Apply On-Line Analytical Processing (OLAP)With Data Mining For Clinical Decision Support

Dr Walid Qassim Qwaider

Majmaah University
College of Science and Humanities in Ghat
Management Information Systems Dept.
Saudi Arabia
wqwaider74@yahoo.com

Abstract

Medicine is a new direction in his mission is to prevent, diagnose and medicate diseases using OLAP with data mining. Are analyzed clinical data on patient population and the wide range of performance management of health care, unfortunately, are not converted to useful information for effective decision-making. Built OLAP and data mining techniques in the field of health care, and an easy to use decision support platform, which supports the decision-making process of caregivers and clinical managers. This paper presents a model for clinical decision support system which combines the strengths of both OLAP and data mining. It provides a knowledge rich environment that cannot be achieved by using OLAP or data mining alone.

Keywords

Medicine, Clinical decision support system, OLAP, OLTP, Data mining, diabetic approach.

1. INTRODUCTION

The healthy environment is generally seen as "informative" after a "poor knowledge" (Lincoln, 1999), and there is a wealth of data available in the health care systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data. The value of knowledge can be discovered by the application of techniques to extract the data in the health care system (Kristen, 2000). Important areas of application of this technology, health care, wellness and disease management, and support for independent living (Stanford, 2002).

The concept of warehousing in 1990. It is a set of techniques that help the managers to make better decisions. Online analytical processing (OLAP) is a key feature supported by most warehousing systems (Lin, 2002; Welbrock, 1998).

All of the data mining and OLAP are powerful tools to support decision making. However, people use them separately for years: OLAP systems focus on efficiency to build cubes OLAP, and applied to any algorithms for mining statistical / data, on the other hand, developed a

traditional statistical analysis for two-way relational databases, and have not been circulated to the structure of the multi-dimensional data OLAP. May combine data mining and OLAP to provide excellent solutions.

OLAP is completely different from its predecessor, for an online transaction processing (OLTP) systems. OLTP focuses on the automation of data collection procedure. Keeping detailed data, consistent and modern, is the most important condition for the application of OLTP (Lane, 2002). Showed the typical structure of OLAP application in the figure. 1.
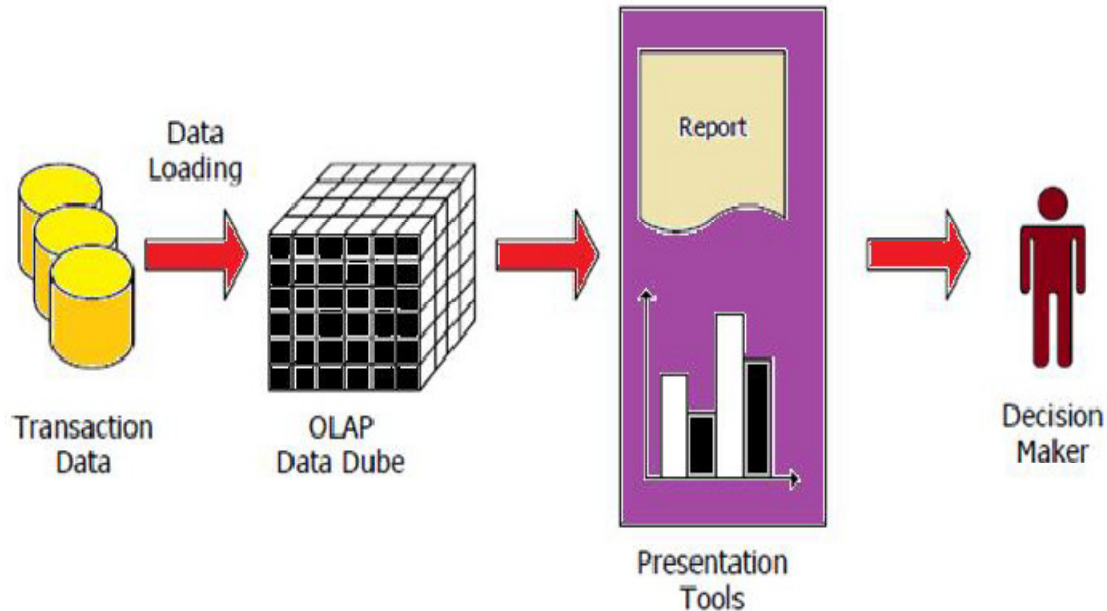

Fig. 1: OLAP Architecture (Lin, 2002)

