ISSN 2456-4400

Ahmed et al., Int J Med Lab Res 2017, 2(2): 60-76

REVIEW ARTICLE INTERNATIONAL JOURNAL OF MEDICAL LABORATORY RESEARCH (IJMLR)

APPLICATION OF DATA MINING AND KNOWLEDGE MANAGEMENT IN SPECIAL REFERENCE TO MEDICAL INFORMATICS: A REVIEW

Yogita Gupta[1], Rana Khudhair Abbas Ahmed[2]*, Dr. Sandeep Kumar Kautish[3]

[1]*Assistant Professor, Aryabhatta Institute of Engineering & Technology, Barnala Punjab, India*
[2]*Alrafidain University College, Baghdad, Iraq*
[3]*Professor – Computer Science, Guru Kashi University, Talwandi Sabo, Bathinda Punjab, India*

Received:11 May, 2017/Revision:29 May, 2017/Accepted:15 July, 2017

ABSTRACT: In recent times, while India is moving towards Digital Revolution, Data Mining and Knowledge Management are two of the areas which attract many researchers as both have high potential in terms of developing new techniques to be applied in different domains of human life. Medical science is one of the areas where new inventions and developments are coming up as the results of integration of Data Mining and Knowledge Management tools. The aim of this paper is to present detailed survey of available literature on recent advancements and notable contributions in the field of applications of Data Mining and Knowledge Management tools especially focused on Medical Informatics. After presenting introduction about the aim and scope of the topic, section two of the paper reestablish the concepts of Data mining and its conventional techniques i.e. Probabilistic & Statistical Models, Rule Induction, Neural Networks and Analytical Learning and the section ends with presenting Knowledge Management concept and its linkage with Data mining and Medical Science field. In section three, all the previous relevant works of Data mining and knowledge management are critically analyzed, explained and categorized on the basis of their applicability which is followed by section four which presents discussion on all the previous works and highlight the advantages and disadvantages of various methods and tools with further scope of future research and limitations. The author concludes the paper while emphasizing on security and privacy concerns of medical data and attract readers' attention towards validation of medical data which is used for medical judgments and decisions.

KEYWORDS*:* Knowledge Management, Data Mining, Medical Informatics, Informatics, Medical Decisions

INTRODUCTION:

Biomedical field is one of the areas from where huge data is being generated on daily basis. The digitization of critical data of biomedical field in form of various reports, anatomic images and patient records has resulted into large amount of data. With the advent of latest computer technologies in databases, we are capable to store such large amount of data at various public and private repositories. However, only storage of data has no use if it is not stored in organized manner as per right context. Conventional techniques of data storage were not capable to manage mass data in efficient manner because of their failure in ensuring the availability of right data on right time.

Corresponding Author:
Rana Khudhair Abbas Ahmed
Alrafidain University College, Baghdad, Iraq

www.ijmlr.com/IJMLR© All right are reserved

Therefore Data Mining comes forward in order to store the huge data categorically and make it available as per requirements. Data Mining can help us out in terms of data storage and retrieval but for once again data itself has no value until we perceive some useful information from the data which is called Knowledge Discovery in Databases (KDD). It helps us to understand complex physical, structural or biological behavior of real world entities that the data belongs to. Knowledge Discovery is one of the aspect of Knowledge Management which deals with effective and efficient access and use of data which helps the perceiver in decision making process and ultimately ensures that the right decisions are made in a given situation. Though we have briefly discussed Data Mining and Knowledge Management, but it's necessary to have a complete understanding about both of them. In the following sections, we will explain about background of Data Mining and Knowledge Management and their emerging collaboration specifically in the field of biomedical and medical informatics.

The key motivation behind the study is the nature of data mining and knowledge tools which supports decision making and are being widely adopted in variety of application areas i.e. military, marketing, banking, education, disaster management, weather forecasting and many more. Authors made an attempt to draw readers' attention on the worthy role and significant presence of data mining and knowledge management tools in numerous fields of medical science. Such tools are not only significantly useful in diagnosis of many diseases like hepatitis, diabetes, breast cancer, skin cancer, lung cancer, kidney failure, kidney stone, heart disease, liver disorder but also in providing forecasting information which helps in curing them. Altogether data mining and knowledge management tools in combination, facilitates prevention of various possible errors.

MATERIALS AND METHODS:

Data mining:

The field of Data Mining is about 25 years old and the very early researchers of the field advocated the need of new technology and tools because conventional tools of data storage and management were unable to handle mass data in desirable way and additionally, their productivity is too low in comparison of their cost. Data mining aims to analyze set of given data or information in order to identify novel and potentially useful patterns (Fayyad et al., 1996). In a very precise manner, we can define Data Mining as the process of discovering the patterns, associations or relationships among data by the use of analytical techniques which results into the creation of a model which will enable to achieve knowledge.

