ISSN 2229-5518

Applications of data mining in health and pharmaceutical industry

Sandhya Joshi, Hanumanthachar Joshi

Abstract— The Data mining is gradually becoming an integral and essential part of health and pharmaceutical industry. Data mining techniques are being regularly used to assess efficacy of treatment, management of ailments, and also in various stages of drug discovery and process research. Off late, data mining has become a boon in health insurance sector to minimize frauds and abuse. A good number of health care companies, medical hospitals and pharmaceutical manufacturing units are employing the data mining tools due their excellent efficiency. The identification and quantification of pharmaceutical information can be extremely useful for patients, physicians, pharmacists, health organizations, insurance companies, regulatory agencies, investors, lawyers, pharmaceutical manufacturers, drug testing companies. This review deals with the major applications of data mining techniques in health care and pharmaceutical industries with relevant examples.

Index Terms—. Data mining, health care, pharmaceutical industry, classification, insurance, drug, Alzheimer's

1 Introduction

Data mining is a fast evolving technology, is being adopted in biomedical sciences and research. Modern medicine generates a great deal of information stored in the medical database. For example, medical data may contain MRIs, signals like ECG, clinical information like blood sugar, blood pressure, cholesterol levels, etc., as well as the physician's interpretation. Extracting useful knowledge and providing scientific decision-making for the diagnosis and treatment of disease from the database increasingly becomes necessary [1]. Data mining in medicine can deal with this problem. It can also improve the management level of hospital information and promote the development of telemedicine and community medicine [2]. The goal of data mining in clinical medicine is to derive models that can use patient specific information to predict the outcome of interest and to thereby support clinical decision-making.

In healthcare, data mining is becoming increasingly popular. Several factors have motivated the use of data mining applications in healthcare. The 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 [3]. Fraud detection using data mining applications is prevalent in the commercial world, for example, in the detection of fraudulent credit card transactions.

- *Associate Professor, Pooja Bhagavat Memorial Mahajana post graduate center, Mysore, India.PH- 09480126436 E-mail: sanjoshi17@yahoo.com*
- *Professor, Sarada Vilas College of Pharmacy, K.M. Puram, Mysore. PH- 09663373701. E-mail: amanjoshi17@mail.com*

Recently, there have been reports of successful data mining applications in healthcare fraud and abuse detection. Another factor is that the huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data [4]. Such analysis has become increasingly essential as financial pressures have heightened the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue, and operating efficiency while maintaining a high level of care [5].

Healthcare organizations that perform data mining are better positioned to meet their long-term needs [6]. Data can be a great asset to healthcare organizations, but they have to be first transformed into information. Yet another factor motivating the use of data mining applications in healthcare is the realization that data mining can generate information that is very useful to all parties involved in the healthcare industry. For example, data mining applications can help healthcare insurers detect fraud and abuse, and healthcare providers can gain assistance in making decisions, for example, in customer relationship management. Data mining applications also can benefit healthcare providers, such as hospitals, clinics and physicians, and patients, for example, by identifying effective treatments and best practices [7,8] The Centers for Medicare and Medicaid Services has used data mining to develop a prospective payment system for inpatient rehabilitation [9] The healthcare industry can benefit greatly from data mining applications. This review is intended to explore relevant data mining applications, understand methodologies involved and their potentials in healthcare and pharmaceutical industry.

ISSN 2229-5518

The term data mining, in its most common use, is very new. The term previously had been used pejoratively by some statisticians and other specialists to refer to the process of analyzing the same data repeatedly until an acceptable result arose [10]. By the early 1990s, a number of forces converged to make data mining a very hot topic. It has subsequently been widely applied in retailing, banking and financial services, insurance, marketing and sales and telecommunications.

At first, parts of the scientific community were slow to embrace data mining. This was at least partially due to the marketing hype and wild claims made by some software salespeople and consultants. However, data mining is now moving into the mainstream of science and engineering. Data mining has come of age because of the confluence of three factors. The first is the ability to inexpensively capture, store and process tremendous amounts of data. The second is advances in database technology that allow the stored data to be organised and stored in ways that facilitate speedy answers to complex queries. Finally, there are developments and improvements in analysis methods that allow them to be effectively applied to these very large and complex databases. It is important to remember that data mining is a tool, not a magic wand. You can't simply throw your data at a data mining tool and expect it to produce reliable or even valid results. You still need to know your business, to understand your data, and to understand the analytical methods you use. Furthermore, the patterns uncovered by data mining must be verified in the real world. Just because data mining predicts that a gene will express a particular protein, or that a drug is best sold to a certain group of physicians, it doesn't mean this prediction is valid in the real world. You still need to verify the prediction with experiments to confirm the existence of a causal relationship. Healthcare Data Mining Applications include evaluation of treatment effectiveness; management of healthcare; customer relationship management; and detection of fraud and abuse.