However, in spite of OLAP is able to provide summary information efficiently, and how to take the final decision is still an art application of knowledge and common sense in some cases, the decision maker few quantitative data mining methods, such as regression or classification, and introduced to the arena OLAP. On the other hand, has been designed, most traditional algorithms for data mining dimension of the data set, and is not involved in the development of OLAP data mining algorithm (Lin, 2002). Figure 2. Shows the integrated model (OLAP with data mining), and consists of several elements. The system is divided into two parts: server side - to build an integrated model, the client side - for inquiries and for the results. And uses of OLAP (slice, dice, roll up, drill down, axis), and mining decision tree algorithm C4.5. Validates the effectiveness of the model test data (Palaniappan et al, 2008).
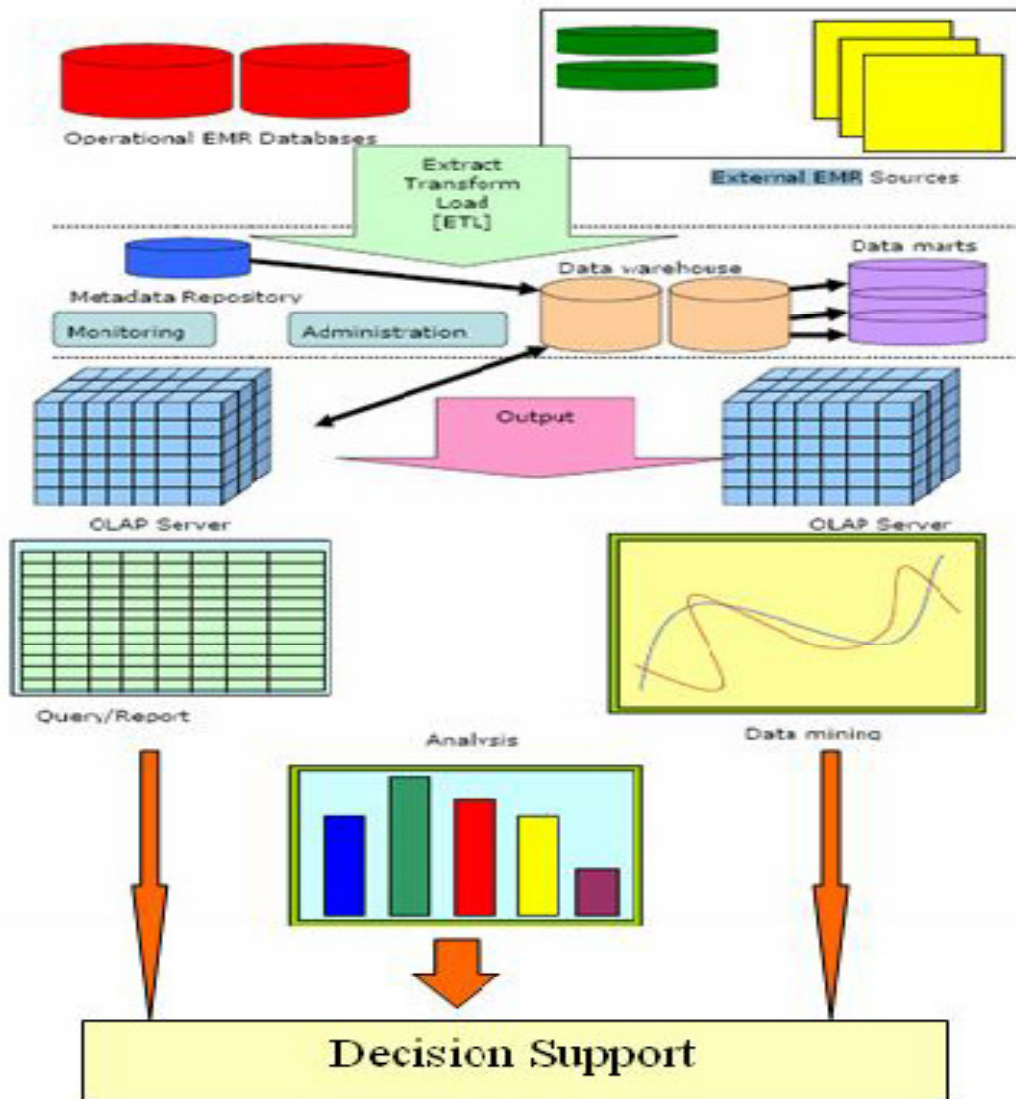
Fig. 2: Integration of OLAP with data mining architecture

Has been the development of clinical decision support systems (CDSS) to overcome these limitations. However, they still do not provide advanced features to help doctors to perform complex queries (Donald,2001; Shim et al, 2002). And advanced technologies can now generate a rich environment to determine the effectiveness of clinical decision-making process. This paper presents a model for clinical decision support system based on OLAP and data mining to solve the problem of data association.

2. LITERATURE REVIEW

Our review covered three areas: on-line analytical processing (OLAP), the application of data mining in health care, data mining and knowledge discovery in databases (KDD). Found that

organizations are generally used instead of OLTP OLAP to build the Dubai Summer Surprises (Surajit, 2001;. Sarwagi et al,1999). Trend now is to use methods combined.

(Tao, 2003) discusses and concludes by discussing the impact of wavelets on the search and data mining to identify trends and possible future research and applications. (Jonathan et al, 1997) In this study, the use of data mining techniques (also known as knowledge discovery in databases) to search for relations in a large clinical database. (Svetlana, 2007) and displays the results of an ongoing project to provide support for OLAP analysis of business processes in an innovative application for modeling operations.

(Surajit et al, 1996) provides an overview of warehousing and OLAP, with a focus on new requirements. We describe the background necessary tools for extracting, cleaning and loading data into the data warehouse, data models, multidimensional model of OLAP; Front client tools end to query and analyze data; server extensions for query processing efficiency, and tools for the management and metadata management repository.