In broader context, Data Mining has two categories, unsupervised and supervised. As the name suggests, unsupervised techniques refers to category in which there is no guidance variable i.e. like hypothesis or any model. Clustering is one of such methods which can be called unsupervised tools. On the other hand, supervised models are tools in which a model is built prior to analysis which is followed by application of data into model for in order to estimate the parameters of model. Regarding biomedical and clinical data, the best part of Data Mining algorithms is that they are predictive i.e. they can learn from past examples. Therefore resulting model presents formalized knowledge which guides good diagnostic options. Few of most popular data mining techniques are given below -

Probabilistic and statistical models

The origin of probabilistic and statistical analysis techniques belongs to 1990s. The statistical analysis keeps and utilizes data analysis and knowledge discovery goals which are quite similar to machine learning methods but the inception of former technique is different from AI research. There are a few popular techniques which are used in biomedical data analysis i.e. time series analysis, multi-dimensional scaling, regression analysis and principal component analysis. These techniques are often referred as most prominent and reliable techniques in comparison of latest machine learning techniques especially in the field of medical data analysis.

In the context of biomedical analysis, the Bayesian Model is considered as most superior model in the field. Having the roots lying in pattern recognition (Duda and Hart, 1973) the Bayesiam Model works on storing the probability of each class, each feature and each feature of each class on the basis of training data and such probabilities are further used to classify new classes. Each class usually has equally independent features and the concept is known as Naïve Bayesian model which is one of the variations of Bayesian model. Naïve Bayesian model has been accepted in different domains on the basis of its simplicity in terms of easy to understand (Fisher, 1987; Kononenko, 1993). In the typical biomedical data mining, the Bayesian model has been used for analyzing the microarray and genomic analysis because it is mathematically rigid and moreover sophisticated in terms of modeling and its implementation. Therefore, it is quite simple and understandable model.

Out of all machine learning techniques, the support vector machines (SVMs) model is very popular in recent times due to its powerful mechanism which is based on statistical learning theory in which two or more classes gets separated with the help of hyperplane (Vapnik, 1998), which is being found from statistical analysis of features of different classes. Several biomedical applications such as biomedical classification problems are used in this model and receives positive results, for example disease state classification or medical diagnosis which are based on genetic variables or patient indicators respectively. Other than this model, SVM model is also used in document classification for its best performance among several learning methods (Joachims, 1998; Yang and Liu, 1999).

Symbolic learning and rule induction

There are number of ways available by which learning can take place but people always choose the easiest one by which they can understand the concept without much efforts. Symbolic learning is one of similar kind of techniques which have been proved very effective in learning process. Symbolic learning is the key process in knowledge data discovery (KDD) and medical data mining. Learning by examples, learning by words being told, rote learning, learning from discovery (Carbonell et al., 1983), these are few common types of symbolic learning. Learning by example got emerged in recent times as one of the most effective tool in the process of symbolic learning specifically in the field of knowledge management and data mining. In learning by example approach, the different classes are defined with different training examples in better manner by the use general concept description. The general concept descriptions are can be generated by using pattern identify algorithms. Patterns are essentials in order to generate general concept description. Couples of indentifying techniques are being used by number of

learning algorithms for recognizing various patterns. In this contrast, few of most popular algorithms are Quinlan's ID3 decision-tree building algorithm (Quinlan, 1983) and decision-tree building algorithm (Quinlan, 1983). A decision tree can be generated in ID3 as per given set of objects which is followed by classification of decision tree which ensures that all the given objects are in proper and right manner. The algorithm executes and produces minimum entropy by which the algorithm can successfully identify the attributes from each step. These attributes further enable us to classify all the objects into different classes which ensure that the ambiguity of information is minimized. This process results into creation of a set of production rules or decision tree.

Although the effectiveness of SVM or neural networks is better than symbolic learning techniques but generally it bears negligence from biologist or physicians due to their "black-box" complex mechanism. On the other hand, the Symbolic learning techniques are widely popular due to their characteristics of results which are understandable and easy to interpret too.

It is essential in biomedical or bioinformatics that the data interpretation techniques are easy to understand and efficient in application and results. It attracts many practitioners and researchers to utilize such techniques for the benefits of society.

Neural networks

An artificial neural network is an abstract of model or graph used in computers which consists of dynamic nodes which are interconnected using weighted links (like neurons from as actual human nervous system). It is equivalent to human nervous system and aspires to attain human like performance. The approach of representation of knowledge can be considered as key difference between symbolic learning and neural networks. Specifically in neural networks, the interconnected neurons, weighted synapses, and threshold logic units (Rumelhart et al., 1986) emphasized the knowledge depiction in comparison of other techniques i.e. decision tree and production rules which are primarily used in symbolic learning techniques. The unknown objects or examples can be classified by learning algorithms. The adjustment of connected weights in the networks is important for categorized unknown examples correctly. In order to retrieve various concepts and knowledge from the network, activation algorithms can be utilized over the nodes (Chen et al., 1995). Several computational models of neural networks have been developed over past years but the most widely model is feedforward or backpropagation. In the feedforward model the information propagates only in forward direction from input layer to output layer through hidden layer (if present) (Rumelhart et al., 1986). There is no loop present in this model so that propagation of information (learning example) through nodes produces a predicted output.

Analytic learning and fuzzy logic

In analytical learning, modeling and representation of knowledge is primarily based on logical rules and reasoning which is achieved by applying either fuzzy logics or other similar techniques. On the basis of logical rules and reasoning, analytical proofs are generated. Analytical problems needs such complex rules (compiled set of proofs) in order to minimize the exhaustive searching process. Analytic learning represents knowledge as logical rules and performs reasoning operations onto such rules for searching related proofs. Further

proofs can be compiled in form of complex rules for future which can be used to solve similar problems with a smaller number of searches required. For an example, if we wish to improve the speed of a parsing System, (Samuelson et al., 1991) proposed the use of analytic learning for representing grammatical rules. In real world and conventionally, there is no difference in between values and classes of analytic learning systems, because traditional analytic learning systems are based on complex computing rules. The fuzzy systems are the ideal solution of such kind of situation. Fuzzy logic and fuzzy system operates the range of real number system from 0 to 1 (Zedah, 1965) which suggests the use of values of true or false. In various new applications, imprecision and approximate reasoning are also performed by fuzzy logic principles. There is necessity of analytical learning and fuzzy logic in biomedical knowledge discovery.

