Development of an Intelligent Approach for Medical Knowledge Discovery and Decision Support

Ali A. Sakr
Dept. Electrical Engineering
Faculty of Engineering,
Kafrelsheikh University,
Kafrelsheikh, Egypt

Diana T. Mosa
Dept. Mathematics
Faculty of science, Mansoura
University,
Mansoura, Egypt

Abdelhafiz Shehabeldien
Head of Neurosurgery Dept.,
Mansoura New General
Hospital, Mansoura, Egypt

ABSTRACT
Electronic knowledgebase (including Electronic Medical Record), together with inference procedures, form an intelligent medical information systems that offers many possibilities for health care providers. It acts as a strong base for scientific upgrading and provides an enormous support for developing new medical decisions. This paper proposes an intelligent neurosurgical decision support system framework. This framework merges the advantages of electronic medical record, rough set theory, clinical pathways, latest available scientific researches, and patient's expectations. We have designed and developed this aiming to get a support system for predicting the ideal treatment method of lumbar disc patients, in addition to evaluating the treatment plan. The system ensures future up-to-date knowledgebase, through permitting upgrading with the most recent innovations in knowledge and discoveries. This framework is expected to improve the quality of health care by providing the necessary requirements of neurosurgery domain.

Keywords
Electronic Medical Record; Rough Set; Patient's Expectations; Medical Decision Support System; Android

1. INTRODUCTION
Decision Support System (DSS) is a class of information systems types which used by the top level management at the hierarchal managerial levels. A major goal is to develop guidelines for designing and implementing systems that can support decision making [1].

Electronic Medical Record (EMR) system offers support in medical decision making, coordination between different healthcare providers and improves the overall quality of care. This is achieved through decreasing redundancy, costs [2], and reducing the likelihood of treatment errors [3].

Patients can participate in the evaluation of primary health care quality in three ways: by defining what is desirable or undesirable (i.e. setting standards of care), by providing information that permits others to evaluate the quality of care; and by expressing satisfaction or dissatisfaction after care [4].

2. RELATED WORKS
2.1 Medical DSSs Approaches
"Clinical Pathways" are used in order to optimize services rendered by hospitals both from the medical and economic perspective. It defines the essential steps during treatment and nursing care along a time line, to determine the method of treatment for a given disease [5].

"Evidence-Based Medicine (EBM) and Shared Decision-Making (SDM)" is the integration between best research evidence, clinical experience and patient values/expectations. The need to make patient preferences for health care decisions and clinical outcomes explicit is one of the key elements in integrating SDM into clinical practice. Patient preferences can be defined as the appraisal by an individual regarding the relative desirability of entities, such as health states, treatment, outcomes of treatment/care, or other aspects of health care [6].

In recent decades, determining the level of patient satisfaction has been found to be the most useful tool for getting their view on how to provide care. This relies on the fact that patients are the best source of information on quality and quantity of the provided medical services. In addition, patients' view are determining factors in planning and evaluating the degree of satisfaction, that is considered as the extent to which the patients feel, their needs and expectations met by the services provided. Donabedian (1980) argued that the measurement of satisfaction is, therefore, an important tool for research, administration and planning. Jiapaul and Rosenthal (2003) found that the level of satisfaction increased with age and there was an inverse relationship between patient satisfaction and education [4].

"Rapid Learning" is the reuse of historical data from routine clinical practice for decisions concerning new patients or to test new hypothesis [7].

2.2 Rough Set Theory as Knowledge Discovery Tool
Rough Set Theory (RST) is a relatively new mathematical and artificial intelligent technique introduced by Zdzislaw Pawlak, Warsaw University of Technology, in the early 1980s. RST is particularly useful for discovering relationships in data, managing uncertainties, and reasoning about imprecise or incomplete data [8, 9, 10, 11, 12, 13]. As these tasks are common in many medical applications, it is only natural that rough sets, despite their relative 'youth' compared to other Knowledge Discovery (KDD) techniques; provide a suitable method in such applications [11]. Abbott defines KDD in healthcare as the process of "the melding of human expertise with statistical and machine learning techniques to identify features, patterns, and underlying rules in large collections of health care data" [14]. In a knowledge representation system, each entity is described by the attribute values of the set of condition and decision attributes. Some condition attributes can be redundant and thus can be eliminated [9].