(Sasso et al, 2001) focuses on one of the analysis conducted by a team of doctors, researchers, computer science, using commercially available on-line analytical processing (OLAP) tool in conjunction with the Royal techniques data mining developed by the researchers HAL.

(Nevena et al, 2002) to explain the integration of data warehousing, OLAP and data mining techniques in the field of health care, and an easy to use decision support platform, which supports the decisionmaking process is the process of building care providers and clinical directors. We offer three case studies, which show that the clinical data warehouse that facilitates evidence-based medicine is a platform for reliable, robust and easy to use to make strategic decisions, which are of great importance to the practice of medicine, and the acceptance of evidence-based.

(Song, 2001) may include both OLAP and data mining solutions is excellent, and this paper, we present a case in point - the OLAP - model-based method of association data. Literature in the data warehouse (Donald et al, 2001), OLAP (Helen, 2001; Robert, 1999) data mining (Fayyad et al, 1996), and the data warehouse and OLAP integration (Ralph 2002) are many.

(Parseye, 1998) examined the relationship between OLAP and data mining, and proposed structure of the integration of OLAP and data mining, and discussed the need for different levels of aggregation of data mining. The (Han et al, 1997) the need for different levels of aggregation for the analysis of the pattern and focused his work on the mining OLAP. Study focused primarily on data mining algorithms, not the integration of OLAP and data mining.

3. PROBLEM DEFINITION

OLTP systems based on relational databases are suitable for recording business transactions. They record information in two dimensions and automate repetitive tasks. I usually use Structured Query Language (SQL) to access the information, and results were presented in the form of doctors' reports that are used in making clinical decisions (Sellappan, 2008).

Shows a simple OLTP entity relationship diagram (ERD), and consists of six to analyze the

relationship between hospitals and patients are shown in Figure. 3 (a). OLTP has a big disadvantage. Large amounts of data in the form of normalization requires a lot even joins to answer simple questions.

For example, to analyze the relationships between the hospital and patients Figure. 3 (b), the table scan query requires several multi-way that could be acceded to the table of lower performance significantly (Surajit, 1997). It requires at least four internal links across five tables. A database of real life and many of the tables and the time required to process the links would be unacceptable (Palaniappan et al, 2008).
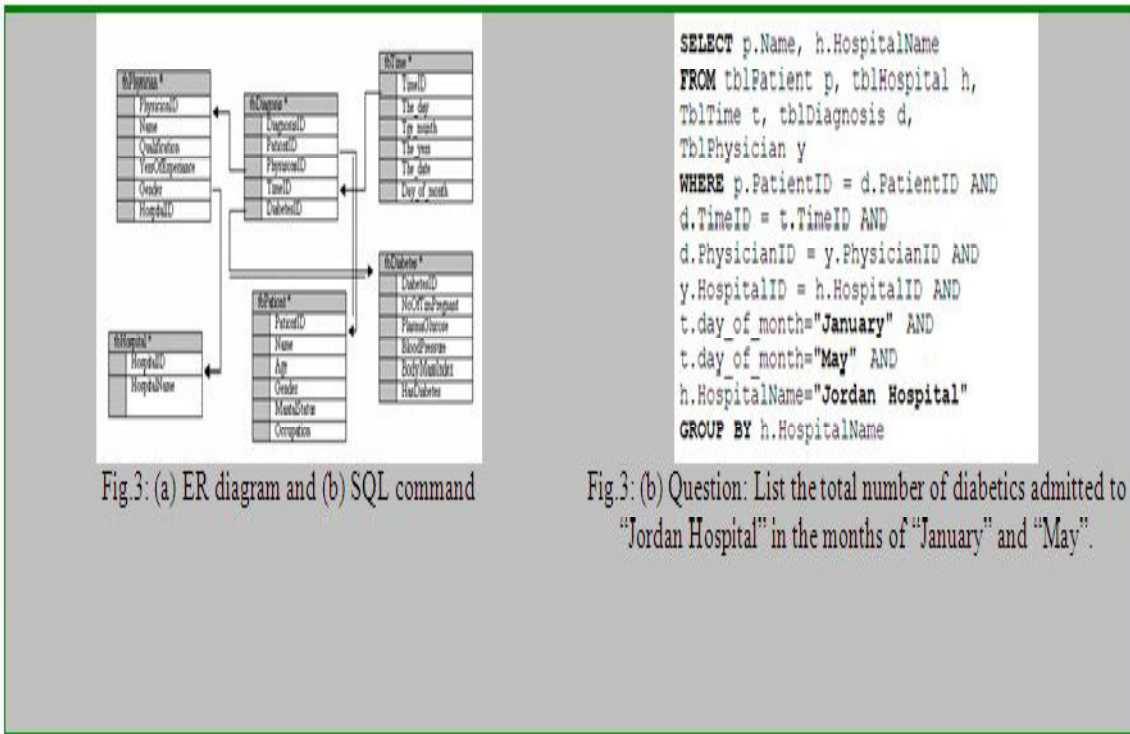

Fig.3: (a) ER diagram and (b) SQL command

Fig.3: (b) Question: List the total number of diabetics admitted to "Jordan Hospital" in the months of "January" and "May".