Big data techniques

Though Big Data concept is not very new, now days it is being used as major buzz in IT industry. Healthcare is one of area in which Big Data technologies has brought huge impact. It is witnessed that three Vs of data i.e. velocity (pace of generation of data), variety, and volume (P. Zikopoulos, 2011) are a native aspect of the data which is produced by Healthcare field. As per McKinsey Global Institute, if US healthcare sector were to use big data effectively then the sector could create more than $300 billion in value every year. Image Processing, Signal Processing and Genomics (Ashwin Belle, 2015) are the major areas where Big Data technologies are widely used in recent times.

Hybrid approach

The hybrid approach in Knowledge discovery is become much admired in now days because several practical biomedical knowledge management and data mining system take up hybrid approaches. The reasons for differentiating paradigms are "more historical than scientific pointed by (Langley and Simon 1995). Many systems have been built to combine different approaches because of indistinguishable limits of different paradigms.

Knowledge management

Researchers (Davenport et al, 1998) have advocated that Knowledge Management is an effective tool to understand various dimensions of organizational performance and productivity. Although there is no universally accepted definition of knowledge but few core characteristics are identified as knowledge is abstract in nature, inferential and essential for decision making. Healthcare data has enough potential for generating massive knowledge and emerging tools of Data Mining and Knowledge Management has served the purpose in effective manner in past one decade.

Knowledge Management is a contemporary philosophy which has successfully utilized in businesses in recent years. Actually it is an extension of three basic trends –

- The expansion of globalization of business world which enforces industry to work towards trying to gain competitive advantage over their competitors and to learn how to make products "better, cheaper and faster".
- The widespread use of digital information-texts, audios, videos and pictures with the help of Internet

which makes all such information available 24*7 in desirable form.

- The increasing complexity of business processes which is multiplied by increasing customer demands and expectations which compels businesses to deliver "the right information at the right time in right manner".

Knowledge Management came into existence in late 1990s which inspired business houses towards applying many revolutionary changes in all business processes i.e. production, supply chain, sales, accounts and finance and even human resource management also. These changes were related to acquire, store, process, interpret and utilize the DATA residing at various levels (including vertical and horizontal) of the business units. Soon the popularity of Knowledge Management got viral and many big business houses i.e. IBM, Shell, British Petroleum started using the same and all of these companies gained huge profits from out of it. Now Knowledge Management is an established discipline that focuses on collaboration; collaboration among people, processes and technologies in such a way that results into "knowledge" which is the driving force of goals of the business.

Specifically in the context of Medical Informatics, the use of knowledge management has been increased drastically since past two decades as a result of heavy investments towards improvement in the quality of health care services. Though this is not sufficient, especially when 60% of world population is still not in position to utilize such advancements in medical science due to poverty, but it is a fact that knowledge management has served as major driving force in recent advances in medical sciences. The use and application of knowledge management in medical informatics can be categorically divided in few major areas. Content Management in first of these in which medical reports, prescriptions and lab reports are the contents which can be effectively managed by the use of knowledge management. It can serve as centralized "library" having various "layers" of information which can disseminated by various professionals working at different levels in medical fields like doctors, persons working in medical labs, medical data experts and researchers. Knowledge Transfer is another tool of knowledge management which is based on knowledge sharing which can change the behavior and outcomes of processes. Knowledge sharing is the simple process of diffusion of innovation within the processes of organization which can be referred as knowledge transfer which will impart the adoption and implementation of critical success factors. The results received after this knowledge transfer process is the next area which needs to be assessed in order to get maximum benefits of Knowledge Management. It is referred to as tracking the results in terms of outcome measures, process measures and satisfaction measures which measures the performances and interpretations of all three measures.

RESULTS & FINDINGS

In this section we have explained all the available and studied literature while highlighting its applicability, uses and limitations (if any). One comparative and Summarized tabular statement has also been given of various recent advances in developments in biomedical field which covers almost all notable developments since from 2000 to present.

Data mining techniques

Prather et al. (1997) used Knowledge Data Discovery (KDD) to indentify factors for improving various parameters of prenatal care of obstetrical patients. Breault et al. (2002) used CART tool onto the data warehouse of diabetic patients and identified factors affecting diabetics. They were the first to discover that younger age predicts bad control over diabetics. Su et al. (2006) also presented a model for diabetic control by using data mining tools. Table 1 also shows one of the researches of similar type. Wilson et al (2004) applied KDD technologies in pharma-covigilance situations for detecting signals earlier than other conventional methods. Lian et al (2003) presented a research about prescription probability and they correlated the specified prescription with a preference function based on the preferences of patients in prior clinical experiences. It was an excellent example and application of probability theory which resulted into dose optimization framework. Susan and Warren (2000) proposed conditional probability (CP) model for optimizing the drug lists with the application of discriminant analysis and multiple linear regression. It demonstrated relationship between diagnosis and medication and proposed a posterior probability based on priori probability where first refers to what medication is needed and the later one refers to what diagnosis has made. This approach is quite similar to Warren et al (1998).

