Data Mining Applications in Healthcare

Hian Chye Koh and Gerald Tan

A B S T R A C T

Data mining has been used intensively and extensively by many organizations. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Data mining applications can greatly benefit all parties involved in the healthcare industry. For example, data mining can help healthcare insurers detect fraud and abuse, healthcare organizations make customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services.

The huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making.

This article explores data mining applications in healthcare. In particular, it discusses data mining and its applications within healthcare in major areas such as the evaluation of treatment effectiveness, management of healthcare, customer relationship management, and the detection of fraud and abuse. It also gives an illustrative example of a healthcare data mining application involving the identification of risk factors associated with the onset of diabetes. Finally, the article highlights the limitations of data mining and discusses some future directions.

K E Y W O R D S

- Healthcare management
- Customer relationship management
- Healthcare applications
- Data mining methodology and techniques
- Data mining applications
- Predictive modeling

Introduction

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.[1] 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]

Data mining is not new—it has been used intensively and extensively by financial institutions, for credit scoring and fraud detection; marketers, for direct marketing and cross-selling or up-selling; retailers, for market segmentation and store layout; and manufacturers, for quality control and maintenance scheduling.

In healthcare, data mining is becoming increasingly popular, if not increasingly essential. 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. Recently, there have been reports of successful data mining applications in healthcare fraud and abuse detection.[2]

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, Benko and Wilson argue.[6] Data can be a great asset to healthcare organizations, but they have to be first transformed into information.

> *"In healthcare, data mining is becoming increasingly popular, if not increasingly essential."*

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]

There are also other factors boosting data mining's popularity. For instance, as a result of the Balanced Budget Act of 1997, the Centers for Medicare and Medicaid Services must implement a prospective payment system based on classifying patients into case-mix groups, using empirical evidence that resource use within each case-mix group is relatively constant. CMS has used data mining to develop a prospective payment system for inpatient rehabilitation.[9]

The healthcare industry can benefit greatly from data mining applications. The objective of this article is to explore relevant data mining applications by first examining data mining methodology and techniques; then, classifying potential data mining applications in healthcare; next, giving an illustration of a healthcare data mining application; and finally, highlighting the limitations of data mining and offering some future directions.

Data Mining

Data mining can be considered a relatively recently developed methodology and technology, coming into prominence only in 1994.[10] It aims to identify valid, novel, potentially useful, and understandable correlations and patterns in data[11] by combing through copious data sets to sniff out patterns that are too subtle or complex for humans to detect.[12]

Cross-Industry Standard Process for Data Mining, or CRISP-DM (see www.crisp-dm.org) proposes the following methodology for data mining: business understanding, data understanding and preparation, modeling, evaluation, and deployment.

Business understanding is critical because it identifies the business objectives and, thus, the success criteria of data mining projects. Further, as the term "data mining" implies, data is a crucial component—no data means no mining. Hence, CRISP-DM includes data understanding and data preparation—in other words, sampling and data transformation—as essential antecedents for modeling.

The modeling stage is the actual data analysis. Most data mining software include online analytical processing; traditional statistical methods, such as cluster analysis, discriminant analysis and regression analysis; and non-traditional statistical analysis, such as neural networks, decision trees, link analysis and association analysis. This extensive range of techniques is not surprising in light of the fact that data mining has been viewed as the offspring of three different disciplines, namely database management, statistics, and computer science, including artificial intelligence and machine learning.

The evaluation stage enables the comparison of models and results from any data mining model by using a common yardstick, such as lift charts, profit charts, or diagnostic classification charts. Finally, deployment relates to the actual implementation and operationalization of the data mining models.

Data mining techniques can be broadly classified based on what they can do, namely description and visualization; association and clustering; and classification and estimation, which is predictive modeling.

Description and visualization can contribute greatly towards understanding a data set, especially a large one, and detecting hidden patterns in data, especially complicated data containing complex and non-linear interactions. They are usually performed before modeling is attempted and represent data understanding in the CRISP-DM methodology.