The OLAP to overcome this problem. While OLTP tables using two-dimensional, OLAP uses a multidimensional data tables are called cubes. The OLAP data can be analyzed in several dimensions. It also includes a query optimization (Surajit, 1997; Torben et al, 2001). Users can filter, slice and dice, drill down and roll up data to search for relevant information efficiently.

Although OLAP provides an effective way to analyze the data, it does not determine the dimensions that may be needed to respond to the specific decision problem. Also, OLAP can predict the future state based on the current properties (Palaniappan et al, 2008).

Figure 4. Shows a simple OLAP cube of eight patients under the care of Dr. 10. And can use this information to answer the question: Which of the eight patients are likely to be diabetic? OLAP and alone cannot answer this question because it does not learn from the data.

Diabetes Diagnosis				
Hospital Name: Jordan Hospital				
		Years 2008	Grand Total	
Name	Patient Name	Total Diabetes	Total Diabetes	
Dr 10	Muhammad Omar	1	1	
	Salim Ali	1	1	
	Lana Ibrahim	1	1	
	Maream Zaby	1	1	
	Istabrag Qwaider	1	1	
	Khaled Waled	1	1	
	Raida Ali	1	1	
	Ahmmad Mussa	1	1	
	Total	8	8	

Fig 4: Table representation of data in an OLAP cube

Extract data from warehousing (as defined in the process of extracting the information is not obvious, there is an implicit reference were not previously known, and can be useful from the data stored in the database) and is now an important research tool (Ming et al, 1996). It can discover knowledge hidden in data and identify problems along several dimensions (Ming, 1996). However, extracting the data alone is not very effective. For example, the information contained in the figure. 5 does not allow doctors to analyze the demographics of patients under the title: Plasma glucose more than 93 and less than or equal to 127.

This paper presents the clinical decision support system using OLAP and data mining that can answer complex questions that cannot be using OLAP or data mining alone.
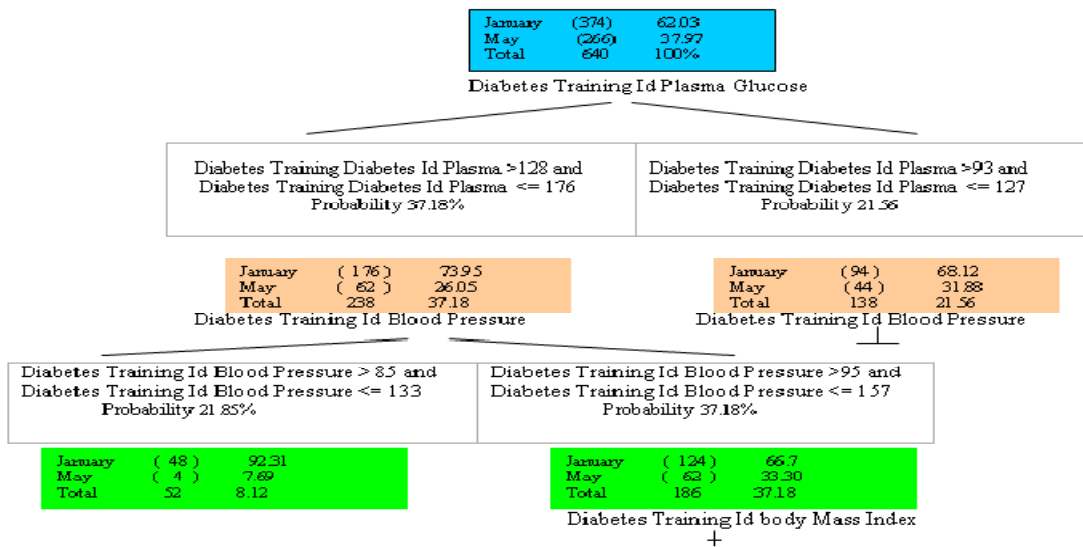

Fig 5: Decision tree mining model

4. THE QUESTIONS

This research shows, "How can an integrated approach, with data mining OLAP provides advanced decision support compared to using OLAP or data mining alone."

Research questions are used for this purpose. Cannot be answered through the use of OLAP or data mining alone, but can be answered using the integrated model. Questions about:

R1.How not OLAP with data mining in real-time indicators such as the promotion of bottlenecks?
R2. How OLAP with data mining to provide improved visualization to detect patterns / trends is likely to be true?
R3. How Data Mining with OLAP detect patterns more accurately in the data provided on the capacity By OLAP or data mining alone?

This paper describes how the integrated approach, with OLAP data mining, decision support and provides advanced compared to using OLAP or data mining alone. Research questions are used for this purpose. Cannot be answered through the use of OLAP or data mining alone, but can be answered using the integrated model.

5. SYSTEM IMPLEMENTATION

Is the first creation of the data mining cube and then began the process of data collection. The cube keep the information and allows browsers on different theoretical levels. Serves as the source of the data for the task of data mining. Can be performed to extract the data on any level or dimension of the cube. After building a model is stored in a cube OLAP. Each representing a dimension of the rule corresponding to the node in the decision tree mining model Fig.7. OLAP operations explain the different states of the system.

