Data Mining and Health Care: Techniques of Application
Ms Manaswini Pradhan
PG Dept of I & CT, F M University, Balasore, Odisha 756019 India
ms.manaswini.pradhan@gmail.com

ARTICLE INFO
Received 15 Nov. 2014
Accepted 20 Dec. 2014
Corresponding Author:
Ms Manaswini Pradhan
PG Dept of I & CT, F M University, Balasore, Odisha 756019 India
ms.manaswini.pradhan@gmail.com

ABSTRACT
The issue of health care assumes prime importance for the society and is a significant indicator of social development. Health is clearly not the mere absence of disease but confers on a person or group’s freedom from illness and the ability to realize one’s potential. Health is therefore best understood as the indispensable basis for defining a person’s sense of well-being. The delivery of health care services thus assumes greater proportion, and in this context the role played by information and communication technology has certainly a greater contribution for its effective delivery mechanism. The application of data mining is specifically relevant and it has been successfully applied in medical needs for its reliable precision accuracy and expeditious beneficial results. The various available application techniques have been discussed and analyzed for the purpose of the paper.

Keywords: Health care, Data mining, Electronic Medical Records (EMR), Medical decision support, Diagnosis and Treatment

INTRODUCTION
In India healthcare is delivered through both the public sector and private sector. The public healthcare system consists of healthcare facilities run by central and state government which provide services free of cost or at a subsidized rates to low income group in rural and urban areas. With the Indian economy enjoying a steady growth, the industry is heading towards growth phase by (M. Radha, 2011). According to (Jeffrey M. Lackner et al. 2013) the introduction of product patents in India is expected to boost the industry by encouraging multinational companies to launch specialized life-saving drugs. Attracted by the advantages such as lower costs of production and skilled workforce that India offers, these companies are looking to set up research and development as well as production centers there. Health and health care need to be distinguished from each other for no better reason than that the former is often incorrectly seen as a direct function of the latter. Heath is clearly not the mere absence of disease by (Ilias Iakovidis, 1998). Good Health confers on a person or group’s freedom from illness - and the ability to realize one’s potential. Health is therefore best understood as the indispensable basis for defining a person’s sense of well-being. The health of populations is a distinct key issue in public policy discourse in every mature society often determining the deployment of huge society. According to (Shivalingappa B Javali et al. 2012) include its cultural understanding of ill health and well-being, extent of socio-economic disparities, reach of health services and quality and costs of care. And current bio-medical understanding about health and illness.

1.1. Definitions of Health Care:
A major challenge facing healthcare organizations is the provision of quality services at affordable costs. Quality service implies correct diagnosis and administering effective treatments to patients. Poor clinical diagnosis can certainly lead to disastrous consequences which are unacceptable (Alok Bhargava, et al. 2005). Hospitals must also minimize the cost of clinical tests. According to Hoosen Coovadia et al. (2009), it can achieve these results by employing appropriate computer-based information and/or decision support systems. Health care data is massive and includes patient centric data, resource management data and transformed data (Cheon-Pyo Lee et al. 2007). Health care organizations must have the ability to computationally analyze data that is stored from treatment records of millions of patients. Thereby, data mining techniques may help in
answering several important and critical questions related to health care.

The healthcare industry has increased in size and content over the years. India will spend €33.8 billion on healthcare in the next five years as the country, on an economic upsurge, is witnessing changes in its demographic profile accompanied with lifestyle diseases and increasing medical expenses (Patricia Pittman et al. 2007). Revenues from the healthcare sector account for 5.2% of the GDP and it employs over 4 million people. Pittman et al. (2007) stated that by 2012, revenues can reach 6.5 to 7.2% of GDP and direct and indirect employment can double. According to Palaniappan S et al. (2008), private healthcare will continue to be the largest component in 2012 and is likely to double to €26.41 billion. It could rise by an additional €6.5 billion if health insurance cover is extended to the rich and middle class. Coupled with the expected increase in the pharmaceutical sector, the total healthcare market in the country could increase to €39.22 – 54 billion (6.2-8.5% of GDP) in the next five years.