Emerging concepts

Many researchers combined artificial intelligence and data mining techniques together and proposed improved models for decision support in clinical situations. Huang (2007) combined Case Based Reasoning (CBR) and data mining for finding a solution model for supporting the chronic disease treatment and proposed a model named chronic prognosis and diagnosis (CDPD) system. Kuo et al (2007) discovered useful rules from different clusters of insurance databases by integrating cluster analysis and association rules mining techniques.

While medical information systems are getting widespread, various types of medical data are emerging including text, documents, audio, speech, hypertext, graphics and images, etc. Similarly new advents in Internet era is another dimension which contributes towards the expansion of versatility of types of data.

- Multimedia data is one of type of data which is considered next generation data mining techniques. Text Mining, image mining, web mining and video mining also part of same category in which many researches are either going on and have been conducted in recent time. Other than medical images and signals, unstructured free texts are highly prominent area in which researchers pursuing their research because it is difficult to interpret such data. The process of extracting useful information and knowledge from textual documents or data is called Text Mining. For example, we can utilize text mining techniques in order to search text related to causes of diabetes in different text documents. Cohen and Hunter define text mining as "*the use of automated methods for exploiting the enormous amount of knowledge available in biomedical literature*" (Cohen and Hunter, 2008). Semantic parsing and hidden vector state model (Zhou et al., 2006) can be used to mine the text on given unstructured text.
- Image mining is another area where huge medical data is available because

medical procedures employ images as most preferred tool for medical diagnosis. Mining work onto images is totally different phenomena as images are most complex data structures in comparison of other types i.e. text and documents. In broader sense, most of the images mining techniques focus on searching, retrieving and comparing the features of similarity and differences of query image and the stored images. One of classic research in image mining presents the use of different data mining techniques for tumor classification in digital mammography (Antonie et al. 2001) and the findings of research suggests that associate rules can provide better results as compared to neural networks. Another research introduced a concept of image sequence similarity patterns (ISSP) (Zhu et al., 2003) for scanning brain images in order to discover the possible Space-Occupying Lesion (PSO). 3D informatics is becoming very popular since last few years and recent advancements in image enhancement techniques have provided support in same. Basically 3D informatics refers to gathering, manipulating, classifying, storing, retrieving, navigating, presenting and displaying extensively complex multi dimensional data. Such representation may include different dimensions which may be independent or dependent both and these dimensions are position, time and scale. 3D medical informatics is relatively very new discipline as compared to any other mining technique. It was started in 20^{th} century when Roentgen discovered x-ray and used them successfully to imaging human body. CAT Scan, tomography, MRI scan came in the latter half of the century and now 3D scanning has presented a refined way to look into human body with multi dimensional view which enables medical practices more effective and efficient.

- Another field which is very much expanded with recent advancements in digital revolution is audio and digital TV. Researchers are encouraged to discover and reveal useful patterns in the video contents and technically, this field is known as video mining. It is observed that technical persons prefer to use camera in order to capture each operation and it leads to good opportunity for applying data mining techniques while using video retrieval methods. One of empirical research in video mining presents video database management framework (Kohavi et al., 2002) for video structure and event mining.
- World Wide Web is undoubtedly largest database that ever exists. Hence Web Mining can be considered as biggest source for medical records and data. More than 80% of such data is available over web in form of electronic documents which widespread and freely accessible for everyone. The only difficulty in using this huge data for the process of knowledge discovery is dynamic nature of web pages i.e. contents of web pages are constantly changed. Technically defined, web mining is a process of using data mining techniques to automatically retrieve and extract information from Internet for the process of knowledge discovery.

Below given table provides a comprehensive comparative statement of research and development in Medical Informatics with the use of data mining tools and Knowledge Management concepts.

Table 1: Comparative Statement

Reference	Summary	Limitation, if any
Megalooikonomou et al. (2000)	Introduced Statistical methods for knowledge discovery of patterns and associations between clinical data and images	Compatibility problems in types of images and their uses
Brossette et al. (2000)	Introduced Data Mining Surveillance System (DMSS) which discovers useful patterns related to infections and antimicrobial resistance from clinical data	No support for Images and their interpretations
Antonie et al. (2001)	Classification of Medical Images and detection of anomalies	
Coulter et al. (2001)	Presented and discovered relation between antipsychotic drugs and myocarditis	Work was limited up to specific kind of drugs
Li et al. (2004)	Discovered cancer detection method with feature detection methods and to compare obtained results in order to detect protein patterns	
A.M. Wilson et al. (2004)	Detection of medical signals earlier using pharma-covigilance methods	
Delen et al. (2005)	They developed prediction models on breast cancer using a dataset by combining two data mining algorithms (artificial neural networks and decision trees) along with statistical method (logistic regression)	
Su et al. (2006)	One of the prominent approaches to predict diabetes while using four data mining methods in combination.	
Phillips-Wren et al. (2008)	Assessment of utilization of healthcare resources by lung cancer patients related to their demographic characteristics like ethnic background, medical histories which will guide in medical decision making and further drafting in public medical policies	
Z.Y. Zhuang (2009)	Intelligent DSS for pathology test ordering by GPs with the use of Case Based Reasoning (CBR)	
Lopez-Vallverdu et al. (2012)	Wearable Sensing systems	Used only for personal health care, Expensive technique
Gaura et al. (2013)	Used Experimental sensor data to perform monitoring non clinical health data	Research performed on non medical data.
Huang et al. (2013)	Online data mining of abnormal period patterns from medical sensor data streams	Applicable only for online tools and difficult to interpret
Clifton et al. (2013)	Proposed personalized e-health monitoring using wearable sensors	Deviations in results of different samples