Effectiveness of Treatment: Data mining applications can be developed to evaluate the effectiveness of medical treatments. By comparing the causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective [11]. For example, the outcomes of patient groups treated with different drug regimens for the same disease or condition can be compared to determine which treatments work best and are most cost-effective [12]. Along this line, United HealthCare has mined its treatment record data to explore ways to cut costs and deliver better medicine [13]. It also has developed clinical profiles to give physicians information about their practice patterns and to compare these with those of other physicians and peer-reviewed industry standards.

Similarly, data mining can help identify successful standardized treatments for specific diseases. In 1999, Florida Hospital launched the clinical best practices initiative with the goal of developing a standard path of care across all campuses, clinicians, and patient admissions.8 Florida Hospital used data mining in its daily activities [7,14]. Other data mining applications are associating the various side-effects of treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub-populations that respond differently from the mainstream population to certain drugs, and determining proactive steps that can reduce the risk of affliction [12].

Group Health Cooperative stratifies its patient populations by demographic characteristics and medical conditions to determine which groups use the most resources, enabling it to develop programs to help educate these populations and prevent or manage their conditions [11]. Group Health Cooperative has been involved in several data mining efforts to give better healthcare at lower costs. In the Seton Medical Center, data mining is used to decrease patient length-of-stay, avoid clinical complications, develop best practices, improve patient outcomes, and provide information to physicians—all to maintain and improve the quality of healthcare [15].

For example, Blue Cross has been implementing data mining initiatives to improve outcomes and reduce expenditures through better disease management. For instance, it uses emergency department and hospitalization claims data, pharmaceutical records, and physician interviews to identify unknown asthmatics and develop appropriate interventions [11]. Data mining also can be used to identify and understand high-cost patients.5 At a Data mining can also facilitate comparisons across healthcare groups of things such as practice patterns, resource utilization, length of stay, and costs of different hospitals [16]. Recently, Sierra Health Services has used data mining extensively to identify areas for quality improvements, including treatment guidelines, disease management groups, and cost management [17].

Customer Potential Management Corp. has developed a Consumer Healthcare Utilization Index that provides an indication of an individual's propensity to use specific healthcare services, defined by 25 major diagnostic categories, selected diagnostic related groups or specific medical service areas by employing data mining techniques [18]. This index, based on millions of healthcare transactions of several million patients, can identify patients who can benefit most from specific healthcare services, encourage patients who most need specific care to access it, and continually refine the channels and messages used to reach appropriate audiences for improved health and long-term patient relationships and loyalty. The index has been used by OSF Saint Joseph Medical Centre to get the right messages and services to the most appropriate patients at strategic times. The end result is more effective and efficient communications as well as increased revenue [18]. According to Miller, [12] the data mining of patient survey data can help set reasonable expectations about waiting times, reveal possible ways to improve service, and provide knowledge about what patients want from their healthcare providers. CRM in healthcare can help promote disease education, prevention, and wellness services [19]. Florida Hospital has used data mining to segment Medicare patients as well as develop commercial applications that enable credit scoring, debt collection, and analysis of financial data [8,16]. Sinai Health System had used of data mining for healthcare marketing and CRM.

Fraud and abuse: Data mining applications that are used to detect fraud and abuse often establish norms and then identify unusual or abnormal patterns of claims by physicians, laboratories, clinics, or others. Among other things, these applications can highlight inappropriate prescriptions or referrals and fraudulent insurance and medical claims. For example, the Utah Bureau of Medicaid Fraud has mined the mass of data generated by millions of prescriptions, operations and treatment courses to identify unusual patterns and uncover fraud [12]. As a result of fraud and abuse detection, ReliaStar Financial Corp. has reported a 20 percent increase in annual savings, Wisconsin Physician's Service Insurance Corporation has noted significant savings [13] and the Australian Health Insurance Commission has estimated tens of millions of dollars of annual savings.

