Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 47 (2) 241-246 (2015)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Case Retrieval Process of CBR Technique Implements on Knowledge-Based Clinical Decision Support Systems (KBCDSS) for Diagnosis of Breast Cancer Disease

S. S. ZIA[++], P. AKHTAR[*], T. J. A. MUGHAL

Faculty of Engineering, Sciences and Technology (FEST), Hamdard University, Karachi- Pakistan.

Received: 23rd September, 2014 Revised 13th April 2015

Abstract: This paper present the brief overview of different CDSSs proposed in order to facilitate the medical practitioners during diagnosis phase, with this we proposed an online KBCDSS. Using this system the medical practitioners of different medical domains gather over one platform from where they can check and verify the reliability of their decision making during diagnosis phase. Aim of the research is to provide guidance to the new medical practitioners as well as to experienced clinicians. Database management systems used as knowledge representation scheme and Case based reasoning technique is applied as an inference mechanism. This research performed data analysis on the Wisconsin breast cancer data set from UCI Machine Learning Repository and implements this medical data set on KBCDSS tool.

Keywords: Clinical Decision Support Systems (CDSS), Knowledge based CDSS (KBCDSS), Knowledge Representation (KR), Database Management Systems (DBMS), Case-based reasoning (CBR), Breast Cancer.

1. INTRODUCTION

Human well being or the safety of human life is the biggest concern of the people these days. In this regard different research and development programs are launched especially in the field of medical sciences.

Research indicates that due to incorrect diagnosis of disease, the incident ratio of human life is on high risk (Charles 2011). The keep going growth of information and communication technology especially in the medical field reduces the risk factor of the human life (Bates, *et al.* 2001) and also support the medical practitioners for their decision making process with the help of domain knowledge of the particular disease that is stored in the database. In health care industry, computer aided information not only increases knowledge regarding diseases in question but also provides the key ideas for the remedial measures taken. Medical consultants could easily access to accurate, complete and on time information of the patient by means of CDSS. This would not only reduce the errors but also increase the quality of decision making process in diagnosis and treating the diseases.

Different approaches and parameters are used to deal with different diseases. Different application oriented CDSSs have designed for diagnosis the specific disease shown in **(Table 1).**

The proposed KBCDSS is a multiple diseases diagnostic system which inaugurates the concept to gather medical practitioners of diverse medical fields over one platform through web from where they can check and verify their findings about the patients. Database management systems (DBMS) are used in our model for representing the medical knowledge and store that knowledge in a knowledge-base (KB). As an inference mechanism, we therefore used schematic cycle of case based reasoning technique. Case based reasoning is an approach to obtain knowledge and make inferences by using previously stored cases.

Table 1: Specific Disease oriented CDSSs

Author	Application Domain
(De Paz, et al. 2009)	Cancer Diagnosis
(Glez-Pena, et al. 2009)	Cancer Classification
(Marling, Shubrook and Schwartz 2008)	Diabetes
(Ahn and Kim 2009)	Breast Cancer Diagnosis
(Obot and Uzoka 2009)	Hepatitis
(Ahmed, et al. 2009)	Stress Management

This paper is organized as follows: Section 2 shows the proposed model of online KBCDSS. Section 3 shows CBR used as an inference engine. Section 4 provides the discussion of the proposed system with the implementation of breast cancer dataset. Finally, Section 5 provides the conclusion of this study.

[++] Correspondence author: Syed Saood Zia, email: saood_zia@hotmail.com , cell: +92-346-2767788
**Pakistan Navy Engineering College, National* University of Science & Technology

2. KNOWLEDGE-BASED CDSS (KBCDSS)

The proposed framework of online KBCDSS launched the concept to "gather" medical practitioners of different medical domains over one platform and provides a graphical user interface environment that facilitates the medical practitioners to perform the knowledge acquisition process in an effective manner for the diagnosis of diseases.