F. E. Dewey et al. (2014)	Proposed actionable recommendations for analyzing genome-scale data	Implications of current public health policies, delivery of care and its associated costs.
T. G. Kannampallil et al. (2014)	Presented rational analysis of information seeking behavior in critical care	
T. Hussain and Q. T. Nguyen (2014)	Proposed Molecular imaging technique for cancer diagnosis and surgery.	
K. Bernatowicz et al. (2015)	Presented the concept of 4D CT (Four dimensional tomography)	The technology is still at inception stage
C. M. C. Tempany et al. (2015)	Proposed Advanced Multimodal Image-Guided Operating (AMIGO) Suite incorporating angiographic X-ray system, MRI, 3D ultrasound, and PET/CT imaging	Used only for cancer therapy
W. Y. Hsu (2015)	Proposed Segmentation-based compression technique	Reducing the volume of data while maintaining important data such as anatomically relevant data
L. Qu, F. Long, and H. Peng (2015)	Presented 3D registration of biological images	Used only for segmentation and annotation of images
J. M. Blum, H. Joo, H. Lee, and M. Saeed (2015)	Implemented hospital wide waveform capture system	Quality and accuracy of data is a concern
M. Attin, G. Feld, H. Lemus et al. (2015)	Inhospital early detection system for cardiac arrest based on Electrocardiogrpah parameters from telemetry along with demographic information	

Knowledge management tools

Artificial Intelligence techniques can be considered as the first ever methods have been used in knowledge management in the field of biomedical techniques. It started in 1970s when *MYCIN* program was developed in order to support medical decision making (Shortliffe et al, 1976). *MYCIN* was an intelligent computer program in which knowledge obtained from experts was represented as set of programming constructs like IF-THEN rules. It worked as the motivation of another type of system called *expert systems* which became very popular knowledge management tool in between 1980 to 2000. The concept of expert system was to feed knowledge into the systems in form of sense of reasoning which the system will utilize while making decisions. Though *MYCIN* was a very effective and was an early success, but it could not be used in actual clinical setups. It had two major reasons; resistance of people because they were unfamiliar with computer technologies. Even many medical practitioners did not believe that computer can perform better than humans. Another reason was heavy costs incurred in establishing such systems because computers were bulky and extensively expensive machines in 1970s. Heathfield et al. advocated that patient record management systems are highly desirable knowledge management tool in clinical setups (Heathfield et al 1999). Dewas et al. also pointed out the major reasons behind the desire as information needs of medical practitioners and clinical information overload (Dawes et al 2003). Hersh et al

(1996) classified clinical information into two broader categories: clinical information specifically related to patients and knowledge based information. Both types of information are growing at a rapid pace as huge number of medical reports, academic research papers, books and technical reports are getting generated on daily basis.

Though early clinical systems were nothing greater than ordinary data storage systems, but they can be considered as first generation knowledge management systems. The use of ontologies in biomedical knowledge management has been widespread since over two decades. An ontology is basically a conceptualization of specification which defines the existing relationships and describes the terminologies in a particular domain. The major use of ontologies is to facilitate knowledge sharing among people, information networks, softwares and knowledge management applications. *HELP* System developed in 1991 supported in monitoring a traditional medical record management system. The SAPHIRE system was enabled to perform automatic indexing of reports related to radiology by utilizing the UMLS Metathesaurus (Hersh et al. 2002).

Other than reporting and formatting clinical information, there is many other knowledge management system tools available in form of research articles and reports and most of them are available thorough digital library techniques. The National Library of Medicine (NLM) provides the PubMed service which contains over 13 million citations for biomedical articles from different medical journals. Regarding application of different models and techniques, MARVIN is one the medical information retrieval systems, (Baujard et al., 1998) which uses machine learning techniques. It was based on multi-agent architecture which works on scrutinizing relevant documents from given sets of web pages using machine learning methods and follows web links in order to retrieve new documents. Shatkay et al. used probabilistic similarity based search in order to retrieve biomedical documents which are similar in terms of themes (Shatkay et al., 2002). Chau and Chen et al. applied artificial neural networks for filtering web pages (Chau and Chen et al., 2004). Three dimensional display (Han and Byun, 2004) for visualization of protein interaction networks. This is very much similar to Virtual Reality which has been successfully implemented in visualizing metabolic networks (Rojdestvesski, 2003).

DISCUSSION AND ANALYSIS:

Effective use of data mining in the medical field is a need of time due to the idiosyncratic nature of the medical profession. It demands widespread changes and transformations in processes which further urge to have throughout understanding of requirements of the healthcare sector. Use of techniques of data mining can lead to requisite knowledge which can further lead successful decisions that will improve the overall success ratio of medical experiments and treatments. It will not only benefit to the health of the patients but also will be an value addition in the service of mankind. This presented review of data mining applications and knowledge management concepts has endowed us with an overview of current practices and further challenges. Health care organizations and agencies could be benefitted with the effective use of these applications and to find useful ideas for knowledge discovery from their own database systems. In further sub-sections we have presented key advantages and disadvantages of Data Mining and Knowledge Management tools

and techniques while highlighting major limitations of the same.