In association, the objective is to determine which variables go together. For example, market-basket analysis (the most popular form of association analysis) refers to a technique that generates probabilistic statements such as, "If patients undergo treatment A, there is a 0.35 probability that they will exhibit symptom Z." Such information can be useful for investigating associative relationships in healthcare. With clustering, the objective is to group objects, such as patients, in such a way that objects belonging to the same cluster are similar and objects belonging to different clusters are dissimilar. In Koh and Leong,[13] clustering is used to group readmitted patients to better profile and understand such patients.

The most common and important applications in data mining probably involve predictive modeling. Classification refers to the prediction of a target variable that is categorical in nature, such as predicting healthcare fraud vs. non-fraud. Estimation, on the other hand, refers to the prediction of a target variable that is metric (i.e., interval or ratio) in nature, such as predicting the length of stay or the amount of resource utilization. For predictive modeling, the data mining techniques commonly used include traditional statistics, such as multiple discriminant analysis and logistic regression analysis. They also include non-traditional methods developed in the areas of artificial intelligence and machine learning. The two most important models of these are neural networks and decision trees. More details on data mining techniques can be found in Berry and Linoff.[14]

Healthcare Data Mining Applications

There is vast potential for data mining applications in healthcare. Generally, these can be grouped as the evaluation of treatment effectiveness; management of healthcare; customer relationship management; and detection of fraud and abuse. More specialized medical data mining, such as predictive medicine and analysis of DNA micro-arrays, lies outside the scope of this paper.

Treatment effectiveness. Data mining applications can be developed to evaluate the effectiveness of medical treatments. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective.[2] 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.[1]

Along this line, United HealthCare has mined its treatment record data to explore ways to cut costs and deliver better medicine.[15] 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] A good account of data mining applications at Florida Hospital also can be found in Gillespie[7] and Veletsos.[16]

Other data mining applications related to treatments include 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.[2]

Healthcare management. To aid healthcare management, data mining applications can be developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims.

For example, to develop better diagnosis and treatment protocols, the Arkansas Data Network looks at readmission and resource utilization and compares its data with current scientific literature to determine the best treatment options, thus using evidence to support medical care.[1] Also, the 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.[1] 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.[17]

As another 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.[1] Data mining also can be used to identify and understand high-cost patients.[5]

Johnson[18] has suggested that, at a higher level, data mining can facilitate comparisons across healthcare groups of things such as practice patterns, resource utilization, length of stay, and costs of different hospitals. Recently, Sierra Health Services has used data mining extensively to identify areas for quality improvements, including treatment guidelines, disease management groups, and cost management.[19]

Data mining can be used to analyze massive volume of data and statistics to search for patterns that might indicate an attack by bio-terrorists.[20] The Lightweight Epidemiological Advanced Detection Emergency Response System (LEADERS) is one such effort. In the past, LEADERS has uncovered several disease outbreaks. Data mining also

Table 1. Significant Variables

Level* / Variables		Chi-square Statistic (p-value)	
1	/ Age	463.84	(< 0.0001)
2	/ Body mass index	184.99	(< 0.0001)
	Body mass index	98.66	(< 0.0001)
3	/ Age	26.72	(< 0.0001)
	Waist hip ratio	32.33	(< 0.0001)
	Waist hip ratio	40.34	(< 0.0001)
4	/ Body mass index	17.63	(< 0.0001)
	Number of times of exercise per week	10.19	(0.0014)
	Age	10.49	(0.0013)

* Level refers to the depth of the decision tree as shown in Figure 1 Decision Tree.

can be used for hospital infection control[12] or as an automated early-warning system in the event of epidemics. A syndromic system, based on patterns of symptoms, is likely to be more efficient and effective than a traditional system that is based on diagnosis. An early warning of the global spread of the SARS virus is an example of the usefulness of a syndromic system based on data mining.[21]

Customer relationship management. While customer relationship management is a core approach in managing interactions between commercial organizations—typically banks and retailers—and their customers, it is no less important in a healthcare context. Customer interactions may occur through call centers, physicians' offices, billing departments, inpatient settings, and ambulatory care settings.

