Data Mining in Health and Medical Information

Peter A. Bath
University of Sheffield

Introduction

Data mining (DM) is part of a process by which information can be extracted from data or databases and used to inform decision making in a variety of contexts (Benoît, 2002; Michalski, Bratka & Kubat, 1997). DM includes a range of tools and methods for extracting information; their use in the commercial sector for knowledge extraction and discovery has been one of the main driving forces in their development (Adriaans & Zantinge, 1996; Benoît, 2002). DM has been developed and applied in numerous areas. This review describes its use in analyzing health and medical information.

Recent *ARIST* reviews of DM have discussed the mining of structured data (Trybula, 1997), textual data (Trybula, 1999), and DM as part of the knowledge discovery process (Benoît, 2002) in different contexts and domains. This chapter complements these reviews by exploring DM in health and medicine and its suitability in these areas. Other recent reviews have discussed DM tools in health and medicine (e.g., Horn, 2001; Lavrač N, 1999a; Maojo & Sanandrés, 2000; McSherry, 1999; Peña-Reyes & Sipper, 2000), and specific reviews of particular tools/methods in this domain have described artificial neural networks (Baxt, 1995; Cross, Harrison, & Kennedy, 1995; Dybowski & Gant, 1995;

Liestol, Anderson, & Anderson 1994; Lisboa, 2002; Tu, 1996), machine learning methods (Lavrač, 1999b), and computer-based clinical decision support systems (Johnston, Langton, Hayes & Mathieu, 1994). This review also considers the importance of statistics in the DM process; numerous general medical statistics texts are, of course, available (e.g., Altman, 1991; Bland, 2000; Daly & Bourke, 2000).

Outline, Scope, and Limitations of the Review

This review provides an overview of the range of DM tools that have been applied in health/medicine and examines the issues that are affecting their development and uptake in routine clinical practice. However, developments in DM in other application areas are beyond the scope of the present chapter. The review also discusses the confusion surrounding definitions of DM and examines the potential of DM in the health/medicine domain. Traditional descriptive and inferential statistical methods of analyzing data are outlined and the importance of statistics in the DM process is discussed. The review considers statistical and non-statistical methods of analyzing data and the relationship between them. Although this chapter emphasizes the importance of using statistical tools to verify results as part of the data mining process, it is beyond the scope of the review to describe detailed applications of statistical methods in health/medicine. Different methods of DM that have been employed in health/medicine and application areas are described. The review discusses challenges that must be overcome for DM techniques to become both widely used in health/medical research and part of routine practice.

The use of DM techniques in related areas, such as analyzing genomic databases, is outside the scope of the the present chapter and has been covered elsewhere (Bertone & Gerstein, 2001; Luscombe, Greenbaum, & Gerstein, 2001; Miller, 2000). The review focuses on DM tools for analyzing numeric quantitative data and does not consider DM tools such as HINT (Hierarchy INduction Tool) and the DEX (decision support tool), which were developed to process qualitative data (Bohanec, Zupan, & Rajkovič, 2000), or the mining of text data in health/medicine (Swanson, 1987; Swanson & Smalheiser, 1999; Trybula, 1999). The application of DM tools in medical/healthcare practice and research is reviewed but

not applications of DM tools in laboratory environments (Dybowski & Gant, 1995) or in clinical trials (Jones, 2001).

A number of the themes that emerge from the review are centered on technical and human issues affecting the development of DM in health/medicine, the potential of this domain for DM, and specific application areas. Technical issues include the importance of mining high quality data, demonstrating the validity of results obtained through DM using statistics, and evaluating the performance of DM tools by comparison with statistical analyses and through their usability. This requires the multidisciplinary collaboration of healthcare professionals (HCPs) in DM development. Other human issues include developing the trust of HCPs and being able to demonstrate the benefits of using DM. The complexity of humans; the importance of health and consequences of disease at individual, group, and population levels; and our capacity to deal with this complexity encourage the development of DM tools for improving diagnosis, prognosis, and decision making and generating hypotheses in health/medicine.

Definitions of Data Mining

Various definitions of, and synonyms for, DM have emerged in recent years. These are not wholly consistent with each other, and, as noted by Benoît (2002), have created some confusion and suspicion in health/medicine. Benoît (2002, p. 265) defined DM as "a multi-staged process of extracting previously unanticipated knowledge from large databases, and applying the results to decision making" within the larger Knowledge Discovery (KD) process (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The relationship between DM and Knowledge Discovery in Databases (KDD) has been presented in detail elsewhere (see, for example, Adriaans & Zantige, 1996; Benoît, 2002). Here it is sufficient to state that DM is the knowledge extraction stage of the KD process, which also includes the selection, cleaning, and merging of appropriate data from various sources, and coding and re-coding of the data, followed by the presentation and reporting of the results of the DM activities. Data mining encompasses a range of techniques selected on the basis of their suitability for a specific task. DM incorporates not only data analysis, but also

involves determining appropriate research questions and interpreting the results (Richards, Rayward-Smith, Sönksen, Carey, & Weng, 2001).

As Benoît (2002) and Trybula (1999) remark, confusion arises through the inappropriate use of various synonyms for DM. These synonyms include "knowledge discovery," which, as indicated above, is the larger process of which DM is a part. Other terms, such as "information extraction," "pattern discovery," and "pattern identification," are all potentially misleading in that they refer to either the end product of the process or one of the DM methods.