Using KBCDSS the medical practitioners of different medical domains can upload the medical cases that are attributed of diseases along with their suggested and reported diagnoses. These piled up cases could be utilized for the diagnoses purpose by the medical experts all over the world through web. The medical practitioners can input the new medical case and then compare them with most similar cases that are already uploaded into the system. On comparing the attributes of disease of the new input case with the stored cases, the medical practitioners extract out the most similar case along with their diagnosis. This comparison provides an aid to the medical practitioners in their decision making during the phase of diagnosis. The proposed model of online KBCDSS is integrated with the architecture of the data warehouse that will complete the KDD process **(Fig. 1),** to get the knowledge for the selected disease.

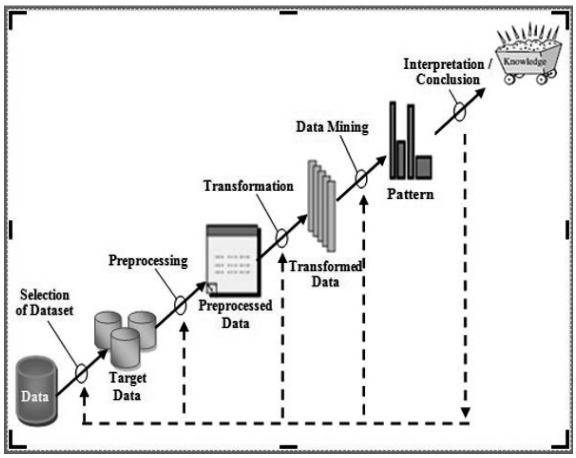

Fig. 1: KDD Process (Fayyad, Piatetsky-Shapiro and Smyth 1996)

Architecture of online KBCDSS

For knowledge data discovery (KDD), the proposed model **(Fig. 2),** pursues the following operations.

The bottom layer can be explained as the potential source of knowledge from where medical experts extract the medical data. Potential sources of knowledge includes medical experts, medical reference books, multimedia documents, flat files, databases (public and private), research reports and information available on the web.

The layer 0 shows the knowledge acquisition process. Knowledge acquisition is the accumulation, transfer and transformation of problem-solving expertise from expert or other knowledge sources to a computer program for construction and expanding the knowledge-base (Turban, *et al.,* 2005). The process of knowledge acquisition follows the following functions:

- **Data extraction:** Extract data from multiple defined sources.
- **Data cleaning:** Detect the mistakes or incorrectness of data from the extracted data and rectifies them when possible.
- **Data transformation:** After the cleaning process, the refined data is transformed into the data warehouse format.
- **Loading**: Load the data in a data base.
- **Refresh:** Spread the updates from the data sources to the data warehouse.

Layer 1 shows the warehouse database server which is relational database system. In the proposed model of KBCDSS, layer 0 performs knowledge acquisition process and as a consequence data is updated from the data sources to the data warehouse. The warehouse database server retains all the different medical records that are stored in relational tables. The warehouse database server also maintain metadata repository, which stores information about the data warehouse and its contents.

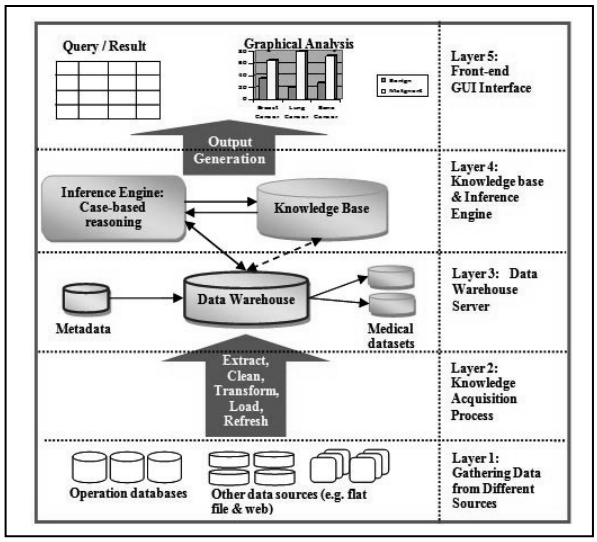

Fig. 2: A Proposed Framework of Knowledge-based Clinical Decision Support System (KBCDSS)

Layer 2 consists of knowledge-base and inference mechanism. Knowledge-base is an organized collection of definite domain knowledge which is required in problem solving phase. In order to pile up and deal with structural knowledge, database management systems

used as knowledge representation scheme in KBCDSS. As an inference mechanism, we therefore used case based reasoning technique in our model. Case based reasoning is an approach to obtain knowledge and make inferences by employing previously stored cases. These cases are consisting of detailed description of the problem along with their way out. In order to solve the new case, we compare previously stored case with the new case and then repossess the similar cases.