As in the case of commercial organizations, data mining applications can be developed in the healthcare industry to determine the preferences, usage patterns, and current and future needs of individuals to improve their level of satisfaction.[4] These applications also can be used to predict other products that a healthcare customer is likely to purchase, whether a patient is likely to comply with prescribed treatment or whether preventive care is likely to produce a significant reduction in future utilization.

Through the use of data mining, 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.[22] 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.[22]

Miller[2] has suggested that 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. Also, Hallick[23] has suggested that CRM in healthcare can help promote disease education, prevention, and wellness services. Kolar[8] and Veletsos[16] also have reported that 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. Rafalski[24] has studied Sinai Health System's use of data mining for healthcare marketing and CRM.

Lastly, 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.[25] Pharmaceutical companies can also apply data mining to huge masses of genomic data to predict how a patient's genetic makeup determines his or her response to a drug therapy.[26]

Fraud and abuse. Data mining applications that attempt 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,

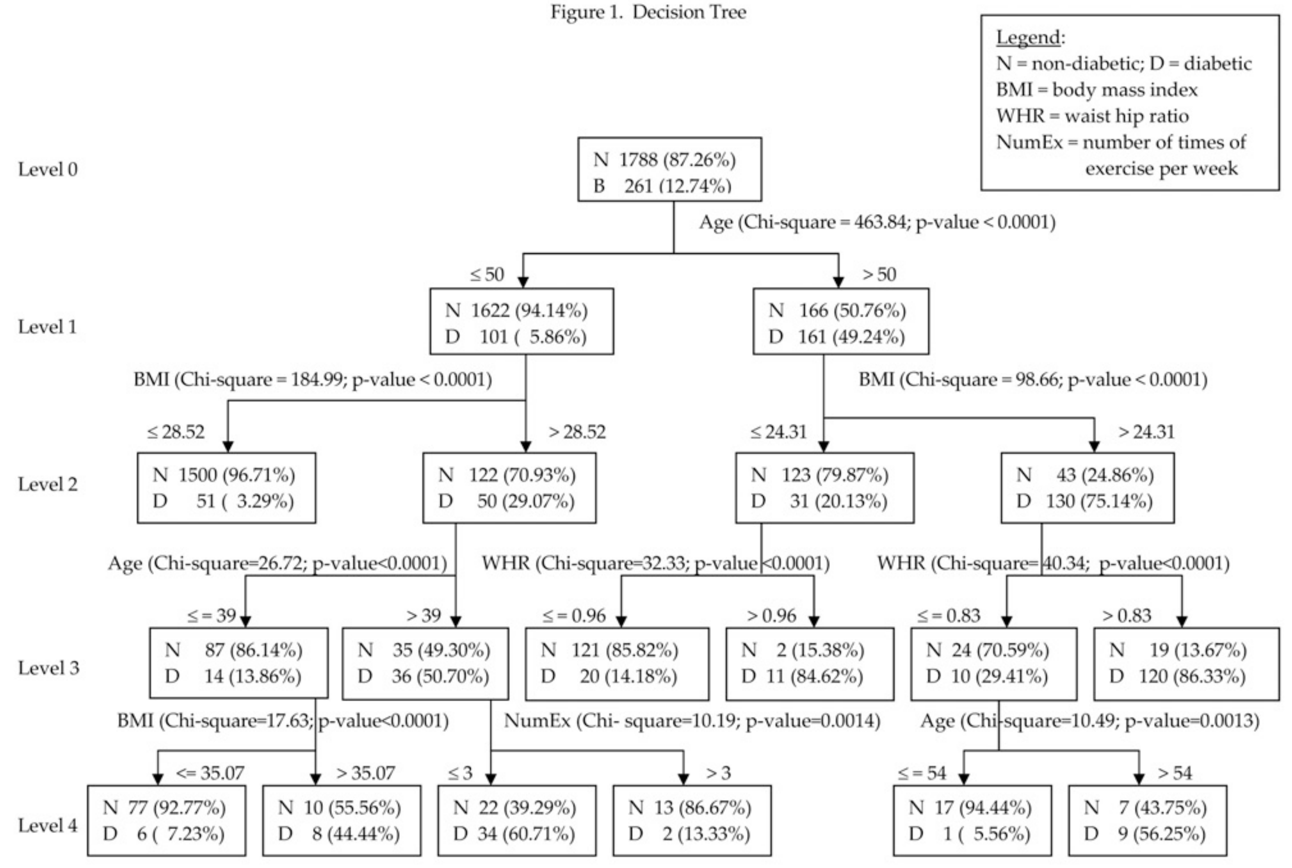

Figure 1. Decision Tree

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.[2]

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,[3] and the Australian Health Insurance Commission has estimated tens of millions of dollars of annual savings. Another successful example of using data mining to detect fraud and abuse is the Texas Medicaid Fraud and Abuse Detection System, which recovered $2.2 million and identified 1,400 suspects for investigation in 1998 after operating for less than a year.[27] In recognition of its success, the Texas system has won a national award for outstanding achievement and top honors for innovative use of technology.

One Example of Data Mining

To illustrate a data mining application in healthcare, suppose that as part of its healthcare management program, HealthOrg (a fictional healthcare organization) is interested in finding out how certain variables are associated with the onset of diabetes. The purpose of this data mining application is to identify high-risk individuals so appropriate messages can be communicated to them.