EXAMPLE OF DATA MINING IN MEDICINE

Alzheimer's disease is a syndrome of gradual onset and continuing decline of higher cognitive functioning. It is a common disorder in older persons and becomes more prevalent in each decade of life. Approximately 10% of adults 65 years and older, and 50% of adults older than 90 years, have dementia. The sample consisted of initial visits of 496 subjects seen either as control or as patients (Table 1). The Neurologists and Neurophysiologists apply the DSM-IV criteria along with some laboratory investigations to classify the subject's mental status as either normal, cognitively impaired i.e. dementia of Alzheimer's type. The data set consisted of 35 attributes broadly classified under five main observations namely age, neuropsychiatry assessments (table 2), mental status examination and laboratory investigations. Entire description of patient's data set is shown in (Table 1). The objective of the present work was on classification of various stages of Alzheimer's disease, exploration of various possible treatments and the management for each stage of AD [2].

Observation	Description	Values
Age	Age in years	Continuous
Neuropsychiatric Assessment	Cognitive changes, Psychiatric symptoms, Personality change Problem behaviours, Family history	Yes / No
Mental Status Examination	MMSE, BIMC, BOMC STMS and AMT	Continuous
Laboratory Investigations	Blood tests, Complete blood count, Serum Chemistry, Thyroid testing, Apolipoprotein E testing, Heavy metal screen for mercury Human Immunodeficiency Virus	Continuous
Stage of disease	Alzheimer's disease	Four stages : Normal/Mild Moderate, Severe

Table 1: Description of Patient Data base

NEUROPSYCHOLOGICAL EXAMINATIONS

The primary symptoms of dementia of Alzheimer's type are impairments in cognition, behavior, and function, a thorough mental status examination is a necessary part of the evaluationA number of instruments have been developed for this purpose. In the present study five important instruments such as Mini-Mental State Examination, Blessed Information Memory Concentration, Blessed Orientation Memory Concentration, Short Test of Mental Status and Functional Activity Questionnaire were employed. These instruments measure the performance in similar areas of cognitive function and take 5 to 10 minutes to administer and score. Each is reliable for ruling out dementia when results are negative.

LABORATORY INVESTIGATIONS

Various laboratory investigations such as complete blood cell count, thyroid-stimulating hormone level, serum electrolytes, serum calcium, and serum glucose to exclude potential infections or metabolic causes for cognitive impairment. Other testing, such as serology for syphilis, human immunodeficiency virus (HIV), urine analysis, culture and sensitivity, heavy metal assays, liver function, serum folic acid level, were considered in the present study and these tests should be performed only when clinical suspicion warrants.

ARCHITECTURE AND MODELING

The Architecture and modelling of the present work is depicted in Figure 2. The work begun with the collection of patient's records, which was followed by pre-processing of data

ISSN 2229-5518

set, includes converting the data in to numeric values. Next stage was evaluation of the attributes by using gain ratio attribute evaluation scheme with ranker search method. The main focus of the work was the classification of various stages of AD using various Machine learning methods. The last and the most important stage are treatment and management of AD, in this stage we suggest different approaches for the management for each of the classified stages of AD were suggested.

PRE-PROCESSING PATIENT'S RECORDS

The collected data was free from outliers, noise and missing data. Some inconsistencies were recorded in the data set and those consistencies were corrected manually by using external references.

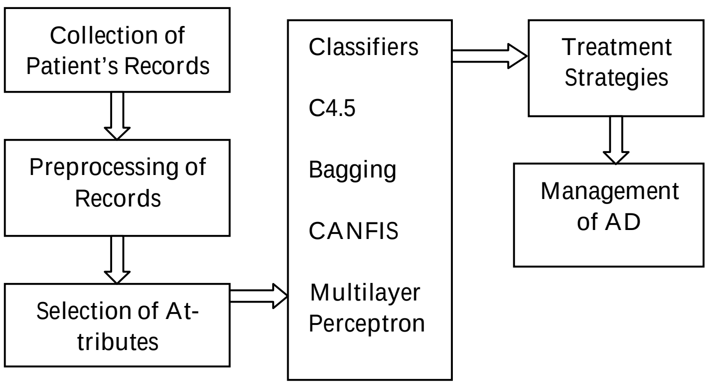

Figure 1: Architecture of Classification of Different Stages of Alzheimer's Disease

The data set in the present study consisted of 496 training cases and 225 testing cases, with 35 attributes broadly classified under five main observations namely age, neuropsychiatry assessments, mental status examination, laboratory investigations and physical examinations (Table 8.5).