Perhaps the most misleading and potentially damaging synonym for DM is "data dredging" (Benoît, 2002; Trybula, 1999), and, in the context of health/medicine, a sharp distinction must be made between these two processes. "Data dredging" is used to describe the process of analyzing a data set to uncover interesting relationships between the variables or patterns among the data. "Dredging" suggests laboriously trawling through a morass in the hope of finding something worthwhile or useful. This analogy suggests that analysts have no clear a priori idea of what they are searching for, but if they search long enough, some relationship or pattern will emerge; in extremis, this has been termed data torturing (Mills, 1993). The problem with this approach in medicine/health is that spurious relationships and patterns can be identified, which arise by chance, and undue importance may be attached to these Type I errors (Altman, 1991). For example if a data set containing 20 variables was analyzed to identify any statistically significant relationships using chi-square tests, then $(n[n - 1]/2)$, or 190, tests would be carried out. If a significance level of $p \leq 0.05$ was used, then, by definition, 1 in 20, or in this example, 9 or 10 test results could appear to be statistically significant purely by chance. Although methods of dealing with such chance findings have been reported (Altman, 1991; Bland & Altman, 1995), there is controversy concerning the precise use of these adjustments in different situations (Bender & Lange, 1999; Perneger, 1998). Data dredging is therefore widely considered inappropriate due to its lack of clear objectives and the potential to yield spurious results. Data dredging has some value in exploratory data analyses and for hypothesis generation, although as Mills (1993, p.119) states, "hypothesis-generating

studies ... should be identified as such." Furthermore, the hypotheses should be tested using appropriate and robust statistical tests.

DM implies drilling down in a much more focused approach with a clear idea of what is being mined for and with a reasonable expectation of retrieving something worthwhile. Data mining suggests that analysts have a good understanding of the data that they are mining and a clear idea, gained through prior knowledge, of the potentially useful and important information that may be retrieved. For example, for a given data set, data dredging could be used to identify any relationships among all the variables; data mining might seek to identify those variables that best predict whether an event will happen (Bath, Morgan, Pendleton, Clague, Horan, & Lucas, 2000), and statistical methods would test whether there is a significant association between a putative risk factor and an event of interest. Although statistical tests can be used in isolation and data mining can be strengthened through statistical tests, it is imperative that data dredging be used for exploratory purposes only, to generate questions and hypotheses for testing by one or both of the others.

DM also implies a systematic approach to the identification of previously hidden associations, patterns, and relationships (Pendharkar, Yaverbaum, Herman, & Benner, 1999); this may involve both hypothesis generation and hypothesis testing. This approach is often more successful when undertaken in collaboration with domain experts and/or statisticians. Using a DM approach might therefore involve identifying a specific research question/hypothesis, for example through a substantive literature review or a discussion with HCPs, and answering/testing this hypothesis using an existing data source by identifying patterns/relationships/associations centered on a limited number of variables. Although this does not wholly eliminate the risk of identifying patterns/relationships/associations that arise purely by chance, nevertheless adopting a focused approach does reduce this risk and is scientifically justifiable. The distinction between data dredging and DM is particularly important in health/medical research: data dredging can produce unreliable and incorrect information, which could adversely affect clinical practice and decision making (Mills, 1993). Data mining may therefore be defined by the approach that the researcher adopts in analyzing the data as well as by the methods that are used.

The Potential of Data Mining in Health and Medicine

In health/medical care, data are routinely generated and stored as part of the care process, for administrative purposes, or for research (Coiera, 1997; Peña-Reyes & Sipper, 2000; Shortliffe & Blois, 2001). A single healthcare episode or research study may yield hundreds of variables and generate large amounts of data. Even though individual data items may be of little value in their own right, valuable information may be contained among them that is not immediately apparent, but that may be extracted and utilized using DM (Kuo, Chang, Chen, & Lee, 2001). This availability of health/medical data and information, coupled with the need to increase our knowledge and understanding of the biological, biochemical, pathological, psychosocial, and environmental processes by which health and disease are mediated, mean that medicine/health is particularly suitable for DM (Shortliffe & Barnett, 2001; Shortliffe & Blois, 2001). This section outlines sources of medical/health data and discusses the suitability of these for data mining.

A contributing factor to the increased availability of medical/health data is the advent of data warehousing and clinical data repositories (CDRs) (Smith & Nelson, 1999), which allow the integration of data from different sources, including patient administration, medical records, and financial systems. Data warehouses are used for storing aggregated data derived from any of these systems and can be used for retrospective analyses for management and financial purposes. CDRs, in contrast, derive data on individuals from separate clinical systems, such as laboratory test results, medical images, and numeric and textual data, and are used for decision making at the patient level (Smith & Nelson, 1999). The development of data warehouses and CDRs is part of the KD process (Benoît, 2002), and although they differ in functionality and the data they contain, DM offers the potential to exploit data obtained from such disparate sources fully.

