test cases the next step is calculating the recall and precision value of the retrieval performance of the CBR system with a threshold value. As different authors [21][8] indicated, there is no common standard threshold for the degree of similarity that has been used for retrieving relevant cases in CBR. Different CBR researchers use different case similarity threshold. For this study, the threshold level of [1.0, 0.8) is adopted. This means cases with global similarity score greater than 80% are retrieved.

Table 5. Performance of CBR module using precision and recall

Test Cases	Precision	Recall
Case 2	0.825	1.0
Case 3	1.0	0.89
Case 4	0.784	0.80
Case 13	0.85	1.0
Case 17	0.85	0.88
Case 19	0.91	0.871
Case 22	0.712	0.75
Case 26	1.0	1.0
Case 29	0.8	0.83
Average	0.856	0.891

As shown in table 3 above, the average recall and precision results are 85.6% and 89.1% respectively which indicates that retrieval was done properly. For every test case more than average is registered for both recall and precision. But, precision is lower compared to the average recall. This may be because of the tradeoff between precision and recall.

In general, the proposed system could perform with average accuracy of 87.5% which indicates that the study was effective in acquiring the required knowledge through data mining for diagnosis and treatment of TB disease.

For testing the user acceptance of a system, a questionnaire is prepared to evaluate the user acceptance of a system and the evaluators fill the questionnaire after they have used the system. The researchers adopted the questions from Audrey Mbogho [1], Solomon [3] and Seblewongel [8]. The Evaluators fill the questionnaires after they have used the system. As a result, the proposed system achieved 89.2% of the user acceptance which is the promising result to implement the proposed system.

VI. Conclusions and Future Work

Intelligent systems can help a great deal in decision making through a display of intelligent behavior that may include learning and reasoning. In this research, a hybrid intelligent system that supports diagnosis of TB disease was developed by integrating data mining techniques as a knowledge acquisition step. The aim of integrating data mining techniques with the hybrid intelligent systems is to reduce the difficulty caused by the 'knowledge acquisition bottleneck' and to obtain low-cost and high-quality knowledge base. The system is evaluated using different evaluation methods and the system has achieved 87.5% on system performance testing and 89.2% on user acceptance testing. The use of data mining techniques to build the knowledge base of the hybrid system can be taken as strong features of the system. However, the system lacks to update rules in the knowledge base of the hybrid system and the user interface need to be enhanced with a better graphical user interface that allows users to choose their language preferences. So, further study is needed to improve user interface of hybrid intelligent system and to design a system that can update rules of the knowledge base.

References

[1] Audrey Mbogho..Knowledge Based Expert System for Medical Advice provision. MSc the sis, department of Computer Science, University of Cape Town, 2012.

[2] Ioannis H. and JIM P. Using a hybrid rule-based approach in developing an intelligent tutoring system with knowledge acquisition and update capabilities. Expert Systems with Applications: An International Journal Pergamon Press, ACM DL, Inc. Tarrytown, NY,2004, USA.

[3] Solomon A. Self-learning KBS for diagnosis and treatment of diabetes. MSc Thesis, Addis Ababa University School of Information Science, 2013. Ethiopia.

[4] Ping-Tsai Chung, Bing-Xing Chen. A knowledge-based decision system for healthcare diagnosis and ADVISORY,

IEEE International Conference on Systems, Man and Cybernetics (SMC), 2011.

[5] Deepti John, Rose John. A Framework for Medical Diagnosis using Hybrid Reasoning. Preceesing of international multi conference of engineers and computer scientists, Vol.1, march 17-19, 2010.

[6] Jim and Ioannis . Integrations of Rule-Based and Case-Based Reasoning. Research Academic Computer Technology Institute, Unpublished, Patras, Greece.

[7] Soundararajan et al. Diagnostics Decision Support System for Tuberculosis using Fuzzy Logic. International Journal of Computer Science and Information Technology and Security (IJCSITS), Vol. 2, No.3, 2012.

[8] Seblewongel . Developing prototype knowledge- based system for anxiety mental disorder diagnosis. MSc Thesis, Addis Ababa University, Information Sciencedept, 2012. Ethiopia

[9] Jose Manuel et al. A Formalism and Method for Representing and Reasoning with Process Models Authored by Subject Matter Experts, IEEE Transactions on Knowledge & Data Engineering, vol.25, no. 9, pp. 1933-1945, 2013.

[10] Jose Enrique Munoz Exposito. Swarm Fuzzy Systems: Knowledge Acquisition in Fuzzy Systems and Its Applications in Grid Computing", IEEE Transactions on Knowledge & Data Engineering, vol.26, no. 7, pp. 1, 2014.

[11] Marta A., Argimiro A. and Ramon X. Forecasting with twitter data. ACM Transactions on Intelligent Systems and Technology (TIST) - -Special Section on Intelligent Mobile Knowledge Discovery and Management Systems and Special Issue on Social Web Mining, Volume 5 Issue 1, Article No. 8, 2013

[12] Kapil K., Durga P. Hybrid Reasoning Model for strengthening the problem solving capability of Expert Systems. International Journal of Advanced Computer Science and Applications, Vol. 4, No. 10, 2013.

[13] Aderonke, et al. An Integrated Knowledge Base System Architecture for Histopathological Diagnosis of Breast Diseases. I.J. Information Technology and Computer Science, V.01, PP.74-84. 2013.

[14] Souad Guessoum et al. Combining Case and Rule Based Reasoning for the Diagnosis and Therapy of Chronic Obstructive Pulmonary Disease. International Journal of Hybrid Information Technology, Vol. 5, No. 3, 2012.