Class	Normal	Mild	Moderate	Severe
Training data set	98	140	138	120
Testing data set	62	68	50	45

Table 8.5: Description of Data Classification

ML Methods	Classification Accuracy	Runtime (msec)	Sensitivity (%)	Specificity (%)
C4.5	98.97	0.023	98.61	97.94
Bagging	98.44	0.025	97.92	98.68
Neural Networks	98.25	0.018	98.62	98.89
Multilayer, Perceptron	98.99	0.015	99.01	98.94
CANFIS	99.55	0.009	99.52	99.93

Table 8.6: Results of Classification Accuracies using Various Machine Learning Methods

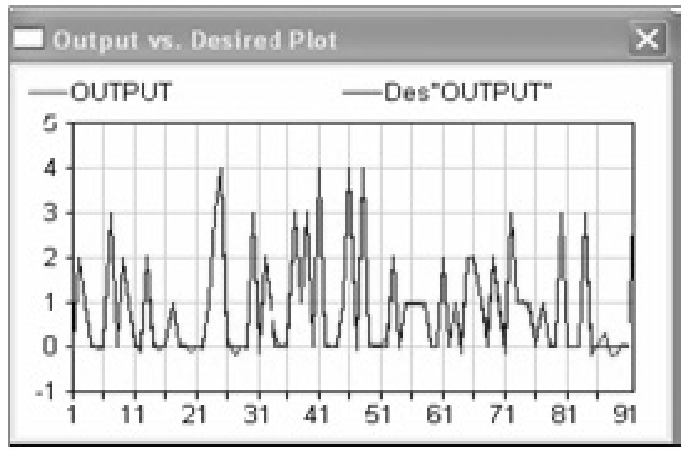

Figure 8.5: Output Vs Desired Plot for the Alzheimer's Database

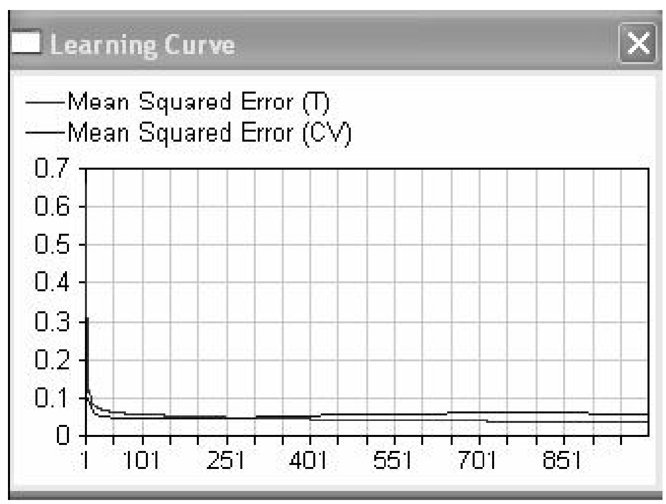

Figure 8.6: Learning Curve for the Alzheimer's Database

Class	Normal	Mild	Moderate	Severe
True	62	67	50	45
False	0	0	1	0

Table 8.7: Classification Matrix for Test set (225 test samples)

Results of classification accuracies using various ML methods are shown in Table 5.6.1. The classification accuracy for CANFIS was found to be 99.55% which was found to be better when compared to other classification methods. Classification accuracy is almost same for C4.5 and multilayer Perceptron, but the runtime for C4.5 is comparatively higher. The classification matrix of C4.5 obtained for 225 testing sample set is shown in table 5.6.2. The classification matrix provides a comprehensive picture of the classification performance of the classifier. The ideal classification matrix is the one in which the sum of the diagonal is equal to the number of samples. Figure 5.6.1 shows the Output Vs Desired Plot for the test set using CANFIS and genetic algorithm. Once the classification is done the next stage would be finding the best possible treatment and management for Alzheimer's diseases.