Medicine and health deal with complex organisms (humans/patients) and with higher-level processes than other branches of science, such as physics and chemistry (Shortliffe & Blois, 2001). Although some of these higher-level processes may be reduced to lower levels of complexity in certain application areas, this can be inappropriate and unhelpful in

medicine/health, where high-level descriptors are necessary to try to encapsulate the complexity of humans (Maojo, Martin, Crespo, & Billhardt, 2002). Therefore, although such traditional computing applications as routine iterative number crunching might be appropriate for the physical sciences, they cannot deal with these complexities, and DM techniques have been adopted and developed for this purpose (Shortliffe & Blois, 2001). Furthermore, the large and complex search spaces that are generated in health and medicine may mean that it is beyond the ability of clinicians to make decisions easily (Peña-Reyes & Sipper, 2000).

The collection, management, analysis, and interpretation of information are fundamental to clinical medicine and healthcare, notably in decision making relating to the categorization, treatment, and management of diseases (Shortliffe & Barnett, 2001). Capture and coding of this information for storage in databases and information systems can reduce some of its complexity and value. However, analyzing and interpreting the encoded data either routinely, or through DM as part of knowledge discovery, can help produce insights into the high-level processes that would not otherwise be possible.

Traditional epidemiological approaches to investigating rates and causes of diseases at a population level (Friedman, 1994) have used descriptive statistics to measure disease and inferential statistics to test hypotheses by investigating the extent to which the variance of a given disease's occurrence can be explained by variables of interest (potential risk factors) relative to other unexplained, or random, variance (Giuliani & Benigni, 2000). Although such studies work well when there is a "single causative agent far exceeding all the others" (Giuliani & Benigni, 2000, p. 308), many diseases and conditions, particularly noninfectious diseases, may have multiple causative agents or many risk factors. In such cases, traditional epidemiological and statistical approaches struggle to discriminate among a range of putative risk factors or causative agents and random variance. In other words, the "signal-to-noise" ratio is too low to be able to elucidate causes effectively (Giuliani & Benigni, 2000). Although proponents have discussed the potential of DM to overcome these limitations, there remains much scepticism among medical statisticians concerning the real value offered by such methods (Schwarzer, Vach, & Schumacher, 2000). In

addition, the low signal-to-noise ratio common in health/medical data means that the potential advantages of flexible, nonlinear, DM tools compared with statistical techniques will not be realized. However, this drawback may be overcome as advances in our understanding of risk factors for disease and health outcomes improve diagnostic and prognostic models (Biganzoli, Boracchi, Mariani, & Marubini, 2002, 1998). The following section discusses traditional statistical methods and their limitations in this regard.

Statistical Methods

Traditional hypothetico-deductive methods of analyzing health/medical data use inferential statistics to test null hypotheses using parametric and non-parametric measures, such as chi-square tests, correlation, and regression (Altman, 1991; Bland 2000). However, these methods have limitations and although they provide a measure of *statistical* significance, do not necessarily indicate *clinical* importance (Last, Schenker, & Kandel, 1999).

Univariate and Multivariate Analyses

Although DM techniques offer little above and beyond univariate or bivariate statistical analyses, such as t-tests, chi-square-tests, and correlation, they can usefully augment multivariate analyses, for example, cluster analysis and regression, which may not deal well with complex interactions among variables. Linear regression estimates the level of association between one or more independent, or predictor, variables and a continuous dependent, or outcome, variable (Altman, 1991; Bland, 2000; Dusseldorp & Meulman, 2001). Simple and multinomial logistic regression permit binary and nominal outcome variables to be used as the dependent variable respectively, through a transformation of the dependent variable (Altman, 1991). Logistic regression is particularly useful in health/medical research because many events of interest can be represented as binary variables; for example, the presence or absence of disease, being alive or dead, or responding to treatment or not (Altman, 1991). Logistic regression is also useful in making predictions and may be used for assisting clinical decision making for diagnosis and prognosis. However, it fails to consider the time at which an event occurs

(Altman, 1991; Bland 2000); survival analyses have been developed for this purpose.

Survival Analysis

Survival analyses account for an event occurring over a period of time within a population or group of interest (Altman, 1991). The term "survival" suggests that the event of interest, particularly in health/medical research, could be death (or not) of the individual, but it could be any event. Parametric methods of analyzing survival by comparing the distribution of survival times of different groups of patients have proved inadequate to deal with the complex relationships between predictor variables and events of interests, due to their assumptions regarding failure time distributions and the effects of the covariates on these distributions (Biganzoli et al., Boracchi & Marubini, 2002). The development of a semi-parametric method has overcome these limitations, but allows the identification only of putative risk factors, through the development of appropriate regression models (e.g., logistic regression and Cox regression) to analyze the effects of variables on survival (Anand, Smith, Hamilton, Anand, Hughes, & Bartels, 1999). Logistic regression is based on whether the event has happened or not, but the Cox proportional hazards regression model (Cox, 1972), or Cox regression, is based on the time elapsed before an event happens and is perhaps the most widely used survival analysis. However, Cox regression has to deal with situations in which the event of interest simply does not happen within a given time period. This is particularly important in health and medicine because of the numerous cases where something changes so that the event of interest cannot happen, for example, a respondent dies following a heart attack so a tumor cannot recur, or it never happens, as when an older person does not fall over, or it simply has not happened yet, as when respondents are still alive at the end of a study. In such circumstances, there is no date for the event of interest occurring, and a cut-off date has to be imposed at which the fact that the event has not occurred is recorded. This process is termed censorship, and because it marks the end of the study, the data are termed right-censored. Analyzing survival for diseases and conditions plays an important role in clinical medicine to enable HCPs to develop prognostic indices

following diagnosis for mortality, disease recurrence, outcomes of treatment, or the risk of adverse health events.