1.2. Scope of the Healthcare Industry: Some of the macro factors for health industry’s growth are the following:

- **Private Sector**

Emerging private sector is more focused on tertiary-level, as well as, preventive and diagnostic healthcare and is sensing a huge untapped opportunity in delivery of quality healthcare to the Indian masses (Leigh Turner et al. 2007). According to Harleen Kaur et al. (2009), public sector is accelerating the prevention and elimination of infectious diseases by providing accessibility basic healthcare facilities to the rural masses. Global private equity and venture capitals are playing a vital and varied role in Indian healthcare delivery for increasing the global footprint of local pharmaceutical companies to aiding the rapidly growing contract research outsourcing industry (Patricia Pittman et al. 2007).

- **Market situation:**

The past decade has witnessed a great change in the availability of healthcare facilities in the country. The number of public hospitals grew from 4600 in the year 2000 to more than 7600 in the year 2006. That is an increase of more than 67% in just six years (Siri Krishan Wasan et al. 2006). The private hospitals also have been growing at a similar pace during these years. The private healthcare boasts of superior quality and facilities as it “accounts for more than 65% of primary care and more than 40% of hospitals, resulting in personnel shortage in the public sector” (K.Srinivas et al. 2010).

- **Medical infrastructure:**

It accounts for the largest portion in the healthcare sector. The availability ratio of bed per thousand population for India stands at 1.03 as against an average of 4.3 in countries like China, Korea and Thailand. Hence, in spite of the phenomenal growth in the healthcare infrastructure, India may very likely attain a bed to thousand-population ratio of 1.85, or at best scenario, a ratio of 2 (Latha Parthiban et al. 2008). Beds in excess of 1 million need to be added to reach a ratio of 1.85 per thousand, out of which about 896,500 beds will be added by the private sector with a total investment of €51 billion over the next six years. According to Varun Kumar et al. (2008) the gains are commensurate in this capital intensive industry, since the revenues generated by private hospitals in the year 2012 will be to the tune of €26.5 billion growing at a rate of 15%. Despite this investment, the bed to thousand population ratio would be far from comparison to other similar developing countries.

- **Telemedicine:**

It allows even the interiors to access quality healthcare and at the same time significantly improves the productivity of medical personnel. In a country of over 1.1 billion people, the healthcare system will have to innovate to double the utilization of its existing resources so as to reach a stage available in developing countries (Aqueel Ahmed et al. 2012). According to Monica Chiariini Tremblay et al. (2009) “telemedicine is one such innovative technology”, and “if used effectively can double utilization of scarce human resources” (Leigh Turner et al. 2007). If telemedicine models are integrated with the healthcare model, such models may become viable. However, standalone telemedicine models may not be feasible, One of the important reasons for the success of telemedicine is it can increase the patient base (Vikram Jeet Singh et al. 2013), which in turn will increase occupancy rates of hospitals in the integrated telemedicine model.

2. **Data Mining:**

Data mining can be considered as a relatively recent developed methodology and technology, coming into prominence (K. R. Lakshmi et al. 2013). It aims to identify valid, novel, potentially useful, and understandable correlations and patterns in data by combing through copious data sets to sniff out patterns that are too subtle or complex for humans to detect. According to Jinn-Yi Yeh et al. (2011), data mining techniques can be broadly classified based upon what they can do, viz: (a) description and visualization; (b) association and clustering; and (c) classification and estimation; and thus can be a predictive modeling.

**Objectives:** The objectives of the present research paper are the following:

1. To enumerate current uses and highlight the importance of data mining in medicine and public health;
2. To find data mining techniques used in other fields that may also be applied in the health sector;
3. To identify issues and challenges in data mining as applied to the medical practice; and
4. To outline some recommendations for discovering knowledge in electronic databases through data mining.