Layer 3 is a graphical user interface environment of the KBCDSS. Using the GUI environment the medical practitioners input the medical cases of the patient in to the system. The system generates the result and is shown in text as well as in graphic form to the medical consultants.

3. INFERENCE TECHNIQUE - CBR

Since last two decades, number of knowledge based techniques is considered to be functional or useful for the diagnosis purpose (Shahina, *et al.* 2011). Case based reasoning is one of the famous cognitive science based procedure for medical knowledge based systems which guides while evaluating the circumstances (Salem 2011). (Aamodt and Plaza 1994) have summarized the task of CBR as follows: **(Fig.3),**

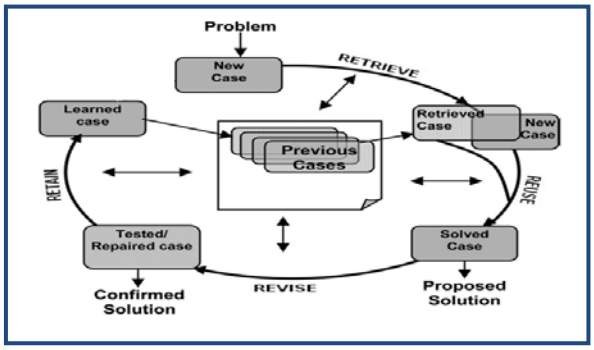

Fig. 3: CBR cycle, introduced by (Aamodt and Plaza 1994)

Retrieve: in the retrieval phase, a new case is compared with the old cases stored in case repository. For solving the problem into the new case we retrieve alike cases from case repository. Reuse: Afterward the closely related solutions are projected with some alterations if essential for solving the new case. Revise: in this phase, we revised and confirmed the preferred solutions. Retain: finally, retain that new case in the case repository for further use (Pal and Shiu 2004), (Aamodt *et al.*, 1994). Using the case retrieval phase of the CBR technique, we extract out the cases from the case library which relates most closely to input case.

3.1. Algorithm applied in Retrieval phase

In the proposed model of KBCDSS medical experts primarily recognize the particular disease (i.e. identify the disease type). Then assign weights to the attributes of the diseases on the basis of their importance. These attributes are the signs and symptoms of the patient. Afterward case retriever input the new case into the case repository and compares its attributes with the cases stored earlier.

Average weighted Euclidian distance is used in our proposed model for computing the distance between the new input case and the stored cases. For easy computation, the range of the distance measures between the new and stored cases can be normalized into 0 to 1. For this purpose, C_{Max} is used for converting the value into normalized form. The weighted Euclidian distance formula for calculating the distance between the input case and the pile up cases that stored in the case library is shown in eq.1 (Zia, et al. 2014).

$$d(C_{New},C_{Old})=\frac{\sum_{i=1}^{n} w_i \times \sqrt{\left|\frac{C_{New\,i}-C_{Old\,i}}{C_{Max\,i}}\right|^2}}{\sum_{i=1}^{n} w_i} \quad (1)$$

Where

- $\mathbf{C_{New}}$ = New referred case from clinicians
- $\mathbf{C_{Old}}$ = Previously stored case in a case repository
- $\mathbf{C_{Max}}$ = Maximum value selected from the new referred case or previously stored.
- **n** = Attributes in each case.
- **i** = is an individual or signal attribute.
- **W** = Weight of each attribute. These weights determine the importance of each attributes and are assigned by field experts.