Limitations of Statistical Methods: Technical and Human Issues

Statistical methods are not able to deal satisfactorily with some problems associated with data generated through clinical practice and medical/health research. The nature of relationships among variables is complex and multivariate (Biganzoli et al., 2002), and interactions among predictor variables occur often; assessing these and their effects on the outcome variable is difficult (Dusseldorp & Meulman, 2001). Furthermore, the preponderance of nonlinear relationships among health/medical data and the nonadditive effects of multivariate relationships between predictor variables and outcome variables (Biganzoli et al., 2002) violate assumptions of linearity implicit in inferential statistical models and make them potentially suitable for DM.

Logistic and Cox regressions are important in generating population-based estimates of survival and for identifying putative risk factors. Logistic regression is also used to test the effectiveness of putative diagnostic and prognostic tools using a classification table that makes predictions on the basis of the values for the predictor variables for each case. These models can be evaluated by comparison with the actual diagnosis/outcome (Altman, 1991; Bland 2000). However, they are not used for making predictions concerning individual patients in a clinical setting (Anand, et al., 1999; Botacci, Drew, Hartley, Hadfield, Farouk, Lee, et al., 1997); HCPs tend to rely on their own knowledge, experience, and judgment, which have their limitations and are prone to human error.

Decision making by HCPs is based on knowledge gained through initial training, updated through continuing professional development and personal learning, and also by development of personal experience (Brause, 2001). Early in their careers, HCPs have limited experience, especially of relatively new or rare diseases/conditions. Humans are better at pattern recognition than at making decisions based on statistical probabilities (Brause, 2001; Lisboa, 2002; Walker, Cross, & Harrison, 1999). Although some of these limitations may be overcome, for example by consulting with more experienced colleagues, decision making may be

flawed by lack of appropriate experience or the ability to deal with complex data. DM may help overcome these problems by identifying patterns that were not previously apparent, or by learning from data to make decisions, predictions, prognoses, or diagnoses (Downs, Harrison, Kennedy, & Cross, 1996). However, to compare the performance of DM and statistical methods, appropriate means of evaluating the performance of diagnostic, prognostic, and other data analytic tools are required.

Evaluation of Methods

A criticism of DM tools developed in health/medicine has been the failure to compare their performance with equivalent statistical methods, a critical step before any data-mining tool can be used in routine clinical practice. For example, the correct diagnosis of diseases and the ability to make an accurate prognosis are vital for effective patient care. When developing and evaluating new methods of diagnosing conditions and making prognoses, it is necessary to compare the predicted diagnosis/prognosis with the true diagnosis or eventual outcome. This can be done using a classification table as shown in Table 7.1 (Altman & Bland, 1994a).

Table 7.1 illustrates that the true diagnosis showed that $n = a + c$ individuals were diagnosed as not having the condition, and of these the new method correctly diagnosed $n = a$ as not having the condition (*true negatives*). The true diagnosis showed that $n = b + d$ individuals/cases were diagnosed as having the condition, and of these the new method correctly diagnosed $n = d$ as having the condition (*true positives*). Overall the new method was correct for $n = a + d$ individuals. Conversely, the new method incorrectly diagnosed $n = b$ individuals as not having the condition (*false negatives*) and it incorrectly diagnosed $n = c$ individuals as having the condition (*false positives*) (Altman & Bland, 1994a; Lavrač, 1999b).

Sensitivity, equivalent to recall in information retrieval (van Rijsbergen, 1979), is the measure of how many of the individuals with the condition the test detects; in other words the proportion or percentage of true positives (Altman & Bland, 1994a). This is calculated by [sensitivity $= d/(b + d)$] and is expressed as a decimal or percentage.

Table 7.1 Time it takes to train and effort involved for different analytical methods in HCI (adapted from Olson & Moran, 1996, p. 281)

	True diagnosis		
Diagnosis by new method	Negative	Positive	Total
Negative	a	b	a+b
Positive	c	d	c+d
Total	a+c	b+d	a+b+c+d

Sensitivity is important in assessing how good the method is at identifying the individuals that have the condition. If the test were used in routine practice, then these people would potentially benefit from any intervention, such as medication or treatment, given to those whom the test identifies. Specificity, on the other hand, is a measure of how many of the individuals without the condition the test detects as not having the condition, that is, the rate of detecting true negatives; it is calculated by [specificity = $a/(a + c)$]. The positive predictive value (ppv) is equivalent to precision in information retrieval (van Rijsbergen, 1979), and is the proportion (percentage) of individuals that the method diagnoses as having the condition who actually have the condition (Altman & Bland, 1994b). It is calculated by [ppv = $d/(c + d)$]. Conversely, negative predictive value (npv) is the proportion of individuals whom the method diagnoses as not having the condition who actually do not have the condition, and is calculated by [negative predictive value = $a/(a + b)$] ($\times$ 100). The final estimate of accuracy is the receiver operating characteristic curve, which plots sensitivity against (1 − specificity) after calculating the sensitivity and specificity of every observed datum (Altman & Bland, 1994c). Although it enables the comparison of sensitivity and specificity in a single graph, giving one of the best estimates of the effectiveness of a procedure, additional calculations need to be incorporated to ensure that the prevalence of the condition in the population is taken into account (Bland & Altman, 1994c; Jefferson, Pendleton, Lucas, & Horan, 1995; MacNamee, Cunningham, Byrne, & Corrigan, 2002). Many DM