A dataset exists in the data warehouse of HealthOrg that contains the following seven variables of particular interest to HealthOrg: gender, age, body mass index (BMI), waist-hip ratio (WHR), smoking status, the number of times a patient exercises per week, and onset of diabetes, which is the target variable, measured by a dichotomous variable indicating whether an individual has tested positive for diabetes. The dataset comprises 262, or 12.78 percent, of positive diabetic cases and 1,778, or 87.22 percent, of negative non-diabetic cases.

After reviewing the work of Breault et al on the data mining of a diabetic data warehouse,[28] HealthOrg decides that the decision tree is an appropriate data mining technique to use to find out how certain variables are associated with the onset of diabetes. Decision trees have the advantage of ease of interpretation and visualization. For the purpose of this illustration, SPSS's Clementine data mining software is used.

Results are summarized in Table 1 (Significant Variables) and Figure 1 (Decision Tree). As shown in Table 1, age, body mass index, waist-hip ratio, and the number of exercise events per week are significantly associated

Table 2. Classification Table

	Classified Status		
Actual Status	Non-diabetic	Diabetic	Total
Non-diabetic	1728	60	1788
Diabetic	80	182	262
Total	1808	242	2050

Computation of Accuracy Rates

1. Rate of classifying actual diabetic cases correctly:
   Sensitivity = 182/262 = 69.47%
2. Rate of classifying actual non-diabetic cases correctly:
   Specificity = 1728/1788 = 96.64%
3. Rate of actual diabetic cases for individuals classified as diabetic cases:
   True positives = 182/242 = 75.21%
4. Rate of actual non-diabetic cases for individuals classified as non-diabetic cases:
   True negatives = 1728/1808 = 95.58%

with the onset of diabetes at a 0.01 significance level. Figure 1 gives a visualization of the decision tree results and facilitates interpretation.

Before proceeding to interpret the results, it is important to evaluate the performance—in other words, the accuracy rates—of the decision tree. For simplicity, only the in-sample performance is discussed here. The classification results are summarized in Table 2 (Classification Table). As shown, the accuracy rates for non-diabetic cases are very high. In particular, of the 1,778 non-diabetic cases in the dataset, the decision tree correctly classifies 1,728 as non-diabetic, an accuracy rate of 96.64 percent. Further, for the 1,808 individuals classified by the decision tree as non-diabetic cases, 1,728, or 95.58 percent, are actually non-diabetic cases.

The accuracy rates for predicting diabetic cases are lower but still can be deemed to be sufficient for the objective of the data mining application. In particular, for the 262 diabetic cases in the dataset, the decision tree correctly classifies 182 as diabetic cases, an accuracy rate of 69.47 percent. Further, for the 242 individuals classified by the decision tree as diabetic cases, 182, or 75.21 percent, are actually diabetic.