Lemma 1: Core and Reducts of Attributes

The core attributes of set of condition attributes C is equal to the set of all indispensable attributes in C. A condition attribute a ∈ C is indispensable if a deletion of a condition attribute from C results in a different equivalence relation from the equivalence relation defined for C. Reducts are all possible minimal subsets of attributes, which lead to the same number of elementary sets as the whole set of attributes [9].

Lemma 2: The dependency of N on A

The dependency of N on A, denoted as γ_A (N), is a plausible measure of how much N depends on A and is defined as follows:

$$\gamma_A (N) = \mid POS_E (N^*) \mid / \mid U \mid \quad (1)$$

Where $E = (U, \tilde{A})$ is the approximation space, and $\mid\;\mid$ denotes the cardinality of a set, and $0 \leq \gamma_A (N) \leq 1$.

Lemma 3: Significance $\sigma_A(N)$

Pawalk (2004), defined the significance of N on a specific condition attribute a can be defined by using the dependencies as follows:

$$\sigma_{\{a\}} (N) = \gamma_C (N) - \gamma_{C-\{a\}} (N) \quad (2)$$

2.3 Knowledge Representation Techniques

Knowledge representation is the methodology that knowledge is encoded by transforming knowledge from one form to another. There are different types of representation, such as logic, rules, frames or semantic nets [14].

2.4 Software Architectures

"Layered systems" use layers to separate different units of functionality. It is possible to define multiple layers at the same level [16]. "Three-Tier architecture", it is client-server architecture consists of three layers: Presentation layer, Application Layer, and Database layer [17].

3. RESEARCH METHODOLOGY

Several visits to Neurosurgery Department of Mansoura New General Hospital defined the problem in establishing an effective intelligent medical system to support decision making process. Access to good data is required to make decision system able to deduce good and accurate decisions. An efficient EMR system provides required information for Medical DSS.

The study was carried out using an intelligent hybrid approach, consists of combination of different medical DSS approaches (rapid learning, evidence-based medicine, and shared decision making) to mine through medical database for the purpose of decision making (Figure 1).
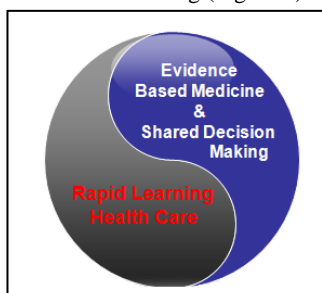

Fig 1: Intelligent hybrid approach

The predicted result of the proposed system is to make an integrated decision based on the reuse of historical data (previous patients have the same condition) for decisions concerning new patients' treatment, ben

efits from latest published scientific research for the same condition, usage of previous experience through clinical pathways, patient perspective and his degree of satisfaction about treatment (Figure2).
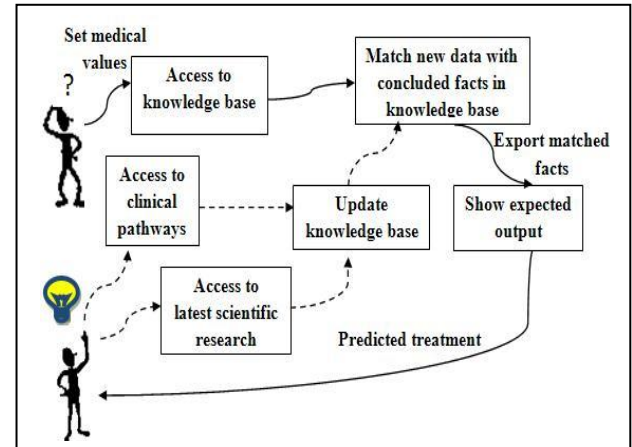

Fig 2: Procedures for decision making process

4. THE PROPOSED FRAMEWORK

The proposed system is to develop an Intelligent Neurosurgical Database System (INDBS), consisting of two main subsystems. The first is the EMR subsystem: its data is the core data source for DSS subsystem. The second subsystem for lumbar disc patients: it predicts treatment for new patients and evaluates treatment planning for old patients. The INDBS architecture is constructed from several components that communicate with each other based on the software's three-tier architecture (Figure 3).