2.2. Data Mining Applications in Healthcare:
Data mining techniques have been used intensively and extensively by many organizations. In healthcare, data mining is gradually increasing popularity, if not by any case, becoming increasingly essential. Data mining applications can greatly benefit all parties involved in the healthcare industry (Petr Hájek et al., 2010). For example, data mining can help healthcare insurers detect fraud and abuse; healthcare organizations can make customer relationship management decisions; physicians can identify effective treatments and best practices; and patients receive better and more affordable healthcare services. Data mining can be defined as “the process of finding previously unknown patterns and trends in databases and using that information to build predictive models” (E.W.T. Ngai et al., 2011). Alternatively, it can be defined as the process of data selection and exploration and building models using vast data stores to uncover previously unknown patterns.

2.3. Why Data Mining Can Aid Healthcare:
In healthcare, data mining is becoming increasingly popular, if not, increasingly essential. Several factors have motivated the use of data mining applications in healthcare. According to Chao-Hui Lee et al. (2010) existence of medical insurance fraud and abuse, for example, has led “many healthcare insurers to attempt to reduce their losses by using data mining tools to help them find and track offenders” (Aqueel Ahmed et al., 2012). According to Monica Chiarini Tremblay et al. (2009), fraud detection using data mining applications is prevalent in the commercial sector, for example, “in the detection of fraudulent credit card transactions” (R.S. Santos et al., 2010). Recently, many cases of successful data mining applications in healthcare insurance fraud and abuse have been detected and reported. Hence, it can be pointed that data mining can lead to the following deliveries:
  • Healthcare insurer’s fraud and abuse detection;
  • Healthcare organization’s decisions for customer relationship management;
  • Physician’s can identify effective treatments and best practices; and
  • Patient’s receive better and more affordable healthcare services.

3. Review of literature for Application with data mining in Healthcare Industry
In this section, an attempt is made to comprehensively and systematically review the literature on healthcare industry and classify the applications with data mining. According to Monica Chiarini Tremblay et al. (2009) it can be easily comprehended and applied. For this, around hundred papers from three leading journals spread over last ten years have been considered.

3.1. Important role of data mining in Healthcare Application:
Despite the differences and clashes in approaches, the health sector has more need for data mining today. There are several arguments that could be advanced to support the use of data mining in the health sector, covering not just the concerns of public health but also the private health sector (Ting-Ting Lee et al., 2011). According to Wen-Yuan Jen et al. (2009), there is a wealth of knowledge to be gained from computerized health records. Yet the overwhelming bulk of data stored in these databases make it extremely difficult, if not impossible, for humans to sift through it and discover knowledge (Syed Sibte Raza Abidi et al., 2001). In fact, some experts believe that medical breakthroughs have slowed down, attributing this to the prohibitive scale and complexity of present-day medical information. Computers and data mining are
best-suited for this purpose. According to Hsu-Hao Tsai et al. (2012), there is evidence-based medicine and prevention of hospital errors. When medical institutions apply data mining on their existing data, they can discover new, useful and potentially life-saving knowledge that otherwise would have remained inert in their databases. For instance, an on-going study on hospitals and safety found that about 87% of hospital deaths in the United States could have been prevented, had hospital staff been more careful in avoiding errors. Hyunjung Shin et al. (2012) state that mining hospital records, such safety issues could be flagged and addressed by hospital management and government regulators.

Table 1: Medical and Non-Medical hospital mining records

<table>
<thead>
<tr>
<th>Medical</th>
<th>1. Picture Archival &amp; Communication System (PACS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. DICOM (Digital Imaging &amp; Communications in Medicine)</td>
</tr>
<tr>
<td></td>
<td>3. Telemedicine</td>
</tr>
<tr>
<td></td>
<td>4. Electronic Medical Records</td>
</tr>
<tr>
<td></td>
<td>5. Clinical Decision Support Systems</td>
</tr>
<tr>
<td></td>
<td>6. Clinical Decision Support Systems</td>
</tr>
<tr>
<td></td>
<td>7. HL7 protocol</td>
</tr>
<tr>
<td></td>
<td>8. Epidemic prevention software</td>
</tr>
<tr>
<td>Non-Medical</td>
<td>1. Integrated hospital information system</td>
</tr>
<tr>
<td></td>
<td>2. Web-enabled Appointment Scheduling</td>
</tr>
<tr>
<td></td>
<td>3. Web-enabled applications for relatives to obtain conditions of the critical patients</td>
</tr>
<tr>
<td></td>
<td>4. Multi-media applications for patient education</td>
</tr>
</tbody>
</table>

4. Gradual evolution of the healthcare infrastructure versus endogenous facility placement:
The policy-making in public health, combined GIS and data mining using among others, Weka with J48, analyzes similarities between community health centers in Slovenia (Ali Serhan Koyuncugil et al. 2012). Using data mining, Ali Serhan Koyuncugil et al. (2012) were able to discover patterns among health centers that led to policy recommendations for their Institute of Public Health.