Advantages of knowledge management and data mining

As the a result of evaluation of above given contents, the conclusion can be stated as knowledge management and data mining has vast scope and advantages which can enable both of techniques to provide better health delivery. The advantages are listed below as –

- Quality of care – Oranzo et al. (2008) has suggested that adoption of knowledge management techniques can lead to enhancement in the quality of care specifically related to health care domain. Chae et al. (2001) has proved that knowledge management tools are very effective in different domains including health insurance. Besides this, the efficiency of work can be enhanced by applying knowledge management tools in day to day work (Davenport et al. 2002). Similar kind of research conducted by Goddard et al. (2004) has proved that knowledge management tools are useful in public health care decision making.
- Cost Reduction – Lamout et al. (2007) has argued that information sharing among stake holders of the organization can lead to cost effective use of health resources. McElroy (2005) suggested that use of knowledge management tools has significantly reduced high cost of medical errors.
- Medical error reduction – Abidi et al. (2001) has studied the impacts of knowledge management practices used by medical practitioners has provided effective decision support in the process of medical error reduction. Case based reasoning can also be used for the same purpose (Montani et al. (2002). Additionally, knowledge management tools has been proved as effective tool for reducing medical prescription errors which leads to cost reduction as well Melymuka (2002).
- Scope of Innovation – In recent times, innovation has been proved as the major driving tool for new developments and advancements in organizations. Medical field is not an exception for same. In fact, innovation is the one of the essential part of knowledge management as it includes sharing of concepts, learning and experiences so that more people can become aware of it and provide further refinement and enhancements in the same tools. Buchan & Hanka (1997) advocated distributed knowledge management paradigms for management of clinical knowledge. Ansel et al. (2007) further identified innovation facilitation methodologies which are being followed by analysis of knowledge flow barriers which may hamper the effective communication in the organization.
- Mobility – The current era can be referred as mobile era and future also belongs to mobile and mobility. Hubert et al. (2006) highlighted the explosion of mobile, interactive devices, e-homecare solutions. In fact, mobile services has emerged in form of virtual communities in which knowledge is captured, shared and disseminated by individuals involved in communication networks which includes patients and medical practitioners.

- **Challenges of knowledge management and data mining**

- Based on extensive review of literature and recent practices, followings are the identified challenges of medical informatics -
- The growing volume of medical data is resulting into growing number of challenges and opportunities which includes data analysis, capturing and managing data and data mining tools. As

the diversity and volume of data getting increased,, so does researchers feeling need to develop innovative and effective ways of capturing, retrieving and using the knowledge from this data. Effective and efficient management of this knowledge is one of emerging challenges of concerned field. Gene expression is one of biomedical concept which extensively uses data mining techniques. Basically medical researchers measure the level of expression of all genes of a particular tissue under a given condition and conduct comparative analysis of frequency levels of same genes in same tissue under different conditions. This process is known as differential gene expression. The major point of concerns and challenge in gene expression is management of experimental data because a single gene expression measurement generates millions of data points.

- Another challenge is data analysis as number of tools available for the purpose does not implement latest algorithms and methods for the same. However few open source collaborations are available for researchers in order to conduct array analysis. Finally we can conclude that there is a need of new advancements to mine gene expression data in effective manner.
- Privacy is another significant challenge of biomedical field which warrants attention. Broadly defined, privacy refers to controlling release of information and eliminating any possibility of intrusion or disturbance. In the context of medical field, we share sensitive data and information with medical practitioners which makes it vulnerable to goes public on which we do not have any control. Medical data and records contains extensively sensitive and confidential information and therefore, it is a very serious concern that such data should be handled with proper caution and care so that their privacy and security can be completely assured and guaranteed. Researchers or data analysts who are using medical records and data do not have an automatic right to use such data and they are bound to have necessary approval and consent (Berman, 2002) before using the data. United States of America has enacted Health Insurance Portability and Accountability Act (HIPAA) which has imposed several rules and standards for handling patient data in electronic form. The EU Data Protection Directive imposed similar kind of regulations in Europe. Such legal provisions ensure the ethical and legal use of medical data for only social concerns and benefits of society.
- Security is one another challenge of the field which is another result of growing use of Internet. However, security concern is relatively new in comparison of other challenges. Numerous threats exist in and outside of the organizations which are part of security concerns such as malicious codes i.e. virus, worms, Trojan horse. Attacks of such malicious codes results into denial of services, theft of information and other kind of intrusions. The major reason of such attacks is often the vulnerabilities in operating systems. However there may be other reasons also responsible for same like lack of trained staff, failure in using updated antivirus software.
- Different types of data being used in medical science are another issue of concern as data is the core at all types of decisions and incorrect or incomplete data cannot lead to successful results. Few of the concerns related to qualitative aspects of data are. In biomedical area, it is very difficult to obtain satisfactory result of data mining acceptance because the data found in biomedicine field is very multifaceted and the quality of data is one of the major challenges on blowing the performance in this industry. The effect of data mining majorly depends on nature (quantity and quality) of existing data so that we can