The bottom tier, it contains the DB layer based on oracle 10g to optimize maintenance, performance and flexibility of databases. At this layer, there is all the data of the system and it consists of two layers: **Sourcing layer,** the runtime application's data repository, in which the user session data is stored. These data are clinical and patient expectations data. **Storage layer,** this layer contains data warehouse, data mart, refined data, knowledge base that stores resulted knowledge, clinical pathways, and scientific research.

The middle tier, it contains the application layer that involves all the system applications:-

- EMR application: concerns with all patients' visit data.
- Android application: determines patient's degree of treatment satisfaction.
- Data mining engine: it is the computational engine that communicates with the lower to retrieve and update data that performs calculations and analyses with the aim of transferring the relevant data to the user interface in the upper tier. This engine is responsible to determine the sufficient attributes, required to predict treatment. Its codes are written in PL/SQL, and Java EE. By using this engine, the physicians can predict the best available treatment to the lumbar disc patients. Hence, the physicians provide only sufficient information and it is the engine who contacts the implied treatment.
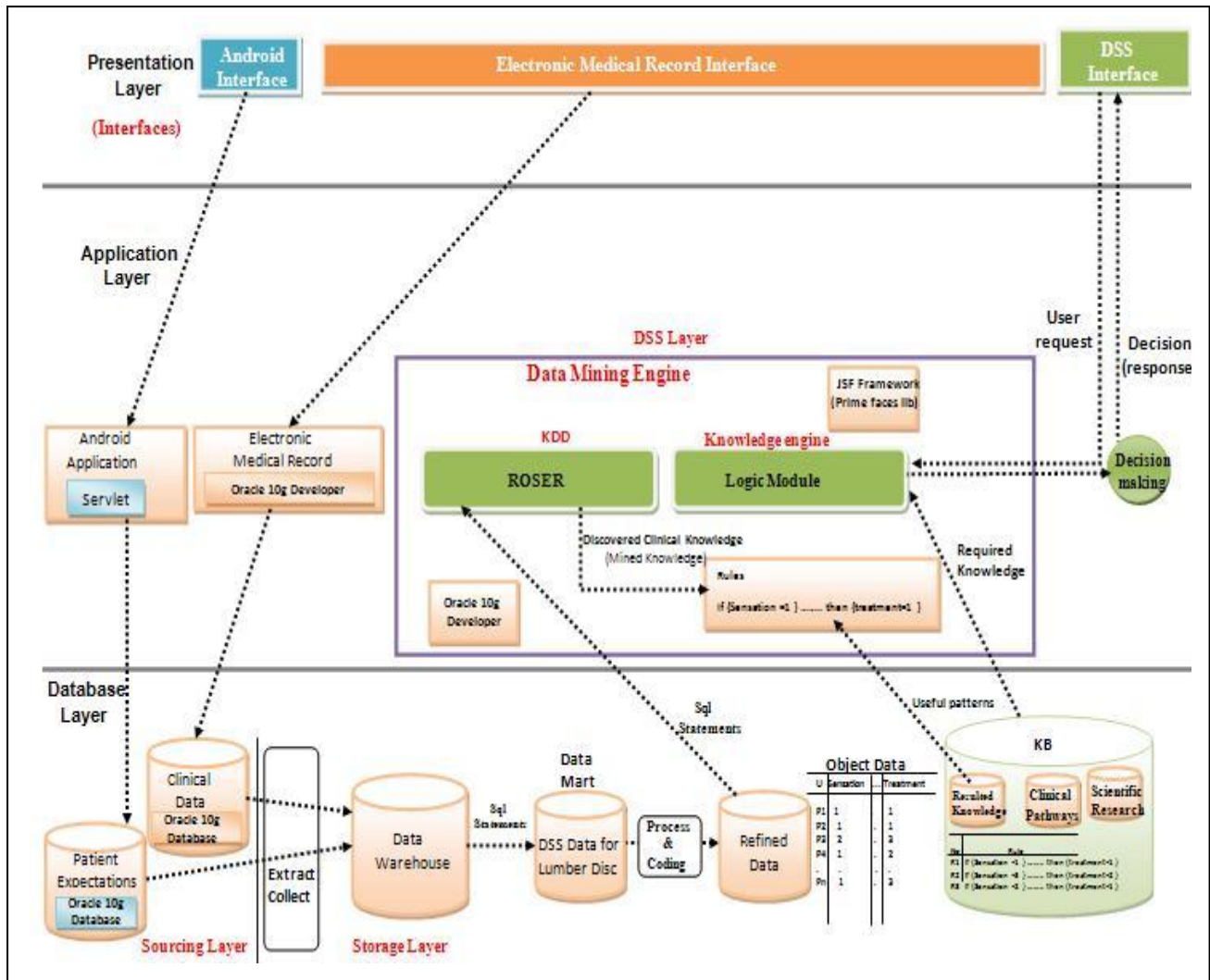
Fig 3: Proposed framework