With the performance of the decision tree deemed adequate, the decision tree can be further interpreted as follows. The results show that age is the most important factor associated with the onset of diabetics (see Level 1 in Figure 1), with individuals older than age 50 showing significant higher risk of diabetes, compared with their counterparts. At the next level (see Level 2 in Figure 1), the body mass index is the next most important factor associated with the onset of diabetes. In particular, individuals younger than 50 with a BMI of less than 28.52 have a very low risk of diabetes (a probability of only 3.29 percent in the cohort). Further, increasing levels of BMI is associated with increasing risk of diabetes. For individuals older than 50 with a BMI greater than 24.31, the risk is 75.14 percent.

> *"The most common and important applications in data mining probably involve predictive modeling."*

As shown at Level 3 of Figure 1, the waist-hip ratio is the next most important factor, with increasing WHR associated with an increased risk of diabetes. For example, the highest-risk individuals in the database are those above 50 years old with BMI of more than 24.31 and WHR above 0.83—their risk probability is 86.33 percent in this cohort.

The remaining nodes in the decision tree can be interpreted in a similar manner. However, nodes further down in the decision tree are usually less important in view of their smaller sample sizes and also because of their more restrictive sub-setting at lower levels.

This decision tree can help HealthOrg identify high-risk individuals so appropriate messages can be communicated to them. For example, HealthOrg can launch a health promotion campaign to educate people that large BMI and

WHR are risk factors associated with the onset of diabetes. It also can scan through its patient databases to identify individuals for further counseling or medical check-ups.

Limitations of Data Mining

Data mining applications can greatly benefit the healthcare industry. However, they are not without limitations.

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. While several authors and researchers have suggested that a data warehouse be built before data mining is attempted, that can be a costly and time-consuming project. On a positive note, a data warehouse has been successfully built by Intermountain Health Care from five different sources—a clinical data repository, acute care case-mix system, laboratory information system, ambulatory case-mix system, and health plans database—and used to find and implement better evidence-based clinical solutions. Oakley[29] has suggested a distributed network topology instead of a data warehouse for more efficient data mining, and Friedman and Pliskin[30] 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.[7] Cios and Moore[31] 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]

"Data mining applications that attempt to detect fraud and abuse often establish norms and then identify unusual or abnormal patterns of claims…"

Thirdly, a sufficiently exhaustive mining of data will certainly yield patterns of some kind that are a product of random fluctuations.[33] 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. Murray[34] and Hand[33] 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.[35] Collectively, the data mining team should possess domain knowledge, statistical and research expertise, and IT and data mining knowledge and skills.

Finally, healthcare organizations developing data mining applications also must make a substantial investment of resources, particularly time, effort, and money. Data mining projects can fail for a variety of reasons, such as lack of management support, unrealistic user expectations, poor project management, inadequate data mining expertise, and more. Data mining requires intensive planning and technological preparation work. In addition, physicians and executives have to be convinced of the usefulness of data mining and be willing to change work processes. Further, all parties involved in the data mining effort have to collaborate and cooperate.[7]

Future Directions

Data mining applications in healthcare can have tremendous potential and usefulness. However, the success of healthcare data mining hinges on the availability of clean healthcare data. In this respect, it is critical that the healthcare industry consider how data can be better captured, stored, prepared, and mined. Possible directions include the standardization of clinical vocabulary and the sharing of data across organizations to enhance the benefits of healthcare data mining applications.

Further, as 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.[36] 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.[37,38]

Finally, the authors hope this paper can make a contribution to the data mining and healthcare literature and practice. It also is hoped that this paper can help all parties involved in healthcare reap the benefits of healthcare data mining.

About the Authors

Hian Chye Koh is an associate professor and vice-dean at the Nanyang Business School, Nanyang Technological University, Singapore. His teaching and research interests

are in data mining and its applications.