Healthcare data mining can be limited by the accessibility of data, because the raw inputs for data mining often exist in different settings and systems, such as administration, clinics, laboratories and more. Hence, the data have to be collected and integrated before data mining can be done. Oakley [21] has suggested a distributed network topology instead of a data warehouse for more efficient data mining, and Friedman and Pliskin[22] have documented a case study of Maccabi Healthcare Services using existing databases to guide subsequent data mining. Secondly, other data problems may arise. These include missing, corrupted, inconsistent, or non-standardized data, such as pieces of information recorded in different formats in different data sources. In particular, the lack of a standard clinical vocabulary is a serious hindrance to data mining. Cios and Moore [23] have argued that data problems in healthcare are the result of the volume, complexity and heterogeneity of medical data and their poor mathematical characterization and non-canonical form. Further, there may be ethical, legal and social issues, such as data ownership and privacy issues, related to healthcare data. The quality of data mining results and applications depends on the quality of data.32 Thirdly, a sufficiently exhaustive mining of data will certainly yield patterns of some kind that are a product of random fluctuations [24]. This is especially true for large data sets with many variables. Hence, many interesting or significant patterns and relationships found in data mining may not be useful. Hand [25] and Murray [26] have warned against using data mining for data dredging or fishing, which is randomly trawling through data in the hope of identifying patterns. Fourthly, the successful application of data mining requires knowledge of the domain area as well as in data mining methodology and tools. Without a sufficient knowledge of data mining, the user may not be aware of or be able to avoid the pitfalls of data mining [27]. Collectively, the data mining team should possess domain knowledge, statistical and research expertise, and IT and data mining knowledge and skills.

DISCUSSION

The healthcare data are not limited to just quantitative data, such as physicians' notes or clinical records, it is necessary to also explore the use of text mining to expand the scope and nature of what healthcare data mining can currently do. In particular, it is useful to be able to integrate data and text mining [28]. It is also useful to look into how digital diagnostic images can be brought into healthcare data mining applications. Some progress has been made in these areas [29.30].

Pharmaceutical companies can benefit from healthcare CRM and data mining, too. By tracking which physicians prescribe which drugs and for what purposes, pharmaceutical companies can decide whom to target, show what is the least expensive or most effective treatment plan for an ailment, help identify physicians whose practices are suited to specific clinical trials (for example, physicians who treat a large number of a specific group of patients), and map the course of an epidemic to support pharmaceutical salespersons, physicians, and patients [31]. Pharmaceutical companies can also employ data mining methods to huge masses of genomic data to predict how a patient's genetic makeup determines his or her response to a drug therapy [32].

The identification and quantification of pharmaceutical information can be extremely useful for patients, physicians, pharmacists, health organizations, insurance companies, regulatory agencies, investors, lawyers, pharmaceutical manufacturers, drug testing companies, etc. For example information on interactions among over-the-counter medicines, interactions between prescription and over-the-counter medicines, interactions among prescription medicines, interactions between any kind of medicine and various foods, beverages, vitamins, and mineral supplements, common characteristics between certain drug groups and offending foods, beverages, medicines, will yield very useful results using data mining techniques [33]. One of the major problems with pharmaceutical data is actually a lack of information. Most health care providers simply do not have the time to fill out reports of possible adverse drug reactions. Furthermore, it is expensive and time-consuming for pharmaceutical companies to perform a thorough job of data collection, especially when most of the information is not required by law [34].

The delivery of health care has always been information intensive, and there are signs that the industry is recognizing the increasing importance of information processing in the new managed care environment [35]. Most healthcare institutions lack the appropriate information systems to produce reliable reports with respect to other information than purely financial and volume related statements [36] stress that. The management of pharma industry starts to recognize the relevance of the definition of drugs and products in relation to management of critical information. In the turmoil between costs, care-results and patient satisfaction the right balance is needed and can be found in upcoming information and communication technology. Research shows that successful decision systems enriched with analytical solutions are necessary for healthcare information systems [37-39].

A user-interface may be designed to accept all kinds of information from the user (e.g., weight, sex, age, foods consumed, reactions reported, dosage, length of usage). Then, based on the information in the databases and the relevant data entered by the user, a list of warnings or known reactions (accompanied by probabilities) should be reported. Note that user profiles can contain large amounts of information, and efficient and effective data mining tools need to be developed to probe the databases for relevant information. Second, the patient's (anonymous) profile should be recordedalong with any adverse reactions reported by the patient, so that future correlations can be reported. Over time, the databases will become much larger, and interaction data for existing medicines will become more complete [34].