The Upper Tier, it contains the presentation layer based on XHTML, Cascading Style Sheets (CSSs), Java script, and java served faces (JSFs) standard, in order to obtain separation between presentation and behavior. At this layer, there are all the interfaces of the proposed system. Users can access INDBS via wireless network according to different access levels, defined for each role with encrypted passwords.

5. THE PHASES OF HYBRID APPROACH

The approach consists of three phases, which are described as the following:

5.1 Data Phase

During this phase, the start point is to construct data component for Medical DSS. There are different data sources: clinical data and patient expectation values. The concept of a data warehouse was deemed the solution through receiving data from multiple data sources to meet the requirements of a system capable of supporting decision making.

The data held in a data warehousing is described as being subject-oriented, integrated, time-variant, and non-volatile [18, 19, 20]. A typical data warehouse sits in the center of a data supply chain. The sourcing process is all the operations associated with Extraction, Transformation, and Loading (ETL) of data in the warehouse, it means sourcing the data, structuring it in appropriate format for the data warehouse schema, and then storing it in the data warehouse [18, 20, 21]. The online transaction processing (OLTP) systems are optimized for a high number of transactions that are predictable, repetitive, and update intensive. The OLTP data is organized according to the requirements of the transactions associated with the business applications and supports the day-to-day decisions of a large number of concurrent operational users [18].

As a data warehouse grows larger, to improve end-user response time, the data mart is created. Data mart is a subset of data warehouse, which supports the decision making requirements of a particular department or business area. The cost of implementing data marts is normally less than that required to establish a data warehouse, as, data marts normally use less data than data warehouse so tasks such as data cleaning, loading, transformation, and integration are far easier. To design database for decision support, we analyzed the collected data by dimensionality modeling technique to support decision making. Dimensionality modeling uses the same diagram notation as ER modeling with some important restrictions. Star schemas can be used to speed up query

performance by denormalizing reference data into a single dimension table [18].

5.2 Knowledge Phase

In this phase, we build knowledge component of Medical DSS. Knowledge can be extracted from data mart through machine learning. In machine learning, models/algorithms are developed that best describe the data and can make predictions for new data. The data mining aspect of rough sets can be regarded as a technique of machine learning.

According to rough set concepts for elimination of dispensable attributes and obtaining core and reducts attributes, the following algorithms as presented in Table 1 and Table 2 based on Data Manipulation Language and Data Definition Language of SQL statements [modified from9 & 22].

The physicians use 21 attributes (condition attributes) to determine lumbar disc treatment. We develop Rough Set Reduction (ROSER) application based on the previous algorithms, to reduce knowledge base.

Using ROSER, 21 attributes were reduced to only 5 attributes which are sufficient to determine treatment for lumbar disc patients as shown in Figure 4. So, ROSER generates new reduced Knowledge base, and presents the set of production rules which control treatment.