efforts are aimed at developing improved methods for making decisions, especially for diagnosis or prognosis; comparing the sensitivity, specificity, ppv, and npv achieved by statistical and DM methods is crucial in the development of tools and indicators. The relative significance of these measures of effectiveness within a particular clinical or health context has an important impact on the development of tools; this topic is discussed in the section on data mining and statistical methods.

Data Mining Tools for Health and Medicine

Data mining tools generally use either supervised or unsupervised learning for classification, making predictions, and other DM activities (Peña-Reyes & Sipper, 2000). A DM tool using supervised learning is trained to recognize different classes of data by exposing it to examples for which it has target answers (a training data set), and then testing it on a new data set, which it classifies (test data set). Unsupervised learning, on the other hand, requires no initial information regarding the correct classification of the data with which it is presented.

Recent reviews by Lavrač (1999a, 1999b) have discussed methods of machine learning for DM in health/medicine. Machine-learning methods include three main types of DM tool: inductive symbolic rule learning, statistical or pattern recognition methods, and artificial neural networks (Lavrač, 1999a). These techniques seek to improve medical diagnosis/prognosis by analyzing test data from previous patients, and from this learning process to predict the diagnosis and/or prognosis for a test set of patients. Lavrač (1999b) categorized DM methods into symbolic methods (e.g., rule induction methods, decision trees, and logic programs) and sub-symbolic methods (e.g., instance-based learning methods such as nearest neighbor algorithms, artificial neural networks, evolutionary methods, Bayesian classifiers, and combined approaches). A key distinction between symbolic and non-symbolic methods is the relative transparency (or "white box") of decision making using symbolic methods compared with the "black box" approaches of non-symbolic methods (Liebowitz, 2001b). This section describes symbolic and sub-symbolic methods of DM.

Inductive Learning of Symbolic Rules

Inductive learning of symbolic rules via rule induction algorithms, decision tree algorithms, and logic programs creates symbolic "if-then" rules from the training set that are used to generalize, and then are applied to classifying the test set of patients (Lavrač, 1999a). The symbolic rules are of the form

IF *Condition(s)* THEN *Conclusion*
or,
Condition(s) → *Conclusion*

in which the *Condition(s)* part includes one or more tests for values of the variables (labeled attributes), A_i, that are being included in which attribute tests, such as $A_i = value$ for discrete (categorical) variables and $A_i < value$ and/or $A_i < value$ for continuous variable. The *Conclusion* part assigns a value to a class of predictions, C_i (Lavrač, 1999b). Although rules derived through this process imply an association between the condition and the conclusion, Richards et al. (2001, p. 216) point out that "there is no implication of cause and effect" between them.

Rule-based approaches have been used in health/medicine for the diagnosis of rheumatic diseases, prognosis following cardiac tests (cited in Lavrač, 1999a), the prediction of early mortality in relation to first hospital visits (Richards et al., 2001), and in analyzing meningitis data (Zhong & Dong, 2002).

Decision Trees

Decision trees, also called tree-based methods, are based on recursive partitioning, which has been used for solving regression and classification problems in health/medical research (Dusseldorp & Meulman, 2001; Kuo et al., 2001). Regression trees model continuous variables to predict specific values for a variable of interest, whereas classification trees are used to model categorical variables in order to predict the group to which an individual or case belongs (Dusseldorp & Meulman, 2001; Kuo et al., 2001). The decision tree model can be used for descriptive purposes as well as for making predictions (Ennis, Hinton, Naylor, Revow, & Tibshirani, 1998; Kuo et al., 2001). The model is presented in the shape

of a tree composed of branches and leaves with decision rules on how the tree was constructed. Kuo et al. (2001) used a decision tree model to code breast cancer tumors as malignant or benign; they showed that the overall accuracy of the decision tree model was better than that of the physician, using measures of sensitivity, specificity, ppv, and npv. Recursive partitioning has been used for identifying interactions among variables by Carmelli, Halpern, Swan, Dame, McElroy, Gelb, et al. (1991), who compared recursive partitioning with Cox regression for examining the relationship between baseline biological and behavioral characteristics and mortality due to coronary heart disease and cancer over 27 years. Although both Cox regression and recursive partitioning were useful in determining risk factors, recursive partitioning enabled the identification of subgroups of individuals with particular characteristics and survival features (Carmelli et al., 1991).