Data mining can provide intelligence in terms of: drug positioning information, patent population characteristics, indications of what the drug is being used for, prescribing physician characteristics, regional preferences, prevalence of diseases, preferred drug for diseases, procedures being performed, disease-related information.

APPLICATIONS OF DATA MINING IN THE PHARMACEUTICAL INDUSTRY

- Clinical data analysis – clinical data analysis evaluates and streamlines from large amount of information. Data mining helps to see trends, irregularity, and risk during product development and launch.
- Marketing and sales analysis –the identification of the most profitable product and allocation of marketing funds. Data mining here helps to examine consumer behavior in terms of prescription renewal and product purchases.
- Customer analysis – using data mining one can develop more targeted customer profiles that focus not only on products, but also on the ability to pay for them by analyzing historical health trends in combination with demographics.
- Identify and target individuals and demographics that could be considered "undiagnosed" with educational campaigns whose goal is to encourage these individuals to get screened and tested for possible issues.
- Combine product sales information with customer groups and customer channel information to analyze what tends to lead customers to fill prescriptions at a more consistent rate or what leads physicians to prescribe certain drugs at a higher rate.
- Operations and financial analysis – analyze the prescription activity in a geographic region or area to make sales force adjustments according to market size or penetration.
- Dissect buying trends from the largest customers (managed care providers and governments) to proactively create price points that benefit both the buyer and the organization.
- Sales and marketing analysis – provide mobile analytics to a sales force that is consistently disconnected, allowing them to answer not only detailed drug information questions, but also historical and trending questions.
- Target physicians who have high prescription rates of a certain drug or treatment with new drug information that treat complementary symptoms or conditions.
- Product analysis – analyze buying tendencies and treatment outcomes to create more drug and product variations tailored directly towards different age groups and risk factors.
- Combine demographics and patient historical trends to target "quality of life" needs of patients (i.e. lifestyle drugs) that improve the day-to-day living standards of patients, especially for non-acute medical conditions.
- Supply chain analysis – improve production schedules through analysis of which pharma products stay on the shelves the longest and how well each pharma product is selling.
- Manage inventories more efficiently based on historical trends and patient behavior to prevent stock-outs at retail and pharmacy locations [34].

DATA MINING IN DISCOVERY OF NEW DRUGS

Here the data mining techniques used are clustering, classification and neural networks. The goal is to determine compounds with similar activity. The reason for this is the compounds with similar activity may behave similarly. This should be performed when we know the compound and are looking for something better or when we do not know the compound, but have desired activity and want to find compound that exhibits similar activity. This can be achieved by clustering the molecules into groups according to the chemical properties of the molecules via cluster analysis [40]. This way every time a new molecule is discovered it can be grouped with other chemically identical molecules. This would help the researchers in finding out with therapeutic group the new molecule would belong to. Mining can help us to measure the chemical activity of the molecule on specific disease say tuberculosis and find out which part of the molecule is causing the action. This way we can combine a vast number of molecules forming a super molecule with only the specific part of the molecule that is responsible for the action and inhibiting the other parts. This would greatly reduce the adverse effects associated with drug actions.

DATA MINING IN CLINICAL TRIALS

Pharmaceutical companies test drugs in patients on larger scale. The company has to keep track of data about patient progress. The data obtained undergoes statistical analysis to determine the success of a trial. Data are generally reported to a food and drug administration department and inspected closely. Too many adverse reactions might indicate a drug is too dangerous. Adverse events are reported to the food and drug administration when a link is suspected. The effectiveness of the drug is often measured by how soon the drug deals with the medical condition [41]. A simple association technique could help to measure the outcomes that would greatly enhance the patient's quality of life such as faster restoration of the body's normal functioning.