Table 1: Algorithm for generating core for set of attributes

Input:	a decision table T(C,D)
Output:	Core-the core attribute of table T

Method:	Set Core = $\varnothing$
	
	 For each attribute $C_i \in C$ {
	
	 If card(Π(C-C_i+D)) $\neq$ card(Π(C-C_i))
	
	 Then Core =Core $\cup$ C_i
	
	 }

Table 2: Algorithm for generating reduct for set of attributes

Input:	a decision table T(C,D)
Output:	A set of minimum attribute subset(REDU)

Method:

(1) Run Algorithm 1 to get the core attributes of the table CO

(2) REDU = CO

(3) AR = C – REDU

(4) Compute the significance values for all attributes of AR

(5) Sort attributes in AR based on significance values in decreasing order

(6) Choose an attribute Cj with the largest significance value (if there are several attributes with the same significance value, choose the attribute which has the least number of combinations with those attributes in REDU.

(7) REDU = REDU $\cup$ {Cj}, AR = AR − {Cj}

(8) If K(REDU ,D) = 1, then terminate, otherwise go back to Step 4.

5.3 Application Design Phase

In this phase, we design the application that use and process the intelligent system. We apply the generated knowledge from the previous phase, by designing the user interfaces (Dialogue components of medical DSS). The developed system is implemented using oracle database 10g, oracle developer 10g, J2EE (JSF framework, Servlet), and android development 3.4 Jellybean which are integrated by using a powerful user interface leads to a powerful intelligent system that can support the decision making process. It is deployed in the Neurosurgical Unit of Mansoura New General Hospital.
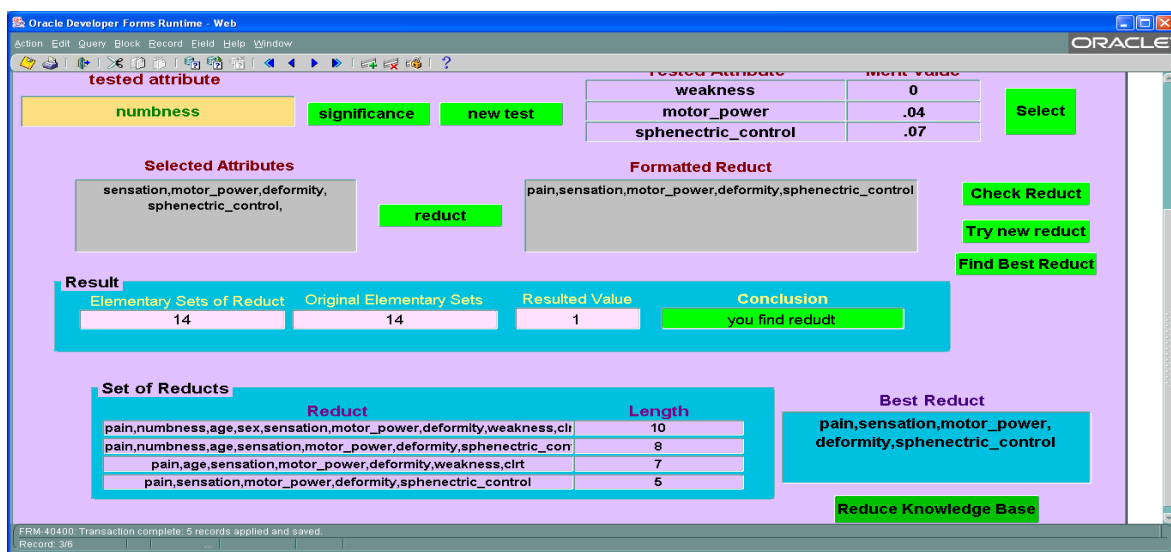

Fig 4: ROSER result sufficient attributes to predict treatment

5.4 System Evaluation

Medical DSSs approaches are the newest direction for today requirements. Questions have been raised about the way that physicians will take their decisions into updatable health care. These rapid changes have a serious effect on decision making.

INDBS is a complete solution, designed to work in an individual or network based environment. Health care at the neurosurgical unit will be improved as the system will improve the physician performance, written guidance of medication, accuracy and easiness of follow up for health status, time spent to retrieve medical record and that for receiving medication.

INDBS is an integrated Medical DSS, differs from other Medical DSS [Tomazewski, Ronald and Bakken, Lambin et al.], and as the decision of treatment is made based on different insights.