Figure 2: Integrated medical database. Source: Ali Serhan Koyuncugil et al. (2012)

The Integrated Emergency, Healthcare, and Medical Information System will be developed in the web-based multimedia environment, mobility and real-time technology. The system provides an integrated medical database, which can provide stakeholders with related medical information. The registered users can log into the system to access or provide medical information based on their accessing privilege (Sumana Sharma et al. 2009). The system will have the capabilities for finding the patient location based and suggest the nearest emergency center, arrange all necessary related patient information to be ready for the
physician when the patient arrives, assigning a doctor to the patient based on the availability of the doctors and list all necessary requirements such as special devices or surgery room (Dursun Delen et al. 2009). The system is an open cross-platform web-based real-time client-server environment with multiple language capabilities. The system provides mechanisms for exchange of image files, shared discussion lists, textual information exchange, access to images and data exported from local data bases, voice and video transmission (Gloria Phillips-Wren et al. 2008). This is represented in the following figure.

Figure 3: Topology of Data Mining Technology in Health Care Application

The various data mining technology used by different authors are detailed below:

<table>
<thead>
<tr>
<th>References</th>
<th>Data Mining Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Aqueel Ahmed et al. 2012) (R.S. Santos et al. 2001)</td>
<td>System supports a function defined by standardized MOH forms can be entered for inclusion in the EMR.</td>
</tr>
<tr>
<td>(Syed Sibte Raza Abidi et al. 2001) (Hyunjung Shin et al. 2012)</td>
<td>Laboratory Information Management Systems</td>
</tr>
</tbody>
</table>

### 4.1 Functional Requirements for EMR Systems:
Electronic medical records (EMR) can be developed to address different goals and health settings, and consequently emerge with different functions and capabilities (Wan-Shiou Yang et al. 2006). However, it is desirable to maintain a core set of functions in each EMR system in order to support similar workflows and encourage best practices in clinical care.

### 4.2 Six key functional areas of EMR:
According to Zhengxing Huang et al. (2012), the functional requirements defined in this chapter can be categorized
into six key functional areas that are critical to the definition of an EMR:

(i) Basic demographic and clinical health information;
(ii) Clinical decision support;
(iii) Order entry and prescribing;
(iv) Health information and reporting;
(v) Security and confidentiality, and;
(vi) Exchange of electronic information.

### Table 3: Technology elements and its function

<table>
<thead>
<tr>
<th>Technology elements</th>
<th>Year</th>
<th>Description of current technology</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMR User interface text template</td>
<td>2010</td>
<td>User can enter pre-written text strings which can be edited</td>
<td>Text based interface prior to introduction of GUI software</td>
</tr>
<tr>
<td>EMR User interface -point-and-click</td>
<td>2008</td>
<td>Several interfaces in use no dominant player yet</td>
<td>Faster and more accurate billing process</td>
</tr>
</tbody>
</table>

4.3. Medical decision support
While arriving at a conclusive medical decision, data mining support assumes a high significance. Analysis of digitized images of skin lesions to diagnose melanoma, computer-assisted texture analysis of ultrasound images aids monitoring of tumor response to chemotherapy, are some of its applications. Further, **Hyunjung Shin et al. (2012)** predicted the presence of brain neoplasm with magnetic resonance spectroscopy. Digital images of tissue sections are analysed to identify and quantify senile plaques for diagnosing and evaluating the severity of Alzheimer’s disease.

4.4. Diagnosis and Treatment:
Data mining could be particularly useful in medicine when there is positive evidence favoring a particular treatment option. According to **Krzysztof J. Cios et al. (2002)**, based on patients’ profile, history, physical examination, diagnosis and utilizing previous treatment patterns, new treatment plans can be effectively suggested.