A data mining contest [42] was conducted held to predict the molecular bioactivity for a drug design; specifically, determining which organic molecules would bind to a target site on thrombin. The predictions were based on about 500 megabytes of data on approximately 1,900 organic molecules, each with more than 130,000 attributes (or dimensions, as they are called in data mining). This was a challenging problem not only because of the large number of attributes but because only 42 of the compounds (2.2%) were active. The relatively small number of cases (organic molecules in this example), compared to the large number of attributes, makes the problem even more difficult. Of the 136 contest entries, about 10% achieved the impressive result of more than 60% accuracy [43], with the winner, Jie Cheng of the Canadian Imperial Bank of Commerce, reaching almost 70% accuracy.

DATA MINING AND THE FUTURE OF PHARMACEUTICAL INDUSTRY

New approaches are needed in order to fully realize the potential of technologies that allow for the creation, acquisition, storage and analysis of databases of unprecedented size and complexity. The modern company needs to view itself as a data machine whose primary business is collecting, processing, analysing and using its data as the primary resource of the company. This trend will continue and accelerate. Genomic and related technologies allow for more data to be collected on each patient both during the development of a drug and after it is marketed. General IT technologies continue to move in a direction that allows more data to be collected, processed and analysed. There are demands for more information about each product from patients, physicians and regulators.

Appropriate data, the structure, job descriptions and likely organizations will change. We are living in an age that emphasizes a great increase in collection and use of data. It is not surprising that the pharmaceutical industry, which has been an information industry for decades, is even more affected by these trends. The fact that some of these changes have happened so quickly and that many in pharmaceutical and medical research did not recognize the previous emphasis on information may have disguised the cataclysmic events taking place. Hence, data mining will play a major and pivotal role in pharmaceutical and health care industry in future.

REFERENCES

[1] Sandhya Joshi, P. Deepa Shenoy, Venugopal K R, L M Patnaik Data Analysis and Classification of Various Stages of Dementia Adopting Rough Sets Theory. *International Journal on Information processing,4(1), 86-99, 2010.*

[2] Sandhya Joshi, P. Deepa Shenoy, Venugopal K R, L M Patnaik A. Classification and treatment of different stages of Alzheimer's disease using various machine learning methods International Journal of Bioinformatics Research, ISSN: 0975–3087, Volume 2, Issue 1, 2010, pp-44-52.

[3] Sandhya Joshi, P. Deepa Shenoy, Venugopal K R, L M Patnaik **"Classification of Neurodegenerative Disorders Based on Major** Risk Factors **Employing** Machine Learning **Techniques"** *International Journal of Engineering and Technology*. ISSN: 1793-8236, Vol.2, No.4, August 2010. pp. 350-355.

[4] Sandhya Joshi, P. Deepa Shenoy, Venugopal K R, L M Patnaik Statistical Classification of Magnetic Resonance Images of Brain Employing Random Forest Classifier. *International Journal of Machine Intelligence*, ISSN: 0975–2927, vol 1, Issue 2, 2009, pp.55-61.

[5] Silver, M. Sakata, T. Su, H.C. Herman, C. Dolins, S.B. & O'Shea, M.J. (2001). Case study: how to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management*, 15(2), 155-164.

[6] Benko, A. & Wilson, B. (2003). Online decision support gives plans an edge. *Managed Healthcare Executive*, 13(5), 20.

[7] Gillespie, G. (2000). There's gold in them thar' databases. *Health Data Management*, 8(11), 40-52.

[8] Kolar, H.R. (2001). Caring for healthcare. *Health Management Technology*, 22(4), 46-47.

[9] Relles, D. Ridgeway, G. & Carter, G. (2002). Data mining and the implementation of a prospective payment system for inpatient rehabilitation. *Health Services & Outcomes Research Methodology*, 3(3-4), 247-266.

[10] Robert D., Small and Herbert A. Edelstein. Data mining in the pharmaceutical industry. Drug Discovery World Fall 2001.

[11] Kincade, K. (1998). Data mining: digging for healthcare gold. *Insurance & Technology*, 23(2), IM2-IM7.

[12] Milley, A. (2000). Healthcare and data mining. *Health Management Technology*, 21(8), 44-47.