We employ a new hybrid approach which combines the principles of rapid learning, EBM and SDM. It is expected to improve the predictability of outcome by reusing clinical data to develop knowledge which can predict treatment with the hypothesis that treatment obtained in the past can be used to predict future results as shown in Figure 5.

Table 3: The Ranges of variable patient clinical attributes recorded during physician's examination

Attribute	Domain Range
Pain	(0,1,2,3),(4,5,6,7),(8,9,10)
Sphincteric control	yes, no
Deformity	(0, 10, 20, 30), (40, 50,60,70)
Sensation	yes, no
motor power	(Grade1, Grade2, Grade3), (Grade4, Grade5)
Treatment Planning	Medical treatment, physiotherapy, Surgery: Open (Laminectomy) / Microscopic / Endoscopic

Also, we involved latest scientific researches, clinical pathways, and patient perspective and his satisfaction about treatment. Figure 6 and Figure 7 show the possibility of continuous change in knowledge base according to new results in medical domain, so the knowledge base of INDBS is always up-to-date. An android device is used as an evaluation device to insert the patient's degree of satisfaction about treatment. The proposed android application would be able to plug into INDBS, allowing doctors to be more easily involved in the process of patients' satisfaction about treatment planning and add any comments. This will have an important role in developing treatment planning, the quality of medical services, and health care in general (Figure 8).

Fig 5: Predicted treatment is physiotherapy: treatment chosen by the system according to the physician given values

Update Lumbar Disc Data

Rule	Sensation	Pain	Sphenectric Control	Deformity	Motor Power	Treatment Planning
1	no	8-10	yes	0-30	G1-G3	phsiotherapy
2	no	4-7	yes	0-30	G1-G3	laminectomy
3	no	4-7	yes	0-30	G4-G5	medical
4	yes	8-10	yes	40-70	G4-G5	phsiotherapy
5	yes	8-10	yes	0-30	G1-G3	phsiotherapy
6	no	8-10	no	40-70	G4-G5	phsiotherapy
7	no	8-10	no	40-70	G1-G3	endoscopic
8	no	4-7	yes	40-70	G4-G5	endoscopic
9	no	4-7	no	40-70	G4-G5	phsiotherapy
10	yes	4-7	yes	40-70	G4-G5	phsiotherapy

Undiserable Attributes

sensation

Add New Attributes | Save Changes
Delete Attributes | Exit

Record: 1/1

Fig 6: Dynamic Knowledge Base: Update treatment planning page by deleting sensation attribute: showing the flexibility of the design of the software according to the future updates of the knowledge base

Update Lumbar Disc Data

Rule	Sensation	Pain	Sphenectric Control	Deformity	Motor Power	Treatment Planning	New1	New2	New3
1	no	8-10	yes	0-30	G1-G3	phsiotherapy	????	????	????
2	no	4-7	yes	0-30	G1-G3	laminectomy			
3	no	4-7	yes	0-30	G4-G5	medical			
4	yes	8-10	yes	40-70	G4-G5	phsiotherapy			
5	yes	8-10	yes	0-30	G1-G3	phsiotherapy			
6	no	8-10	no	40-70	G4-G5	phsiotherapy			
7	no	8-10	no	40-70	G1-G3	endoscopic			
8	no	4-7	yes	40-70	G4-G5	endoscopic			
9	no	4-7	no	40-70	G4-G5	phsiotherapy			
10	yes	4-7	yes	40-70	G4-G5	phsiotherapy			