Data are taken for this study from the UCI repository machine learning databases (Blake, 1998). Data from a database with Jordan Hospital for Diabetes, kidney colic base. Also be declassified in the data, we have added many features and information such as a fake doctor and the patient and the hospital.
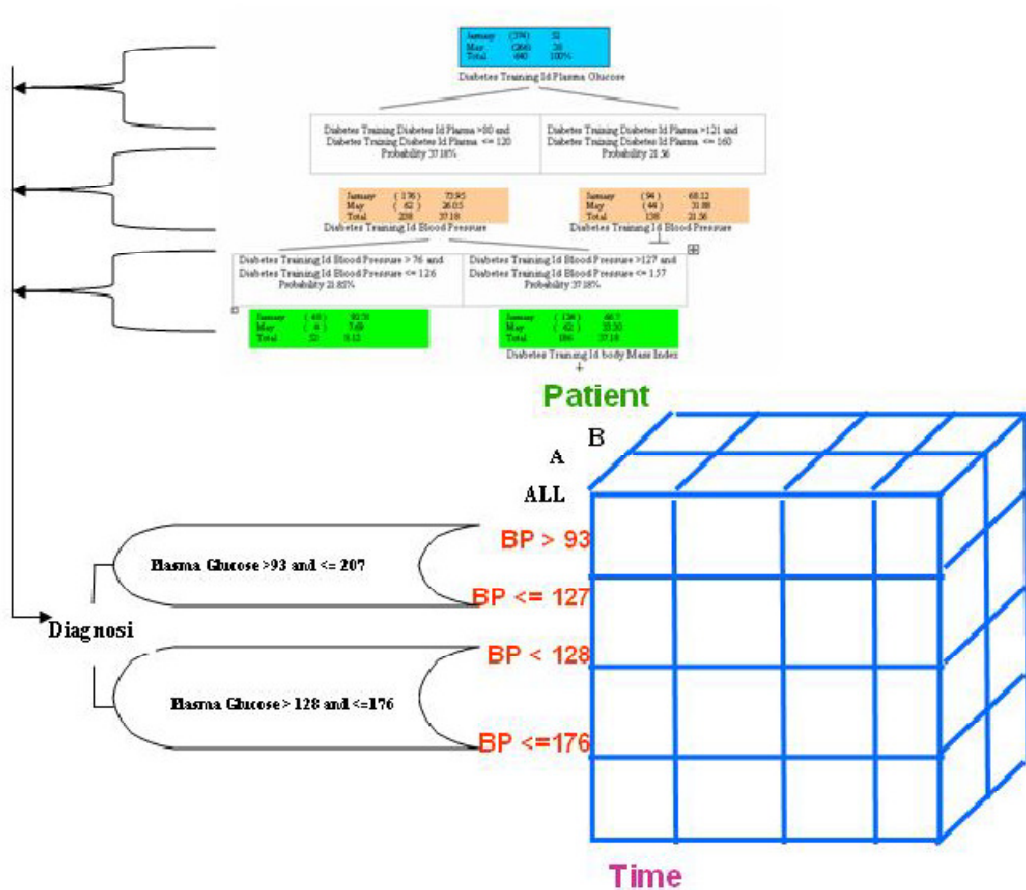
Fig 7: A logical view of representation of decision tree mining model in an OLAP cube

6. FINDING AND DISCUSSION

System can predict the future state, and generate useful information for effective decision making. Can answer all research questions.

R1. Integrated model enhances the indicators in real time using information on the use of a room at Shifa Hospital for patients after surgery. It allows officials at the hospital to discover any bottlenecks that may exist. It allows them to solve problems related to using the hospital room. Results in Figure 8. Show that a total of 8 patients are likely to be discharge from the hospital, Jordan.