4.5. Healthcare Resource Management:
**David F. Motta Cabrera et al. (2013)** use logistic regression models to compare hospital profiles based on risk-adjusted death with 30 days of non-cardiac surgery. According to **A´lvaro Rebuge et al. (2012)**, neural network system is used to predict the disposition in children presenting to the emergency room with bronchiolitis. Further, data mining is relied while predicting the risk of in-hospital mortality in cancer patients with nonterminal disease.

4.6. Prediction of inpatient length of stay
A key problem in the healthcare area is the measurement of flow of patients through hospitals and other health care facilities (**Thanh Kim Dao et al. 2010**). If the inpatient length of stay (LOS) can be predicted efficiently, the planning and management of hospital resources can be greatly enhanced. Hence, it can effectively manage the resource allocation by identifying high risk areas and predicting the need and usage of various resources.

4.7. Customer Relation Management (CRM):
In determining the customer relationship management (CRM), the focus shifts away from the breadth of customer base (product oriented view, mass marketing) to the depth of each customer’s needs (customer-oriented view, one-to-one marketing) (**Dursun Delen et al. 2009**). CRM is built on an integrated view of the customer across the whole organization. Customers have a fractured view of an enterprise; the enterprise has a splintered view of the customer. **Kohli et al. (2010)** demonstrate a web-based Physician Profiling System (PPS) to strengthen relationships with physicians and improve hospital profitability and quality. Development of total customer relationship in healthcare includes several tenets. According to **S.Vijiyarani et al. (2013)** a helping profession, the ultimate judge of performance is the person helped. Most people, including sick people, are reasonable for most of the time. Different people have different, legitimate needs. Pain and fear produce anxiety in both the victim and the helper (**Dursun Delen et al. 2009**). Meeting needs without waste is a strategic and moral imperative. Some demographic characteristics and institutional characteristics consistently have a significant effect on a patient’s satisfaction scores. Chronic illnesses require self-management and a collaborative patient-physician relationship. The principles of applying of data mining for customer relationship management in the other industries are also applicable to the healthcare industry (**Dursun Delen et al. 2009**). The identification of usage and purchase patterns and the eventual satisfaction can be used to improve overall customer satisfaction (**N. Aditya Sundar et al., 2012**). The customers could be patients, pharmacists, physicians or clinics. In many cases prediction of purchasing and usage behavior can
help to provide proactive initiatives to reduce the overall cost and increase customer satisfaction.

4.8. Unhealthy insurance practices: Bolton and Hand (2002) briefly discuss healthcare insurance fraud by mentioning two unhealthy insurance practices. Prescription fraud: claims for patients who do not exist Upcoding: claims for a medical procedure which is more expensive or not performed at all.

The ability to detect anomalous behavior based on purchase, usage and other transactional behavior information has made data mining a key tool in variety of organizations to detect fraudulent claims, in appropriate prescriptions and other abnormal behavioral patterns (Motilal C. Tayade et al. 2003). Another key area where data mining based fraud detection is useful is detection and prediction of faults in medical devices.

5. Data quality and completeness:
Data quality and completeness are critical to the success of any information system. Achieving high standards is a particular challenge in sites with limited computer literacy and experience (Aqueel Ahmed et al. 2012). It is important to design systems that are easy to use and have good instructions and training. The system should collect the minimum data necessary for the task, and data items should be structured and coded where possible to simplify data checking and optimize reuse. This does not mean that free text must be excluded; doing so prevents the system from capturing any data that do not fit the normal pattern (Ting-Ting Lee et al. 2011). Such data will either be lost or recorded in hard-to-locate paper records. Structured data such as laboratory test results might benefit from double entry (Dursun Delen et al. 2009). In some instances physicians and other staff enter data directly. This has the advantage of avoiding transcription errors, and also allows order entry systems to be deployed to check for potential medical errors.