3.2. Computing Similarity

Once the distance between the new input case and the stored cases are computed in normalized form then apply similarity measurement function that will show the most similar cases with the new input case. The calculation performs for computing similarity measurement function shows in eq. 2 and eq. 3 (Zia, *et al.* 2014).

$$Sim(C_{New}, C_{Old})=1-d(C_{New}, C_{Old}) \quad (2)$$

$$Sim(C_{New}, C_{Old})=\left[1-\frac{\sum_{i=1}^{n} w_i \times \sqrt{\left|\frac{C_{New\,i}-C_{Old\,i}}{C_{Max\,i}}\right|^2}}{\sum_{i=1}^{n} w_i}\right] * 100 \quad (3)$$

4. RESULTS AND DISCUSSION

Online KBCDSS is deployed as an effective prototype application in the medical field. Medical practitioners can use this application during patient evaluation phase it's because this system has the competency to provide detailed structured knowledge to its users. In order to accomplish an effective functioning of this system certain course of action is followed for data analysis on the Wisconsin breast cancer dataset from UCI Machine Learning Repository implements this medical data set on proposed KBCDSS tool **(Table 2).**

Breast cancer is classified into benign (not cancerous tissue) or malignant (cancerous tissue). The study tested the condition of breast cancer tumor for diagnosing if it is benign or malignant (Zia, 2014).

Table 2: Wisconsin Breast Cancer Dataset

#	Attribute Name	Possible Value
1	Case ID	Id Number
2	Clump Thickness	1 - 10
3	Uniformity of Cell Size	1 - 10
4	Uniformity of Cell Shape	1 - 10
5	Marginal Adhesion	1 - 10
6	Single Epithelial Cell Size	1 - 10
7	Bare Nuclei	1 - 10
8	Bland Chromatin	1 - 10
9	Normal Nuclei	1 - 10
10	Mitoses	1 - 10
11	Class	(2 for Benign, 4 for Malignant)

The number of instances inside the dataset is 699. Each record contains ten attributes plus the class attribute. Table 2 shows the attributes and their possible values. 65.5% of the elements belong to the benign class and 34.5% to the malignant class. 54 elements are incomplete (an attribute is missing) and have been excluded from the database. The medical practitioners used the proposed system for classification and diagnosis the disease of the patient. They first perform the knowledge acquisition process in which gather the medical data from different potential sources like own medical cases of their patients, different medical data web sites, research journals and hospitals.

(Fig. 4) shows the knowledge acquisition process in the processed online KBCDSS tool. In this process the medical practitioners identify the main diseases like Cancer and also mentioned the sub category like Breast Cancer then medical practitioners upload the breast cancer medical data set that is in (*.csv) format. Accuracy in storing cases in the case repository or knowledge-base (KB) is essential to make accurate decisions during diagnosis and treatment phases of patient care. Once the medical cases are uploaded then the proposed system initially cleans all the inappropriate data from the current uploaded file.

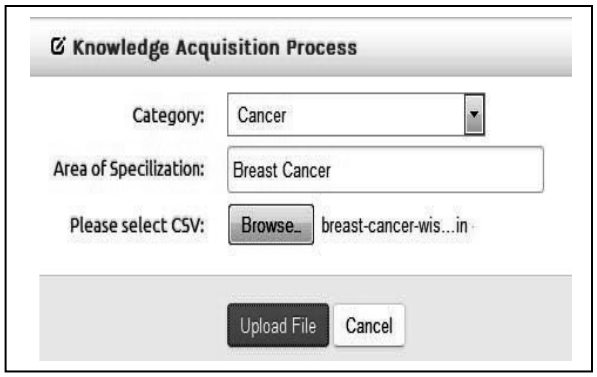

Fig. 4: Knowledge Acquisition Process in KBCDSS

This cleaning process has minimized the chances to make less accurate and incorrect decisions. After cleaning the garbage data from the uploaded file, medical practitioners now focus onto the formation of knowledge-base (KB). **(Fig. 5)** basically shows the case repository or knowledge-base or a case definition of the specific diseases.