Undiserable Attributes

Add New Attributes | Save Changes
Delete Attributes | Exit

Record: 1/1

Fig 7: Dynamic Knowledge Base: Update treatment planning page by adding new attributes according to new added discoveries
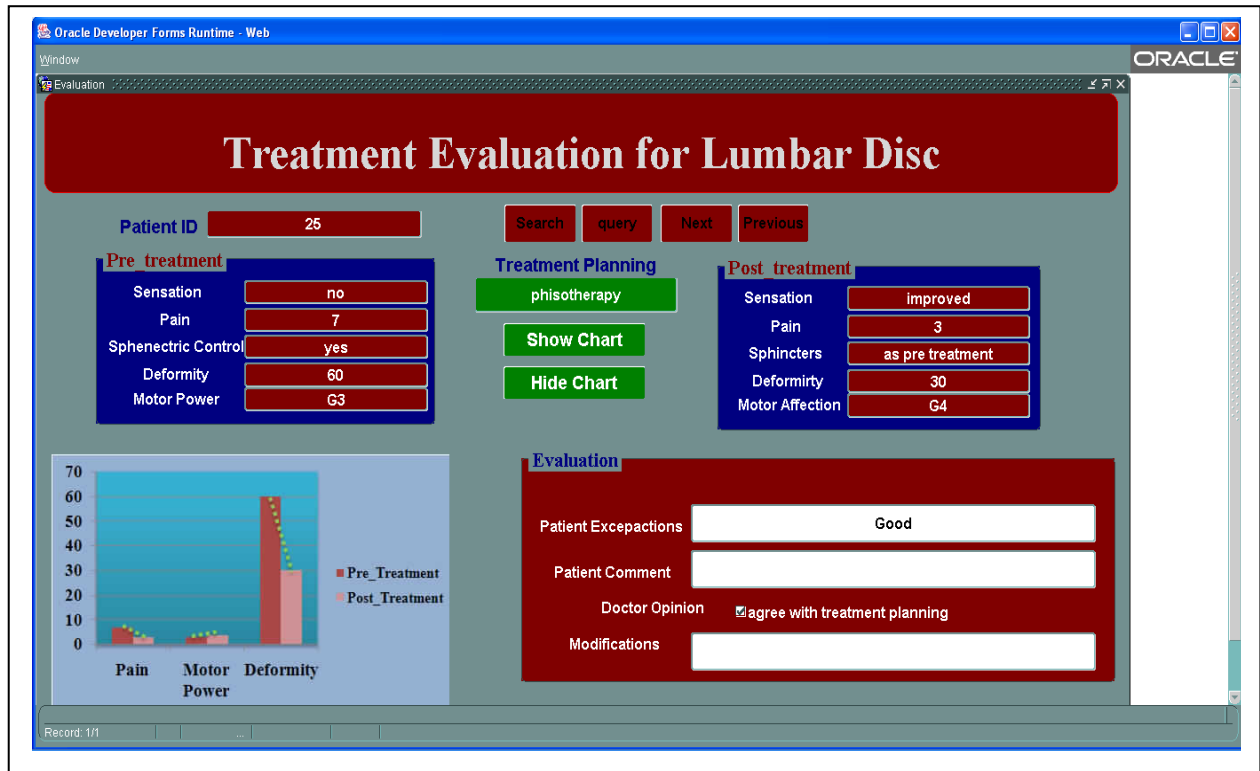
Fig 8: Physician's final decision: Confirmation of the selected treatment based on the post-treatment evaluation and the aftercare degree of patient satisfaction

6. CONCLUSION AND FUTURE WORK

This research ensures that implementation of EMR; together with Medical DSS has an important role for physicians to support their decisions that is expected to expand by time. INDBS is a new system that combines EMR system and Medical DSS. It was developed according to machine learning technique of data mining aspect of rough set which is classification of empirical data and subsequent decision making. Decision making is used in lumbar disc treatment prediction environment for the developed system and decision making in evaluating treatment planning. The developed system uses android technology to help patients in recording their opinions about treatment by selecting degree of satisfaction and record any comments, which in some times may be important notes for physicians. In addition, it uses the lumbar disc treatment predication environment to predict the lumbar disc treatment for new patients.

The developed version of the work presented in this paper may further help in other neurosurgical cases (e.g. brain tumors). Also, the number of treatment options is increasing. So, hybrid approach which used in the INDBS can be suitable in making treatment decisions. Both of Vague rough set model [23] and intuitionistic fuzzy rough technique [24] can be utilized as alternative techniques for RST in the developed versions.

7. REFERENCES

[1] Turban, E. 2001. Decision Support and Expert Systems. Management Support Systems. Macmillan Publishing Company. USA.

[2] Alhajj, R. 2003. Extracting the Extended Entity-Relationship Model from A Legacy Relational Database. Information Systems 28, 597–618. Elsevier Science.