An administrator can use this information to customize the rooms on the basis of their characteristics. For patients with more than 60 years, may be a decision to carry out in the older wing, or transferred to another hospital or the best of their house. This indicator is useful to perform "what if" analysis on the availability of the hospital room.
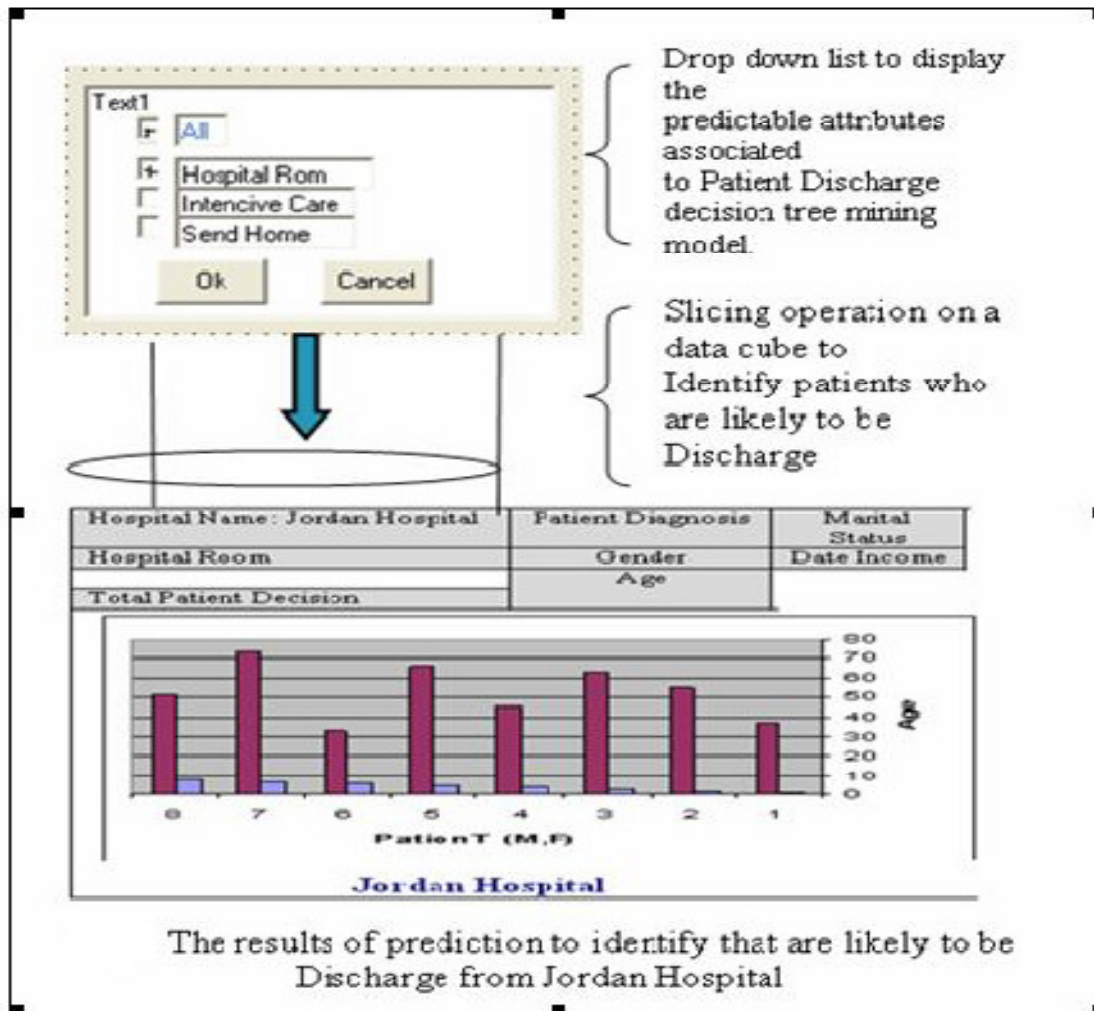
Fig 8: The results of prediction to identify that are likely to be Discharge from Jordan Hospital

R2. Integrated model improves the visualization of information. He discovers the general trends that are likely to be missed by the use of OLAP or data mining alone. Figs. 9 (a) shows that all states in the Jordan, and the largest number of patients diagnosed with renal colic. Apart from this, the best specialists renal colic are from Jordan Hospital shape. 9 (b). The report provides an analysis of more comprehensive and easier decision-making process through the allocation of doctors to under-represented geographic areas. It allows to improve the quality of doctors in the areas of representation.
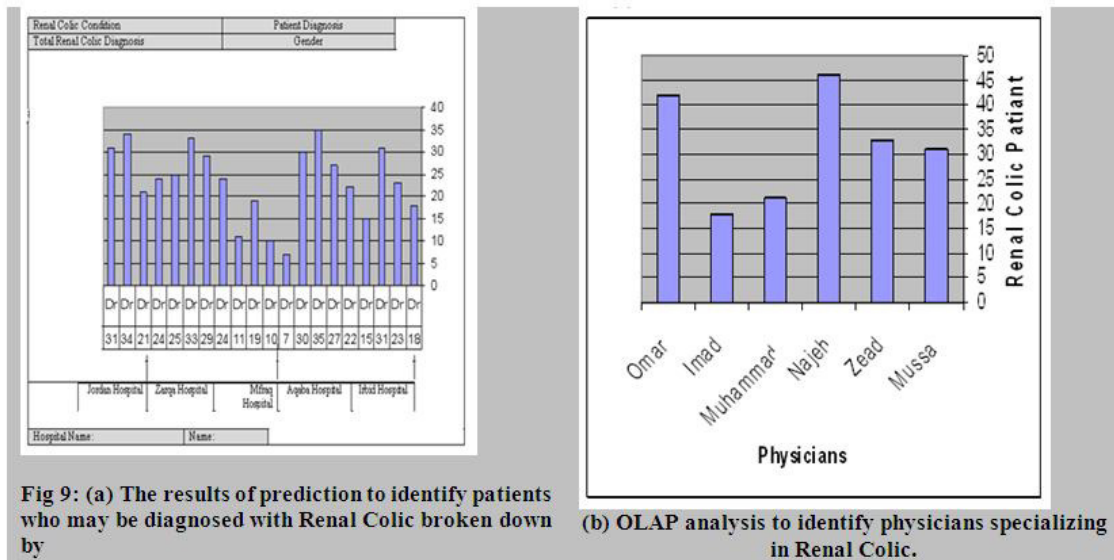
Fig 9: (a) The results of prediction to identify patients who may be diagnosed with Renal Colic broken down by

(b) OLAP analysis to identify physicians specializing in Renal Colic.