[13] Young, J. & Pitta, J. (1997). Wal-Mart or Western Union? United HealthCare Corp. *Forbes*, 160(1), 244.
[14] Veletsos, A. (2003). Getting to the bottom of hospital finances. *Health Management Technology*, 24(8), 30-31.
[15] Dakins, D.R. (2001). Center takes data tracking to heart. *Health Data Management*, 9(1), 32-36.
[16] Johnson, D.E.L. (2001). Web-based data analysis tools help providers, MCOs contain costs. *Health Care Strategic Management*, 19(4), 16-19.
[17] Schuerenberg, B.K. (2003). An information excavation. *Health Data Management*, 11(6), 80-82.
[18] Paddison, N. (2000). Index predicts individual service use. *Health Management Technology*, 21(2), 14-17.
[19] Hallick, J.N. (2001), Analytics and the data warehouse. *Health Management Technology*, 22(6), 24-25.
[20] Anonymous. (1999). Texas Medicaid Fraud and Abuse Detection System recovers $2.2 million, wins national award. *Health Management Technology*, 20(10), 8.
[21] Oakley, S. (1999). Data mining, distributed networks and the laboratory. *Health Management Technology*, 20(5), 26-31.
[22] Friedman, N.L. & Pliskin, N. (2002). Demonstrating value-added utilization of existing databases for organizational decision-support. *Information Resources Management Journal*, 15(4), 1-15.
[23] Cios, K.J. & Moore, G.W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, 26(1), 1-24.
[24] Chopoorian, J.A. Witherell, R. Khalil, O.E.M. & Ahmed, M. (2001). Mind your own business by mining your data. *SAM Advanced Management Journal*, 66(2), 45-51.
[25] Hand, D.J. (1998). Data mining: statistics and more? *The American Statistician*, 52(2), 112-118.
[26] Murray, L.R. (1997). Lies, damned lies and more statistics: the neglected issue of multiplicity in accounting research. *Accounting and Business Research*, 27(3), 243-258.
[27] McQueen, G. & Thorley, S. (1999). Mining fool's gold. *Financial Analysts Journal*, 55(2), 61-72.
[28] Cody, W.F. Kreulen, J.T. Krishna, V. & Spangler, W.S. (2002). The integration of business intelligence and knowledge management. *IBM* Systems *Journal*, 41(4), 697-713.
[29] Ceusters, W. (2001). Medical natural language understanding as a supporting technology for data mining in healthcare. In *Medical Data Mining and Knowledge Discovery*, Cios, K. J. (Ed.), Physica- Verlag Heidelberg, New York, 41-69.
[30] Megalooikonomou, V. & Herskovits, E.H. (2001). Mining structurefunction associations in a brain image database. In *Medical Data Mining* and *Knowledge Discovery*, Cios, K. J. (Ed.), Physica-Verlag Heidelberg, New York, 153-180.
[31] Brannigan, M. (1999). Quintiles seeks mother lode in health "data mining." *Wall Street Journal*, March 2, 1.
[32] Thompson, W. & Roberson, M. (2000). *Making predictive medicine possible*. R & D, 42(6), E4-E6.
[33] Hian Chye Koh and Gerald Tan. Data Mining Applications in Healthcare. Journal of Healthcare Information Management — Vol. 19, No. 2, 5-11.
[34] Jayanthi Ranjan Data mining in pharma sector: Benefits , International Journal of health care quality assurance, 22(1), 2009, pp. 82-92.
[35] Morrisey, J. (1995), "Managed care steers info systems", Modern Healthcare, Vol. 25 No. 8.
[36] Prins, S. and Stegwee, R.A. (2000), "Zorgproducten en geintegreerde infromatie systemen", Handboek sturen met zorgproducten, F3100-3, Bohn Stafleu Van Loghum, Houten, December (in Dutch).
[37] Zuckerman, A.M. (2006), Healthcare Strategic Planning, Prentice-Hall of India, New Delhi.
[38] Armoni, A. (2002), Effective Healthcare Information Systems, IRM Press, Hershey, PA.
[39] Rada, R. (2002), Information Systems for Healthcare Enterprises, Hypermedia Solutions,
a. Liverpool.
[40] De Cooman, F. (2005), Data Mining in a Pharmaceutical Environment, Belgium Press, Brussels.
[41] Graedon, J. and Graedon, T. (1996), People's Guide to Deadly Drug Interactions, St Martin's Press, New York, NY.
[42] Knowledge Discovery in Databases: 10 Years After. SIGKDD Explorations, ACM SIGKDD V.1 59-61, 2000.
[43] Edelstein, Herb. Data Mining Technology Report: 2002, Forthcoming, Two Crows Corporation.