Artificial Neural Networks

Artificial neural networks (ANNs) have emerged relatively recently as a useful and effective means of tackling a range of DM problems, including pattern recognition, prediction of outcomes, classification, and partitioning of multivariate data (Bath & Philp, 1998; Haykin, 1999). They have been applied in a variety of domains (Benoît 2002; Dayhoff, 1990; Trybula 1999), including health and medicine (Baxt, 1995; Brause, 2001; Cross et al., 1995; Dybowski & Gant, 1995). ANNs are so called because they have structures and processes that are modeled on the architecture and learning processes found in biological nervous systems. ANNs have the potential to extract information that is complementary, rather than an alternative, to that obtained using statistical methods; they are closely linked to regression (Cross et al., 1995; Sarle, 2002). For example, feed-forward neural nets can be regarded as a form of nonlinear regression, and Kohonen nets are a form of cluster analysis.

ANNs differ from statistical methods in being adaptive; that is, the data are presented to the ANN iteratively as the network "learns" and then revises the predictions or classifications it has made. During these iterations the network is trained to "recognize" patterns in the data; as a result of the training, the ANN can make predictions or classifications (Lipmann, 1987).

ANNs use supervised and unsupervised learning to mine data. ANNs employing unsupervised learning, such as Kohonen self-organizing maps, are able to analyze multi-dimensional data sets to discover natural patterns, or clusters and sub-clusters, that exist within the data (Kohonen, 1995; Lipmann, 1987). ANNs using this technique are able to identify their own classification schemes based upon the structure of the data provided. Unsupervised pattern recognition is similar to traditional methods of cluster analysis and is based on measures of similarity. ANNs using supervised learning, such as multi-layer perceptrons and radial basis function networks, learn from a training data set and then use a test data set to make predictions or classifications based on this learning. Supervised learning is more commonly used in modeling data derived from health/medicine (Lavrač, 1999b). Feed-forward networks, in which information is fed from the input layer through to the output layer, can become trapped in local minima and fail to reach an optimal solution (Cross et al., 1995). Back propagation can help to overcome this problem by comparing the output from the network with the true results, and then feeding this back through the network to refine the parameters of the net.

Artificial neural networks have been used in numerous clinical applications, including diagnosis, risk assessment, analyzing medical images and wave forms, and treatment selection and predicting outcomes; pharmacological applications include prediction of drug activities and responses to medication (cited in Baxt, 1991, and in Lavrač, 1999b). Artificial neural networks have been used for diagnosing a wide range of health/medical problems including myocardial infarction (Baxt, 1991; Baxt & Skora, 1996; Ennis et al., 1998), different forms of cancer (Pendharkar et al., 1999), detecting ischemia (Papaloukas, Fotiadias, Likas, & Michalis, 2002), appendicitis, back pain, dementia, psychiatric emergencies, pulmonary embolism, sexually transmitted diseases, skin diseases, and temporal ateritis (Baxt, 1995). Improved methods of diagnosis for myocardial infarction are necessary because, although the disease incidence is low, the consequences of a myocardial infarction not being diagnosed are potentially fatal (Baxt, 1995). Clinicians therefore tend to diagnose to avoid the risk of missing diagnosis of myocardial infarction. Although they may have a high sensitivity, the specificity of their diagnoses is relatively low and results in unnecessary hospital

admissions. Baxt (1995) identified a number of conditions, including recovery from surgery, for which artificial neural networks had been used in prognosis; these include predicting outcomes following surgery in intensive care units and orthopedic rehabilitation units (Grigsby, Kooken, & Hershberger, 1994); recovery from prostate, breast, and ovarian cancer (Downs et al., 1996); cardiopulmonary resuscitation and liver transplantation (Doyle, Dvorchik, Mitchell, Marino, Ebert, McMichael, et al., 1994); and rehospitalization following stroke (Ottenbacher, Smith, Illig, Linn, Fiedler, & Granger, 2001). Neural networks have also been used extensively for analyzing survival data (Biganzoli et al., Boracchi, Mariani, & Marubini, 1998; Biganzoli et al., 2002; Cacciafesta, Campana, Piccirillo, Cicconetti, Trani, Leonetti-Luparini, et al., 2001; Cross et al., 1995; Downs et al., 1996) and for predicting outcomes for providing policy information in the management of hypertension (Chae, Ho, Cho, Lee, & Ji, 2001).

ANNs have a number of advantages over statistical techniques that make them particularly suitable for mining health/medical data. ANNs are non-parametric and do not make assumptions about the underlying distributions of the data that statistical methods do (Lipmann, 1987). ANNs therefore may be more robust and perform better when data are not normally distributed or where there is a nonlinear relationship between predictor variables and an outcome variable. Artificial neural networks are able to analyze the higher-order relationships frequently present in health/medical data that traditional statistical tools are less capable of dealing with (Cross et al., 1995). However, the black-box nature of ANNs, in which data are fed in and results are obtained but with very little understanding of the reasons for the decision (Tu, 1996), is one of the fundamental limitations and explains why their use has been regarded with suspicion and mistrust within the medical community. Downs et al. (1996, p. 411) discussed the need to supplement the use of neural networks with the extraction of symbolic rules to "provide explanatory facilities for the network's 'reasoning'" and developed symbolic rules to try to explain the reasoning behind the decisions. Andrews, Diederich, & Tickle (1995) have developed techniques that permit this function.