R3. With data mining, and doctors can predict patients who may be diagnosed with diabetes. OLAP provides the answer focused using historical data. However, by combining, we can improve the current operations and to detect patterns more accurately, for example, by analyzing the demographics of patients.

Fig 10. Show that patients who have plasma glucose is more than 154.25 to have the highest probability to be diagnosed with diabetes compared with those patients who have plasma glucose between 127.25 and 154.25 (Palaniappan et al, 2008). It also shows that patients who are single females are more likely to have diabetes compared with male patients.

Interestingly, for patients who have reached a certain stage in the process of care, and doctors can (endocrine) recommends that it be subject to examination of the retina. This can help reduce or prevent blindness before it reaches the critical stage. System that allows users to implement advanced data analysis and queries, custom ad hoc queries. It also generates reports in multiple formats.
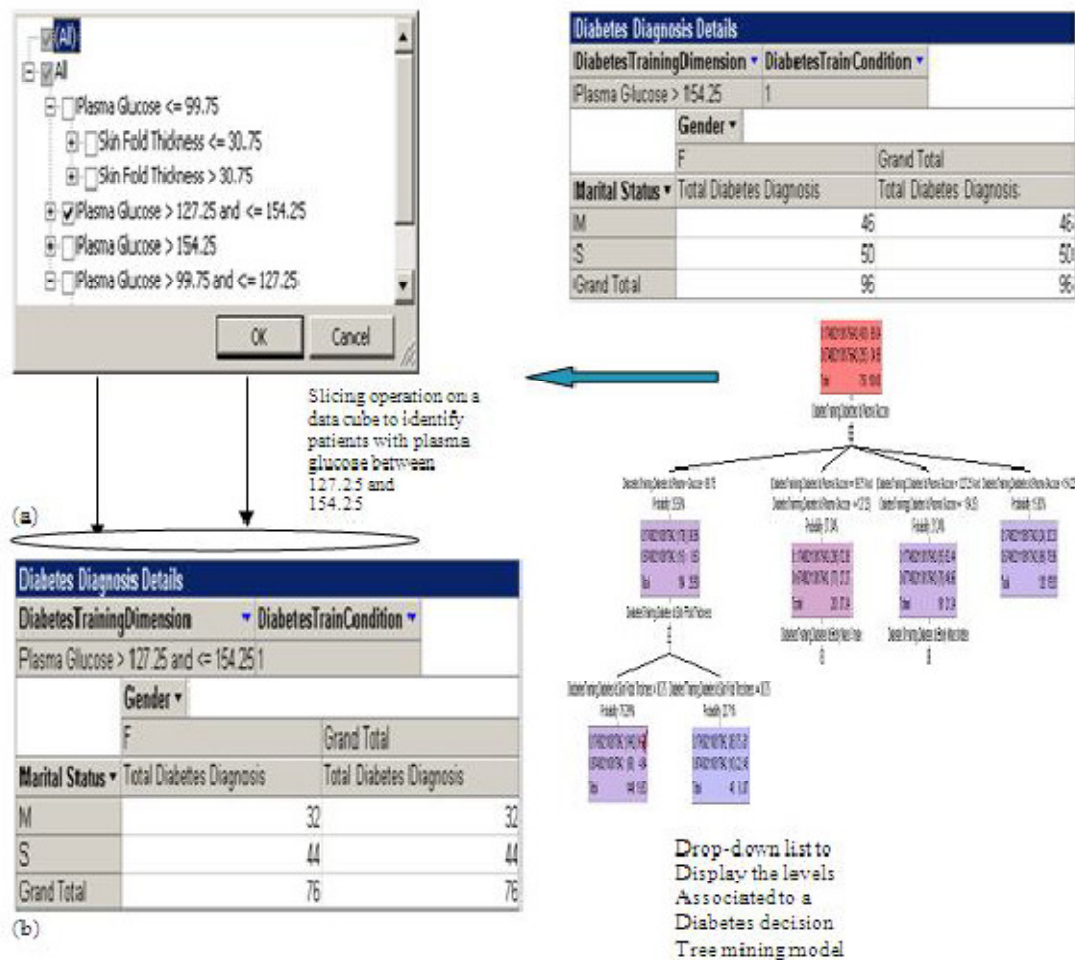
Fig 10: Comparison results between patients whose plasma glucose (a) between 127.25 and 154.25 and (b) plasma glucose more than 154.25 (Palaniappan, et al. 2008)

7. CONCLUSION

In this paper, we provide a new form of association data. This model combines all of the concepts and OLAP mining model the idea of disclosure. Function is defined and the degree of outlier cells and an OLAP cube that measures the extremeness of the cell. Also, the CDSS based on OLAP with data mining. Cannot be answered through the use of OLAP or data mining alone, but can be answered using the integrated model. Questions about the system strong for the following reasons:

• It discover hidden patterns in data.
• It enhances the real-time indicators and discover bottlenecks.
• It improves the visualization of information.
• Combining OLAP and data mining provides a potentially powerful tool to solve a real problem the world.