[3] Sangjae, L., and Ingoo, H. 2000. EDI Controls Design Support System Using Relational Database System. Decision Support Systems. vol. 29, 169–193. Elsevier.

[4] Al-azmi, S., Mohammed, A., and Hanafi M. 2006 Patients' Satisfaction with Primary Heath Care in Kuwait after Electronic Medical Record Implementation. J Egypt Public Health Assoc. vol. 8, NO. 5& 6.

[5] Tomaszewski, W. 2012. Computer-Based Medical Decision Support System Based On Guidelines, Clinical Pathways and Decision Nodes. Acta of Bioengineering and Biomechanics. vol. 14, NO. 1.

[6] Ruland, C., and Bakken, S. 2002. Developing, Implementing, and Evaluating Decision Support Systems for Shared Decision Making in Patient Care: A Conceptual Model and Case Illustration. Journal of Biomedical Informatics. vol. 35, 313-321.

[7] P. Lambin, et al. 2013. Rapid Learning Health Care in Oncology – an Approach Towards Decision Support Systems Enabling Customized Radiotherapy. Radiotherapy and Oncology.

[8] Hu, X., Han, J. and Lin, T. 2004. A New Rough Sets Model Based on Database Systems, Fundamental Informatics. 1-18.

[9] Munakata, T. 2008. Fundamentals of the New Artificial Intelligence Neural, Evolutionary. Fuzzy and More, 2nd ed.

[10] Yao, T., and Herbert, J. 2007 Web-Based Support Systems with Rough Set Analysis. Springer-Verlag Berlin Heidelberg. 360–370.

[11] Hassanien, A., Abraham, A., Peters, J., and Schaefer G. 2009. Rough Sets in Medical Informatics Applications. Applications of Soft Computing. AISC 58, 23-30. Springer- Verlag Berlin Hiedelberg.

[12] Pawlak, Z., and Skowron, A. 1994. Rough membership functions Advances in the Dempster -Shafer theory of evidence,NewYork,NY,John Wiley&Sons, 251-271.

[13] Perlin, J., Kolodner, R., and Roswell, R. 2004. The Veterans Health Administration: Quality, Value, Accountability, and Information as Transforming Strategies for Patient Centered Care, The American Journal of Managed Care, vol. 11.

[14] Lee, S., and Abbott, P. 2003. Bayesian Networks for Knowledge Discovery in Large Datasets: Basics for Nurse Researchers. Journal of Biomedical Informatics. vol. 36, 389-399, Elsevier.

[15] Kuan-Liang, K., and Chiou-Shann F. 2011. A Rule-Based Clinical Decision Model to Support Interpretation of Multiple Data in Health Examinations. Journal of Medical Systems, vol. 35 (6), 1359-1373.

[16] Riad, A., El-Bakery, H., and El-Ghareeb H. 2009. Mapping different software architecture paradigms to different integration techniques: highlighting driving and restraining force for each paradigm. journal of convergence information technology, vol. 4.

[17] http://ntier.com

[18] Connolly, T., Begg, C., and Holowczak, R. 2008. Business Database Systems, Addison-Wesley.

[19] Inmon, W. 1996. Building the Data Warehouse, 2nd ed., John Wiley & Sons.

[20] Pinet, F. 2012. Entity-Relationship and Object-Oriented Formalisms for Modeling spatial Environmental Data. Environmental Modelling & Software, vol. 33,.80-91, Elsevier.

[21] Gibson, M., and Arnott, D. 2005. The Evaluation of Business Intelligence: A Case Study in a Major Financial Institution. Australasian Conference on Information Systems, Sydney, Australia.

[22] Kumar, A. 1998. New Techniques for Data Reduction in Database Systems for Knowledge Discovery Applications. Journal of Intelligent Information Systems, vol. 10(1), 31-48.

[23] Singh, K., Thakur, S., and Lal, N. 2008. Vague Rough Set Techniques for Uncertainty Processing in Relational Database Model. Informatica, vol. 19, NO. 1, 113–134.

[24] Gangwal, C., and Bhaunik, R. 2012. Intuitionistic Fuzzy Rough Relational Database Model. International Journal of Database Theory and Application, vol. 5, NO. 3.