A further problem with ANNs is that their performance on a test data set is often worse than that achieved through the training set (Brause,

2001) due to the network over training and adapting to any biases in the training set. Solutions to this difficulty include using a training data set that is representative of the test set, e.g., by randomly allocating training and test data from an original data set and checking that there are no significant differences between training and test data sets. However, the training and test data are not then independent of each other and subtle differences between training and test data sets may lead to a deterioration in performance, notably when the network is used on a truly independent data set, as in a clinical environment (Brause, 2001). In health/medicine, the problem that rare or unique cases may occur also can reduce the capacity to generalize. An additional problem is that ANNs may be over trained on the random variation present within populations or groups and be unable to generalize to other data sets. This problem can be overcome by halting the training at various points to ensure that the network does not train beyond the required level (Cross et al., 1995). Cross et al. (1995) commented that there was less rigorous development of artificial neural networks compared with that of conventional statistical tests and advised large-scale clinical trials to evaluate their use statistically before ANNs are accepted as a diagnostic tool. Additional limitations of DM tools are discussed in the section on challenges and solutions for DM.

Evolutionary DM Tools

Evolutionary DM tools encompass those computational techniques that are based on the principles and processes of evolution in nature, particularly those of reproduction, mutation, and selection (Goldberg, 1989; Peña-Reyes & Sipper, 2000). Evolutionary tools are methods for searching through the high-dimensional space of possible solutions to a given problem in order to find an optimal solution. They are particularly suitable for DM in health/medicine, given the preponderance of variables and multivariate relationships discussed previously. In this section, the concepts of evolution and how they are applied in these methods are discussed before genetic algorithms (GAs), genetic programming, and combined methods are presented.

Evolution is the theory of how living organisms developed over million of years from more primitive life forms. The manifestation of each individual (i.e., its phenotype) within a population is determined ultimately

by its genetic makeup or genome (genotype), which is encoded on chromosomes via genes. This genetic information is unique to each individual and reproduction, the process by which new individuals are created, involves the development of a new genome for that individual. Sexual reproduction involves the development of an entirely new genotype by recombination of the genetic material of the parents. This is supplemented by mutation, in which small random changes arise in the genetic material. The offspring from sexual reproduction undergo selection in which the Darwinian "survival of the fittest" occurs, so that those individuals that are best suited to the environment survive long enough to reproduce and pass their genetic material to the following generation. Over many generations, success in this process will permit the adaptation of the species to ensure its survival within the environment.

In evolutionary computing the environment represents the problem situation of interest, and the individuals within the population in this environment represent possible solutions to this problem (Goldberg, 1989). The algorithms for the various types of evolutionary computing tools are based on a common procedure in which the initial population is generated randomly or by using heuristics (Peña-Reyes & Sipper, 2000). The features or attributes of each individual are encoded via genes on a chromosome; associated with each chromosome is a fitness function, which measures its suitability to the environment or problem situation. The population undergoes a series of generations in which individuals (chromosomes) within the population undergo sexual reproduction to create new individuals (chromosomes) with new genotypes containing genetic material from the parents' crossover to create new genotypes, which are also subject to mutation. The offspring from this process, each with a fitness function associated with its genotype, then join the population. The fitness of each individual is determined by decoding and evaluating the genotype according to predefined criteria dependent on the problem being addressed. The strength of this fitness function will determine whether the individual survives to reproduce and pass on its genetic material to the next generation. Individuals (chromosomes) having the highest fitness functions will form a mating pool for the next generation, and the individuals (chromosomes) having lower fitness functions will be lost from the population. This selection process ensures

that the fittest individuals pass their genes to the next generation. The crossover ensures that new combinations of genetic material are introduced and "move towards promising new areas of the search space" (Peña-Reyes & Sipper, 2000, p. 23). Mutation prevents the process from converging in local optima that do not represent globally optimal solutions, and the new individuals then enter the environment and the next generation commences. Thus, similar to natural evolution, over a number of generations the population should adapt to the environment and a good approximation to an optimal solution to the problem should emerge. The process is terminated after a specified number of generations or when a predefined level of fitness is achieved.

An advantage of evolutionary computational tools over more traditional methods is that they combine coverage of all the available search space with the capacity to search the most promising areas (Peña-Reyes & Sipper, 2000). The results of the searches in these spaces can then be combined via crossover in reproduction and new areas of the search space can be investigated through mutations. This combination of targeted and stochastic search techniques means that evolutionary tools require less knowledge of the search space and make relatively few assumptions about it (Peña-Reyes & Sipper, 2000). Key considerations when using evolutionary DM include how to encode the features of possible solutions into genes and how to measure the fitness of the individuals and chromosomes. These issues depend on the specific problem and its particular features (Peña-Reyes & Sipper, 2000).

Genetic Algorithms

Much similarity is evident among the different types of evolutionary DM tools, and all are based on the principles and process of evolution. The most commonly used type of evolutionary tools are genetic algorithms (GAs), which represent the genome (genotype) of the individual (phenotype) using a fixed-length binary string (Peña-Reyes & Sipper, 2000). Although GAs can be used to generate solutions to almost any problem if the genotype can be represented in this way, care must be taken to ensure that no two genotypes encode the same phenotype (redundancy) in order to achieve a good solution (Peña-Reyes & Sipper, 2000). Using GAs, the number of individuals (population) is kept constant. During each generation these are decoded, their fitness is evaluated, and the fittest are

selected for reproduction. As mentioned earlier, GAs are particularly useful for DM in medicine because of their ability to search high-dimensional spaces to find an optimal solution to a problem.