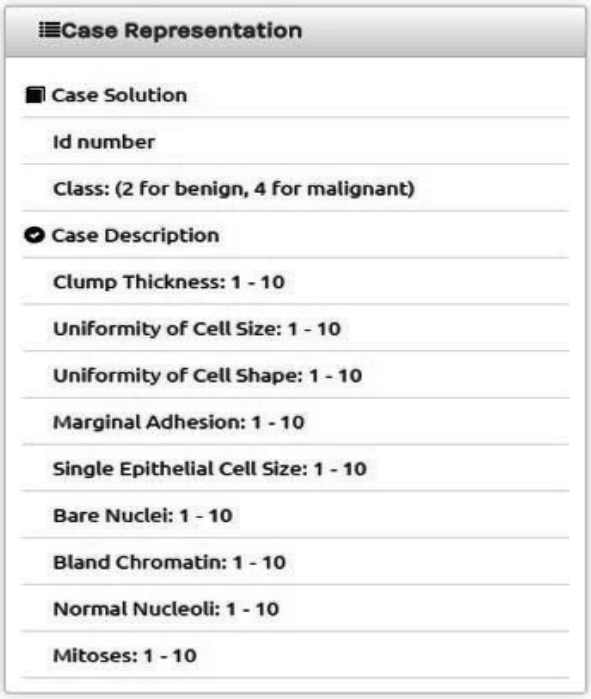

Fig. 5: Knowledge-base of a Breast Cancer disease

While building the case repository or the knowledge-base of the specific disease, the medical practitioners will go for the diagnosis phase. In this phase, the medical practitioners select the main category of the disease and then select the sub category of the disease along with the percentage of the threshold value.

Then the proposed system will retrieve the attributes of the selected disease. **(Fig. 6)** shows the diagnosis process of the breast cancer dataset. In this screen the medical practitioners assigned the weights to the attributes of disease that starts from [1 to 10] i.e. 1 is minimum and 10 is the maximum weight of the feature value. The medical experts input the pathological report value of the patient to the proposed system and that will be treated as a new case. After that, the medical expert will assign the weights of each attribute. These weights determine the importance of each attributes and will be different from the opinion of the different medical experts. The percentage of the threshold value is basically help the medical expert for retrieving the similar record that will not be below the identified threshold value. After assigning the weights of each attributes along with the input of the new case value, the proposed system will retrieve the most similar cases from the knowledge base or case repository.

(Fig. 7) shows retrieve cases from the case repository. With the help of graphical representation, the medical practitioners can easily evaluate the highest similar case that will relate to the new input case.