conclude that data mining is a data driven approach. However, the nature of the biomedical data is still quite complex. Therefore, some concerns are indentified as the key findings of the chapter in order to improve the functions of data mining techniques as follow:

a. Volume of data - This is very difficult for all the data mining methods to achieve desirable results using raw data because biomedicine area is one in which has huge data is being generated on daily basis. It is required that the biomedical experts pre-process the data prior to apply data mining technique. In this context, pre-processing also has different aspects to be handled by the medical professionals and these aspects makes the data handling processes extensively rigorous and time consuming.
b. Nature of data - The dynamic nature of data makes it difficult to handle and manage as well. Adding new information, updating (new versions) of data is constant process which cannot be avoided if we wish to gain maximum benefit of data mining. Medical data has a common property of changing values i.e. new tests may have new type of results which makes medical data dynamic in nature. For data collection techniques, It is very difficult to makes noise free database (elimination of entire noise). To address this, the amount of noise in database to be collected in the future by the different data mining methods ,that will be approximately same as that in the current data. Therefore, data mining methods should be less sensitive to noise.
c. Missing attribute values -The methods of data mining generally facing the problems regarding missing attribute values, because the methods require a preset dimension for each data object in data base. In the context of medical database, the information assortment depends on collection of patient care activities not on organized research protocols. As an example, The University of Wisconsin Hospitals encountered the problem of missing values in some large medical databases such as breast cancer data sets. In order to address this issue, they skipped over the records having missing attribute values. We can have another approach to eliminate this difficulty is to substitute the missing values with some similar kind of. We can change over the missing values as a process by gathering values from existing values through Artificial intelligence methods (e.g. case-based reasoning).
d. Redundant, insignificant data, or inconsistent data - Data redundancy, inconsistency or insignificant data leads to data corruption. In such type of situation, data objects and attributes contain repeated non relevant values in data sets and it must be avoided by designing the data base in correct form. Pre-processing of data is the ideal way to get useful and desirable data. In Biomedical Informatics field the databases contains huge amount of data in textual and numeric format thus, data should be pre-processed in order to eliminate possibility of redundancy and inconsistency of data. As an example, in the medical terminology one condition or prescription may be generally referred to by many names like (i.e. stomach and abdominal pain), this condition commonly occurred by the misspelled of medical terms.

In perspectives of data quality in the database, several considerations are also

been made. Quality of learning mechanism is one of them which refer to the affects of mining process by the mean of several ways. One of them is under and over learning. This condition occurred when the human's preferences misunderstand by the learning mechanism techniques and for accomplish the target, they need human being. Quality of knowledge representation is another perspective. Representation of knowledge or information in biomedical area is a significant tool for patient data or clinical trials. The aim of Knowledge representation is to understand things in easy manner. If the machine is not capable to store the knowledge discovered and to represent the same; then, the machine cannot be referred as intelligent machine. Nature of problem is playing an important role in data mining methods. Sometimes the intelligent system or machine has insufficient time or knowledge to solve the problem and provide correct result

CONCLUSION:

In this paper, we provided a detailed and refined overview of various technologies and tools of Data Mining and Knowledge Management with emphasis on tools and techniques used especially in medical informatics. Nonetheless, the paper presents a timeline and flowing summary of all major tools and applications of data mining and knowledge management which are used in medical informatics. Since past two decades, it has been observed that data mining and knowledge management has served as major driving force for advancements in medical informatics field and we have witnessed numerous developments in the field of bioinformatics. However, it is also recommended that medical data and information should be handled with due care in order to ensure necessary security and privacy concerns. It must at priority that patient's records and details must be used for any illegal or unethical practice. Additionally confidentiality and privacy of such records must not be compromised at any level due to recent advancements in medical fields. Another point which needs attention is the interpretations and judgments which are based upon the findings discovered by computers using data mining and knowledge management tools. Such interpretations must be validated in similar ways just like any other knowledge generated by human. Incorrect interpretations should be propagated through media so that people understands the consequences and actual facts of the situations.

REFERENCES:

1. Abidi, S. S. R. (2001). "Knowledge Management in Healthcare: Towards 'Knowledgedriven' Decision-support Services," International Journal of Medical Informatics, 63, 5-18.
2. Antonie M. L., Zaiane O. R., and Coman A. (2001), "Application of data mining techniques for medical image classifica-tion," in Proceedings Second International Workshop on Multimedia Data Mining, pp. 94–101.
3. Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian (2015), "Big Data Analytics in Healthcare," BioMed Research International, vol. 2015, Article ID 370194, 16 pages, 2015. doi:10.1155/2015/370194
4. Brossette S E, Sprague A. P., Jones W. T., and Moser S. A., (2000), "A data mining system for infection control sur-veillance," Methods of Information in