[15] Rajeswari P. and V. Prasad. Hybrid Systems for Knowledge Representation in Artificial Intelligence. International Journal of Advanced Research in Artificial Intelligence, Vol. 1, No. 8, 2012

[16] Mariana and Ernesto. Integration of Rule Based Expert Systems and Case Based Reasoning in an Acute Bacterial Meningitis Clinical Decision Support System. International Journal of Computer Science and Information Security, Vol. 7, No. 2, 2010.

[17] T. Bruland et al. Architectures integrating Case-Based Reasoning and Bayesian Networks for Clinical Support System. Z. Shi and al, pp. 82–91, 2010.

[18] Yongjian Fu. Data Mining: Tasks, Techniques and Applications, in Introduction to Data Mining and its Applications. Berlin, 2006, Germany:

[19] Sondak et al. New directions for medical artificial intelligence.Computers &amp; Mathematics with Applications 20(4-6): 313-319, 1990.

[20] Anthony Farrugia et al. Medical Diagnosis: Are Artificial Intelligence Systems able to Diagnose the Underlying Causes of Specific Headaches?.Developments in eSystems Engineering (DeSE), Sixth International Conference. IEEE, Abu Dhabi, 2013.

[21] Shortliffe and E. H. Computer Programs to Support Clinical Decision Making. Journal of the American Medical Association, Vol. 258, No. 1.

[22] Toor, Atif Imran. Decision Support System for Lung Diseases (DSS), 2006.

[23] Rudiger Wirth. towards a standard processmodel for data mining. In Proceedings of the Fourth International Conference on the Practical Applications ofKnowledge Discovery and Data Mining, pages 29–39, 2000.

[24] Adane Nega. Localized hybrid reasoning system for TB disease diagnosis. IEEE Africon (2015) Conference, Addis ababa, 2015, Ethiopia.

[25] Mihaela OPREA. on the Use of Data-Mining Techniques in Knowledge-Based Systems. Economy Informatics, pp.1-4. 2006.

[26] Danubianu M. Combining the Power of Data Mining with Expert System for Efficiently Personalize Speech Therapy". Unpublished. 2012.

[27] I.G.L. da Silva et al. Integration of Data Mining and Hybrid Expert System. FLAIRS-02 Proceedings,AAAI. 2002.

[28] Kittisak and kittaya. Bridging Data Mining Model to the Automated Knowledge Base of Biomedical Informatics . International Journal of Bio-Science and Bio-Technology Vol. 4, No. 1. 2012.

[29] Khademolqorani, Shakiba, and Ali Zeinal Hamadani. "An adjusted decision support system through data mining and multiple criteria decision making." Procedia-Social and Behavioral Sciences, PP. 388-395,2013.

[30] Sellppan P. and Rafiah A. Intelligent heart disease prediction system using data mining techniques. Proceeding, AICCSA'08 Proceedings of the 2008 IEEE/ACS Internatioanl Conference on Computer Systems and Applications. 2008.

[31] Luxmi Verma et al. Integration of rule based and case based reasoning system to support decision making .IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). Ghaziabad, 2014.

[32] Farzad V et al. Fuzzy rule based expert system for diagnosis of lung cancer". Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC), 2015 Annual Conference of the North American , IEEE and Redmond, WA, 2015.

[33] Tesfamariam M and Tibebe B. Integrating Data Mining Results with the Knowledge Based System for Diagnosis and Treatment of Visceral Leishmaniasis. International Journal of Advanced Research in Computer Science and Software Engineering,Vol.5, Issue 5, 2015.

[34] Claudio and Ramon. Towards integration of knowledge based systems and knowledge discovery systems. JCS&T Vol. 7 No. 1, 2007.

[35] Ahamed A. A Method for Classification Using Data Mining Technique for Diabetes: A Study of Health Care Information System International Journal of Healthcare Information Systems and Informatics archive ,ACM DL, Vol. 10, Issue 2, 2015.

[36] Sujata D., et al. A Hybrid Data Mining Technique for Improving the Classification Accuracy of icroarray Data Set. I.J. Information Engineering and Electronic Business. Vol.2,pp.43-50,2012.

[37] Abhijit A. ,et al. Study of Data Mining Techniques used for Financial Data Analysis. international Journal of Engineering Science and Innovative Technology(IJESIT) Volume 2, Issue 3, May 2013.

[38] Folorunsho O. Adesesan B. Application of Data Mining Techniques in Weather Prediction and Climate Change Studies. I.J. Information Engineering and Electronic

Business, Vol. 1, pp. 51-59, 2012.

Authors' Profiles

Mr. Adane Nega obtained his M.S.c degree in information technology from university of Gondar and B.S.c degree in computer science from Bahir Dar University, Ethiopia. He is currently working as a lecturer at the faculty of computing, Bahir Dar University. His area of research interests include Artificial intelligence and soft computing, Data mining, big data analysis, machine learning.

Mr. Alemu Kumilachew obtained his M.S.c degree in information science from Addis Ababa University and B.S.c degree in Information Technology from Jimma University, Ethiopia. He is currently working as a lecturer at the faculty of computing, Bahir Dar University. His area of research interests include information storage and retrieval, Artificial intelligence, question answering systems, machine learning.

How to cite this paper: Adane Nega, Alemu Kumlachew,"Data Mining Based Hybrid Intelligent System for Medical Application", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.9, No.4, pp.38-46, 2017. DOI: 10.5815/ijieeb.2017.04.06