Fig. 6: Diagnosis Screen of the proposed system

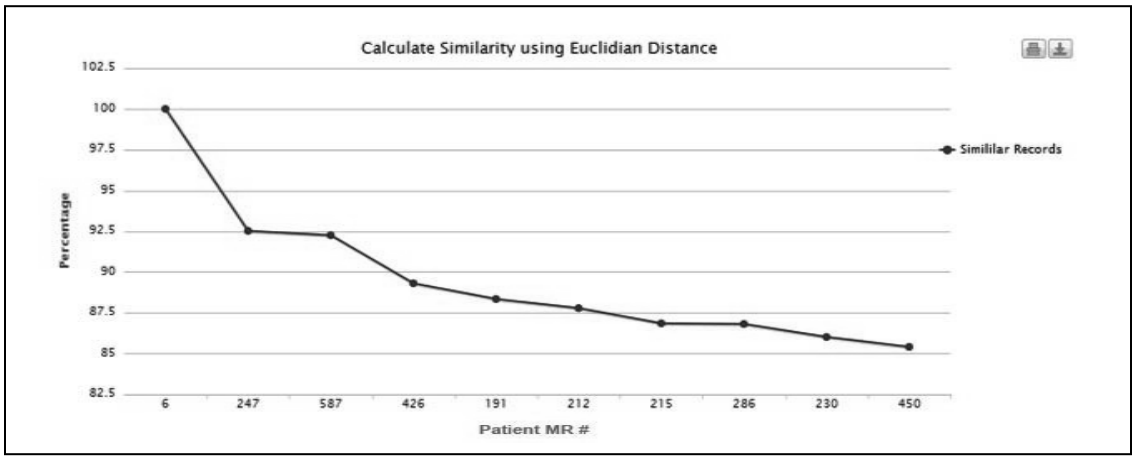

Fig. 7: Retrieve the most similar cases from the case repository

5. CONCLUSION

KBCDSS is proposed as a reliable decision support system for diagnosis multiple diseases by the medical practitioners all over the world through web. It's a step wise procedure which helps medical practitioners in their diagnoses process when they input the case (attributes of the disease) and compared it with the cases stored into the case library. These cases are compared on the basis of similarity in attributes of disease and the closely related stored cases along with its diagnoses are then selected. Data is being initially collected from different potential sources of data. Accumulated data is then undergoing the knowledge acquisition process which updates data from data sources to data ware- house. Now data warehouse have records of all the medical dataset of multiple diseases. Afterward database management systems (DBMS) is used as a knowledge representation scheme for the knowledge base. Inference mechanism then done through Case base reasoning technique. Using case retrieval phase we extract out the closely related cases from the case library which helps in diagnosis the problem of input case. The system can used these stored cases are utilized as a diagnostic tool for new cases.

Future work will focus on introducing the case adaption to reuse the closely related case. Hybrid reasoning approach will be used for this purpose.

REFERENCES:

Aamodt, A. and E. Plaza (1994) "Case-based reasoning: Foundational issues, methodological variations, and system approaches." *AI communications* 7, no. 1: 39-59.

Ahmed, M. U., S. Begum P. Funk N. Xiong. (2009)."A multi-modal case-based system for clinical diagnosis and treatment in stress management." *In 7th Workshop Case-Based Reason. Health Sci.* Seattle, Washington,

Ahn, H., and K. J. Kim. (2009) "Global optimization of case-based reasoning for breast cytology diagnosis." *Expert Systems with Applications* 36, no. 1 724-734.

Bates, D., M. Cohen, L. Leape, J. Marc Overhage, M. Michael and T. Sheridan. (2001) "Reducing the frequency of errors in medicine using information technology." *Journal of the American Medical Informatics Association*,vol. 102: 299-308.

Charles, Vincent. (2011).*Patient Safety.* Chichester, West Sussex,: John Wiley & Sons,

De Paz, J F., S. Rodriguez, J. Bajo, and J. M. Corchado. (2009) "Case-based reasoning as a decision support system for cancer diagnosis: A case Study." *International Journal of Hybrid Intelligent System* 6, no. 2: 97-110.

Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. (1996) "Knowledge Discovery and Data Mining: Towards a Unifying Framework." *In KDD* 96 82-88.

Glez-Pena, D., F. Diaz, J. M. Hernandex, J. M. Corchado, and F. Fdez-Riverola. (2009) "geneCBR: a translation tool for multiple-microarray analysis and integrative information retrieval for aiding diagnosis in cancer research." *BMC bioinformatics* 10, no. 1, 187Pp.

Obot, O. U. (2009) "A framework for application of neuro-case-rule base bybridization in medical diagnosis." *Applied Soft Computing* 9, no. 1 245-253.

Pal, S. K., and S. C.K Shiu. (2004) *Foundations of soft case-based reasoning.* Vol. 8. John Wiley & Sons, 2004.

Salem, A M. (2011): "Intelligent Technologies and Methodologies for Medical Knowledge Engineering." *ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer*, 25Pp.

Shahina B., M. U. Ahmed, P. Funk, N. Xiong, and M. Folke. (2011) "Case-based reasoning systems in the health sciences: a survey of recent trends and developments." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*, 421-434.

Turban, E., J. Aronson, and T. Liang. (2005) *Decision Support Systems and Intelligent Systems.* 7. Pearson Prentice Hall, 2005.

Zia, S. S., P. Akhtar, T. Javid. A. Mughal, and I. Mala. (2014) "Case Retrieval Phase of Case-Based Reasoning Technique for Medical Diagnosis." *World Applied Sciences Journal* 32, no. 3 451-458.