May also include a model complex data objects, and spatial data and multimedia data. Along with the decision tree, and can also use other techniques can explore and extract the data.

FUTURE WORK

Can be done to strengthen the system. For example, you can add features to allow doctors to query the data cubes on trade issues and translate automatically to these questions are multi-dimensional expression queries. May also include a model complex data objects, and spatial data and multimedia data. Along with the decision tree, and can also use other techniques can explore and extract the data.

REFERENCES

Book

[1] Blake, C.L. & Merz, C.J, 1998. UCI Repository of Machine Learning Databases, University of California, Department of Information and Computer Science.

[2] Robert, S.C et al, 1999. Microsoft Data Warehousing: Building Distributed Decision Support Systems, London: Idea Group Publishing.

[3] Ralph, K. and Margy, R, 2002. The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling (2nd ed.). Canada: John Wiley & Sons, Inc.

Journal

[4] Donald, J.B. et al, 2001. Healthcare Data Warehousing and Quality Assurance. IEEE Computer, pp56-65. Helen, H. and Peter, H, 2001. Using OLAP and Multidimensional Data for Decision Making. IEEE IT Professional, 44-50.

[5] Lin, S and Brown, D, 2002. Outlier-based Data Association: Combining OLAP and Data Mining. Department of Systems and Information Engineering University of Virginia, Charlottesville, VA 22904, SIE 020011, December, 2002.

[6] Ming-Syan, C et al, 1996. Data Mining: An Overview From a Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866-883, December.

[7] Torben, B.P. and Christian, S.J, 2001. Multidimensional Database Technology. IEEE Computer, 34(12), 40-46.

[8] Surajit, C. et al, 2001. Database Technology for Decision Support Systems. IEEE Computer, 34(12), 48-55.

[9] Stanford, Vince Stanford, 2002. Pervasive Health Care Applications Face Tough Security Challenges. IEEE Pervasive Computing.

[10] Palaniappan, S, and Chua, S, 2008. Clinical Decision Support Using OLAP With Data Mining. VOL.8 No.9, September.

[11] Shim, J.P. et al, 2002. Past, Present and Future of Decision Support Technology. Elsevier Science B. V., 33, 111-126. Tao, L. et al, 2003. A Survey on Wavelet Applications in Data Mining. Computer Science Univ. of Rochester. Volume 4, Issue 2 - page 49.

[12] Nevena Stolba and A Min Tjoa, 2002. The relevance of data warehousing and data mining in the field of evidencebased medicine to support healthcare decision making. ESF, under grant 31.963/46- VII/9.

[13] Song Lin & Donald E. Brown, 2001. Outlier-based Data Association: Combining OLAP and Data mining. Department of Systems and Information Engineering University of Virginia, Charlottesville, VA 22904.

[14] Svetlana, M. et al, 2007. LAP and Data Mining: Bridging the Gap. Database Programming and Design. DaWaK 2007, LNCS 4654, pp. 111–122.

[15] Parseye, K, 1998. OLAP and Data Mining: Bridging the Gap. Database Programming and Design, 10, 30-37.

[16] Saso, D. et al, 2001. Using Data Mining and OLAP to Discover Patterns in a Database of Patients with Y-Chromosome Deletions. Jozef Stefan Institute, Jamova 39, Ljubljana, Slovenia.

[17] Welbrock, P.R., Strategic Data Warehousing Principles Using SAS Software, Cary, NC: SAS Institute inc., 1998.

Conference paper or contributed volume

[18] Fayyad, U. et al, 1996. From Data Mining to Knowledge Discovery in Databases. AI Magazine, 37(3), pp 37-54.

[19] Jonathan, C.P. et al, 1997. Medical Data Mining: Knowledge Discovery in a Clinical Data Warehouse. Proceedings of the American Medical Informatics Association Symposium. Philadelphia, United States of America, pp101-105.

[20] Han, J, 1997. OLAP Mining: An Integration of OLAP with Data Mining. Proceedings of 1997 IFIP Conference on Data Semantics (DS-7). Leysin, Switzerland, pp 1-11.

[21] Han, J et al, 1997. DB Miner: A System for Data Mining in Relational Databases and Data Warehouses. Proceedings of the 1997 Conference of the Centre for Advanced Studies on Collaborative research. Ontario, Canada, 1-12, November.

[22] Sarwagi, S. 1999. Explaining Differences in Multidimensional Aggregate. Proceedings of the 25th International Conference on Very Large Data Bases. Scotland, United Kingdom, 42-53.

[23] Surajit, C. and Umeshwar, D, 1996. An Overview of Data Warehousing and OLAP Technology. VLDB Conference.

[24] Surajit, C. and Umeshwar, D, 1997. An Overview of Data Warehousing and OLAP Technology. ACM Sigmod Record, 26(1), 65-74.

[25] Han, J. and Kamber, M. 1997. Data Mining Concepts and Techniques. San Diego, USA: Morgan Kaufmann Publishers, pp. 294- 296.