Gerald Tan is the manager of products and services at SPSS Singapore. His professional and consultancy expertise lies in the area of statistics, data mining, and business intelligence and solutions.

References

1. Kincade, K. (1998). Data mining: digging for healthcare gold. *Insurance & Technology*, 23(2), IM2-IM7.
2. Milley, A. (2000). Healthcare and data mining. *Health Management Technology*, 21(8), 44-47.
3. Christy, T. (1997). Analytical tools help health firms fight fraud. *Insurance & Technology*, 22(3), 22-26.
4. Biafore, S. (1999). Predictive solutions bring more power to decision makers. *Health Management Technology*, 20(10), 12-14.
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. Trybula, W.J. (1997). Data mining and knowledge discovery. *Annual Review of Information Science and Technology*, 32, 197-229.
11. Chung, H.M. & Gray, P. (1999). Data mining. *Journal of Management Information Systems*, 16(1), 11-16.
12. Kreuze, D. (2001). Debugging hospitals. *Technology Review*, 104(2), 32.
13. Koh, H.C. & Leong, S.K. (2001). Data mining applications in the context of casemix. *Annals, Academy of Medicine*, 30(4, Supplement), 41-49.
14. Berry, M.J.A. & Linoff, G.S. (1997). *Data Mining Techniques: For Marketing, Sales and Customer Support.* New York: John Wiley & Sons Inc.
15. Young, J. & Pitta, J. (1997). Wal-Mart or Western Union? United HealthCare Corp. *Forbes*, 160(1), 244.
16. Veletsos, A. (2003). Getting to the bottom of hospital finances. *Health Management Technology*, 24(8), 30-31.
17. Dakins, D.R. (2001). Center takes data tracking to heart. *Health Data Management*, 9(1), 32-36.
18. Johnson, D.E.L. (2001). Web-based data analysis tools help providers, MCOs contain costs. *Health Care Strategic Management*, 19(4), 16-19.
19. Schuerenberg, B.K. (2003). An information excavation. *Health Data Management*, 11(6), 80-82.
20. Piazza, P. (2002). Health alerts to fight bioterror. *Security Management*, 46(5), 40.
21. Brewin, B. (2003). New health data net may help in fight against SARS. *Computerworld*, 37(17), 1, 59.
22. Paddison, N. (2000). Index predicts individual service use. *Health Management Technology*, 21(2), 14-17.
23. Hallick, J.N. (2001), Analytics and the data warehouse. *Health Management Technology*, 22(6), 24-25.
24. Rafalski, E. (2002). Using data mining and data repository methods to identify marketing opportunities in healthcare. *Journal of Consumer Marketing*, 19(7), 607-613.
25. Brannigan, M. (1999). Quintiles seeks mother lode in health "data mining." *Wall Street Journal*, March 2, 1.
26. Thompson, W. & Roberson, M. (2000). *Making predictive medicine possible.* R & D, 42(6), E4-E6.
27. Anonymous. (1999). Texas Medicaid Fraud and Abuse Detection System recovers $2.2 million, wins national award. *Health Management Technology*, 20(10), 8.
28. Breault, J.L, Goodall, C.R. & Fos, P.J. (2002). Data mining a diabetic data warehouse. *Artificial Intelligence in Medicine*, 26(1), 37-54.
29. Oakley, S. (1999). Data mining, distributed networks and the laboratory. *Health Management Technology*, 20(5), 26-31.
30. 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.
31. Cios, K.J. & Moore, G.W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, 26(1), 1-24.
32. 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.
33. Hand, D.J. (1998). Data mining: statistics and more? *The American Statistician*, 52(2), 112-118.
34. 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.
35. McQueen, G. & Thorley, S. (1999). Mining fool's gold. *Financial Analysts Journal*, 55(2), 61-72.
36. 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.

37. 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.

38. Megalooikonomou, V. & Herskovits, E.H. (2001). Mining structure-function associations in a brain image database. In *Medical Data Mining* and *Knowledge Discovery*, Cios, K. J. (Ed.), Physica-Verlag Heidelberg, New York, 153-180.