5.1 Obstacles for Data Mining in Healthcare
One of the biggest problems of data mining in medicine is that the raw medical data is voluminous, and heterogeneous. These data can be gathered from various sources such as from conversations with patients, laboratory results, review and interpretation of doctors (Ali Serhan Koyuncugil et al. 2012). All these components can have a major impact on diagnosis, and prognosis of the patient, and should not be ignored. The content and complexity of medical data is one of the obstacles for successful data mining. Missing, incorrect, inconsistent or non-standard data, such as pieces of information saved in different formats from different data sources, create a major obstacle to successful data mining (Dursun Delen et al. 2009). It is very difficult for people to process gigabytes of records, although working with images is relatively easy, because doctors are being able to recognize patterns, to accept the basic trends in the data, and formulate rational decisions. Stored information becomes less useful if they are not available in easily apprehensible format (Gloria Phillips-Wren et al. 2008). The role of visualization techniques is increasing in this, as the picture are easiest for people to understand, and can provide plenty of information in a snapshot of the results.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Consequences</th>
<th>Application Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E.W.T. Ngai et al.) (Petr Hájek et al.)</td>
<td>2010 2011</td>
<td>Dirt, moisture, and bird droppings can contaminate window air conditioners, which can then introduce infectious material into hospital rooms.</td>
<td>EMR User interface text template</td>
</tr>
<tr>
<td>(Ting-Ting Lee et al.) (Syed Sibte Raza Abidi et al.)</td>
<td>2011 2001</td>
<td>Several interfaces in use – no dominant player yet</td>
<td>EMR User interface - point-and-click</td>
</tr>
<tr>
<td>(Chao-Hui Lee et al.) (Hyunjung Shin et al.)</td>
<td>2010 2012</td>
<td>Socioeconomic, public spending and healthcare system</td>
<td>Laboratory Information Management Systems</td>
</tr>
<tr>
<td>(Aqueel Ahmed et al.) (Hsu-Hao Tsai et al.)</td>
<td>2012 2012</td>
<td>Particles in the air are larger than the openings between the filter fibers, resulting in gross removal of large particles.</td>
<td>Evaluation and Management Services</td>
</tr>
<tr>
<td>(R.S. Santos et al.) (Hsu-Hao Tsai et al.)</td>
<td>2010 2012</td>
<td>Particles collide with filter fibers and remain attached to the filter. Fibers may be coated with adhesive.</td>
<td>Document Management System</td>
</tr>
<tr>
<td>(Wen-Yuan Jen et al.) (Syed Sibte Raza Abidi et al.)</td>
<td>2009 2001</td>
<td>Small particles, moving in erratic motion, collide with filter fibers and</td>
<td>Continuity of Care Record</td>
</tr>
</tbody>
</table>
5.2 Data security and confidentiality:
Views of medical data security and confidentiality vary in different developing countries. In some sites, the use of electronic databases is treated with great suspicion (Chao-Hui Lee et al. 2010); in other sites staff thinks nothing of emailing sensitive medical data. Patients can face serious risk if their communities discover their HIV status or other sensitive medical information. For this data security and confidentiality gains significance for data mining technology.

5.3 Estimation of current HIT adoption and related factors:
For the present research study primary data source was the Healthcare Information and Management Systems Society -Dorenfest survey which represents a broad canvassing of acute care hospitals, chronic care facilities, and ambulatory practices on their adoption and plans to adopt various HIT components (Chao-Hui Lee et al. 2010). They included in the adoption category the provider organizations that had contracted for but not yet installed an EMR system (Sumana Sharma et al. 2009). To examine the factors related to differences in adoption, emerged additional data about the providers and then performed pro bit regression analysis (Petr Hájek et al. 2010). Our lower-bound estimate of HIT adoption as summed an integrated system that had an EMR, clinical decision support, and a central data repository from the same vendor to ensure interoperability They adjusted the estimates according to the known under-representation of smaller providers in this survey (E.W.T. Ngai et al. 2011).

5.4 Challenges in Data Mining for Healthcare: Data mining in health care faces challenges from various arenas. It becomes a complex process for collection, retrieval and analysis. It has to be systematically entered and stored for future application utilizations. A standardization programme has to be evolved in this direction that is to be uniformly adhered. Some of the challenges are discussed below.