Medicine, Vol. 39, No. 4-5, pp. 303–310
5. Carbonell, J. G. Michalski, R. S., Mitchell, T. M. (1983). "An Overview of Machine Learning," in R. S. Michalski, J. G.
6. Clifton, L.; Clifton, D.A.; Pimentel, M.A.F.; Watkinson, P.J.; Tarassenko, L. (2013) Gaussian processes for personalized e-health monitoring with wearable sensors. IEEE Trans. Biomed. Eng., 60, 193–197.
7. C. M. C. Tempany, J. Jayender, T. Kapur et al. (2015), "Multimodal imaging for improved diagnosis and treatment of cancers,"Cancer, vol. 121, no. 6, pp. 817–827
8. Coulter D. M., Bate A., Meyboom R. H. B., Lindquist M., and Edwards R. (2001), "Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: Data mining study," British Medical Journal, 322, 1207–1209.
9. Delen D., Walker G., and Kadam A. (2005), "Predicting breast cancer survivability: A comparison of three data mining methods," Artificial Intelligence in Medicine, 34(2), 113–27.
10. Duda, R. 0.and Hart, P. E. (1973). Pattern ClassiJication and Scene Analysis, New York: John Wiley and Sons.
11. F. E. Dewey, M. E. Grove, C. Pan et al. (2014), "Clinical interpretation and implications of whole-genome sequencing," JAMA, vol. 311, no. 10, pp. 1035–1045
12. Fisher, D. H. (1987)."Knowledge Acquisition via Incremental Conceptual Clustering," Machine Learning, 2, 139-172.
13. Gaura, E.; Kemp, J.; Brusey, J. (2013), Leveraging knowledge from physiological data: On-body heat stress risk prediction with sensor networks. IEEE Trans. Biomed.Circuits System.
14. Huang, G.; Zhang, Y.; Cao, J.; Steyn, M.; Taraporewalla, K. (2013), Online mining abnormal period patterns from multiple medical sensor data streams. World Wide Web, doi:10.1007/s11280-013-0203-y.
15. Hubert, R. (2006). Accessibility and usability guidelines for mobile devices in home health monitoring. *SIGACCESS Accessibility and Computing* (84), 26-29.
16. J. M. Blum, H. Joo, H. Lee, and M. Saeed (2015), "Design and implementation of a hospital wide waveform capture system," Journal of Clinical Monitoring and Computing, vol. 29, no. 3, pp. 359–362
17. K. Bernatowicz, P. Keall, P.Mishra, A. Knopf, A. Lomax, and J. Kipritidis (2015), "Quantifying the impact of respiratory-gated 4D CT acquisition on thoracic image quality: a digital phantom study," Medical Physics, vol. 42, no. 1, pp. 324–334
18. Kohonen, T. (1995).Self-organizing Maps, Springer-Verlag, Berlin.Kononenko, I. (1993). "Inductive and Bayesian Learning in Medical Diagnosis," Applied Artificial Intelligence, 7,3 17-337
19. Langley, P. and Simon, H. (1995). "Applications of Machine Learning and Rule Induction," Communications of the ACM, 38(1 I), 55-64.
20. Li L., Tang H., Wu Z., Gong J., Gruidl M., Zou J., Tockman M., and Clark R. (2004), "Data mining techniques for cancer detection using serum proteomic profiling," Artificial Intelligence in Medicine, 32(2), 71–83.
21. L. Qu, F. Long, and H. Peng (2015), "3D registration of biological images and models: registration of microscopic images and its uses in segmentation and annotation," IEEE Signal Processing Magazine, vol. 32, no. 1, pp. 70–77
22. M. Attin, G. Feld, H. Lemus et al. (2015), "Electrocardiogram characteristics prior to in-hospital cardiac arrest," Journal of Clinical Monitoring and Computing, vol. 29, no. 3, pp. 385–392
23. Megalooikonomou V., Ford J., Shen L., Makedon F., and Saykin A. (2000), "Data mining in brain imaging,"

Statistical Methods in Medical Research, Vol. 9, No. 4, pp. 359–394.

24. Prather, J. C., Lobach, D. F., Goodwin, L. K., Hales, J. W., Hage, M. L., and Hammond, W. E. (1997)."Medical Data Mining: Knowledge Discovery in a Clinical Data Warehouse,"cin Proceedings of the AMIA Annual Symposium Fall 1997, 101-105.
25. Philips-Wren G., Sharkey. P., and Morss. S. (2008), "Mining lung cancer patient data to assess healthcare resource utiliza-tion," Expert Systems with Applications: An International Journal, 35(4), 1611–1619.
26. P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis (2011), Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill Osborne Media
27. Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986a). "A General Framework for Parallel Distributed Processing," in D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), Parallel Distributed Processing, pp. 45-76, Cambridge, MA: The MIT Press.
28. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986b). "Learning Internal
29. Representations by Error Propagation," in D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), Parallel Distributed Processing, pp. 318-362, Cambridge, MA: The MIT Press.
30. Su, C. T., Yang, C. H., Hsu, K. H., and Chiu, W. K. (2006), "Data mining for the diagnosis of type II diabetes from three- dimensional body surface anthropometrical scanning data," Computers & Mathematics with Applications, 51(6–7), 1075–1092.
31. T. Hussain and Q. T. Nguyen (2014), "Molecular imaging for cancer diagnosis and surgery," Advanced Drug Delivery Reviews, vol. 66, pp. 90–100
32. T. G. Kannampallil, A. Franklin, T. Cohen, and T. G. Buchman (2014), "Sub-optimal patterns of information use: a rational analysisof information seeking behavior in critical care," in Cognitive Informatics in Health and Biomedicine, pp. 389–408, Springer, London, UK
33. W. Y. Hsu (2015), "Segmentation-based compression: new frontiers of telemedicine in telecommunication," Telematics and Informatics, vol. 32, no. 3, pp. 475–485

CONFLICT OF INTEREST: Authors declared no conflict of interest