5.4.1 Data sets from various data sources
1: Patient referral data can vary extensively between cases because structure of patient referrals is up to general practitioner who refers the patient
2: Dipti Patil et al. (2012) use neural networks to predict preterm birth on a heterogeneous maternal population

3: Traditional clinical-based prognosis models were discovered to contain some restrictions to address the heterogeneity of breast cancer.

5.4.2 Data from heterogeneous sources present challenges:
1. Sampling bias: Clinical studies use diverse collecting methods, inclusion criteria, and sampling methods.
2. Referral bias: Data represent a preselected group with a high prevalence of disease.
3. Selection bias: Clinical data sets include patients with different demographics.
4. Method bias: Predictors have varied specifications, granularities, and precisions.
5. Clinical spectrum bias: Patient records represent varied severity of a disease and co-occurrence of other medical problems.

5.4.3 Missing values, noise, and outliers
1. Cleaning data from noise and outliers and handling missing values, and then finding the right subset of data, prepares them for successful data mining.
2. Transcription and manipulation of patient records often result in a high volume of noise and a high portion of missing values.
3. Missing attribute values can impact the assessment of whether a particular combination of attribute-value pairs is significant within a dataset.

5.5 Advantages of Data Mining Application in Healthcare:
Information technologies in healthcare have enabled the creation of electronic patient records obtained from monitoring of the patient visits (Dursun Delen et al. 2009). This information includes patient demographics, records on the treatment progress, details of examination, prescribed drugs, previous medical history, lab results, etc. Information system simplifies and automates the workflow of health care institution (Sumana Sharma et al. 2009). Privacy of documentation and ethical use of information about patients is a major obstacle for data mining in medicine. According to Motiial C. Tayade et al. (2013) data mining to be more exact, it is necessary to make a considerable amount of documentation. Health records are private information, yet the use of these private documents may help in treating deadly diseases (V. Krishnaiah et al. 2013). Before data mining process can begin, healthcare organizations must formulate a clear policy concerning privacy and security of patient
records (Gloria Phillips-Wren et al. 2008). This policy must be fully implemented in order to ensure patient privacy. Health institutions are able to use data mining applications for a variety of areas, such as doctors who use patterns by measuring clinical indicators, quality indicators, customer satisfaction and economic indicators, performance of physicians from multiple perspectives to optimize use of resources, cost efficiency and decision making based on evidence, identifying high-risk patients and intervene proactively, optimize health care, etc (Ishtake S. H et al. 2012).

6. CONCLUSION:
Data mining has great importance for area of medicine, and it represents comprehensive process that demands thorough understanding of needs of the healthcare organizations. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve success of healthcare organization and health of the patients. Data mining requires appropriate technology and analytical techniques, as well as systems for reporting and tracking which can enable measuring of results. Data mining, once started, represents continuous cycle of knowledge discovery. For organizations, it presents one of the key things that help create a good business strategy. Today, there have been many efforts with the goal of successful application of data mining in the healthcare institutions. Primary potential of this technique lies in the possibility for research of hidden patterns in data sets in healthcare domain. These patterns can be used for clinical diagnosis. However, available raw medical data are widely distributed, different and voluminous by nature. These data must be collected and stored in data warehouses in organized forms, and they can be integrated in order to form hospital information system. Data mining technology provides customer oriented approach towards new and hidden patterns in data, from which the knowledge is being generated, the knowledge that can help in providing of medical and other services to the patients. Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments. With the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in the medical data.

Acknowledgments
This research work is supported by the grant offered as financial assistance for research project under “FAST TRACK SCHEME FOR YOUNG SCIENTISTS— (Computer Science Engineering)” for the year 2013-14 F. No. SB/FTP/ETA-246/2012 Dated 21-06-2013 of SERB, Department of Science and Technology, Govt. of India, Technology Bhavan, New Mehrauli Road, New Delhi-110016

References:
10. Leigh Turner, “First World Health Care at Third World Prices’: Globalization, Bioethics and Medical Tourism”,


35. Zhengxing Huang, Xudong Lu and Huilong Duan “On mining clinical pathway patterns from medical