GAs have been used for analyzing sleep patterns (Baumgart-Schmitt, Herrmann, & Eilers, 1998), diagnosis of female urinary incontinence and breast cancer (cited in Peña-Reyes & Sipper, 2000), development of prognostic systems for colorectal cancer (Anand et al., 1999), selection of features for recognizing skin tumors (Handels, Rob, Kruesch, Wolff, & Pöppl, 1999), prediction of depression after mania (Jefferson, Pendleton, Lucas, Lucas, & Horan, 1998a), predicting outcomes after surgery, predicting survival after lung cancer (Jefferson, Pendleton, Mohamed, Kirkman, Little, Lucas et al., 1998b), improving response to warfarin (Naranyan & Lucas, 1993), survival after skin cancer, and estimation of tumor stage and lymph node status in patients with colorectal adenocarcinoma (cited in Peña-Reyes & Sipper, 2000).

Genetic Programming

Work by Koza (1990a, 1990b) developed and extended the idea of evolutionary computational tools by using genetic programming. Although the basic evolutionary principles of GAs and genetic programming are similar, the features by which these tools carry out their tasks are fundamentally different (Peña-Reyes & Sipper, 2000). Genetic programming encodes possible solutions to problems as computer programs rather than as binary strings; to achieve this outcome, they use parse trees and functional programming languages, unlike GAs, which use line code and procedural languages. Genetic programming allows both asexual reproduction, in which the individuals with the highest fitness survive intact to the succeeding generation, as well as sexual reproduction, in which randomly selected points in the parse trees are selected and the sub-trees beneath these points are exchanged between the parents (Peña-Reyes & Sipper, 2000). Genetic programming tools have been less widely adopted in health/medical research than GAs, but have been used to identify causal relationships among children with limb fractures and on spinal deformation (Ngan, Wong, Lam, Leung, & Cheng, 1999), to classify brain tumors into meningioma and non-meningioma classes (Gray, Maxwell, Martinez-Perez, Arus, & Cerdan, 1998), learning rules

from a fractures database (Wong, Leung, & Cheng, 2000), and for the diagnosis of chest pain (Bojarczuk, Lopes, Freitas, 2000).

Other Methods of Evolutionary Computation

Evolutionary strategies and evolutionary programming have had little use in mining health/medical data (Peña-Reyes & Sipper, 2000). Their use has been restricted to analyzing sleep patterns (Baumgart-Schmitt et al., 1998), detecting breast cancer using histologic data (Fogel, Wasson, & Boughton, 1995) and radiographic features (Fogel, Wasson, Boughton, & Porto, 1997), and optimizing electrical parameters for therapeutic stimulation of the carotid sinus nerves (Peters, Koralewski, & Zerbst, 1989).

Combined Approaches

Evolutionary computing techniques have been used in combination with other tools for mining health/medical data. GAs have been combined with statistical and non-statistical methods to optimize the variables for inclusion in models. GAs have been combined with neural networks for detecting and diagnosing breast cancer (Abbass, 2002; Fogel et al., 1995), predicting response to warfarin (Naranyan & Lucas, 1993), outcomes following surgery (Jefferson, Pendleton, Lucas, & Horan, 1997), hemorrhagic blood loss (Jefferson et al., 1998b), depression following mania (Jefferson et al., 1998a) and for predicting falls and identifying risk factors associated with falls in older people (Bath et al., 2000). Fogel et al. (1995) used evolutionary artificial neural networks for analyzing histological data to detect and diagnose breast cancer. Fogel et al. (1997) used evolutionary programming to train artificial neural networks to detect breast cancer using data from radiographic features and patient age. As mentioned earlier, artificial neural networks can become stuck in local optima, and although increasing the number of nodes and weights associated with them can help overcome this problem, it becomes computationally intensive. Combining GAs with artificial neural networks can help the network overcome local optima and improve the topology of the neural network (Fogel et al., 1997). GAs have been used in combination with Bayesian networks to predict survival following malignant skin melanoma (Sierra & Larrañaga, 1998). Ngan et al. (1999) also used

genetic programming combined with Bayesian networks to identify rules for limb fracture patterns and for classifying scoliosis. Holmes, Durbin, and Winston (2000) combined a genetic algorithm with a rule-based system for epidemiologic surveillance. Peña-Reyes and Sipper (1999) combined GAs with a fuzzy system for diagnosing breast cancer. Although these studies represent attempts to combine evolutionary computing techniques with DM tools, little work has been conducted combining evolutionary computing methods with statistical methods to optimize the variables used in predictive models (Jefferson, 2001), indicating the potential for further work in this area.

Application of DM Tools in Diagnosis and Prognosis

Data mining tools have been used for a range of tasks, but particularly for diagnosis and prognosis of diseases and, in this section, their application in the diagnosis and prognosis of breast cancer is discussed.

Breast cancer has attracted considerable interest from data miners, particularly in relation to diagnosis. Reasons for this include its high incidence and high mortality rates in the developed world relative to other diseases and cancers (Alberg, Singh, May, & Helzlsouer, 2000) and, as Abbass (2002, p. 265) suggests, because of the very high "economic and social values" associated with it. An additional factor is the importance of early diagnosis, which has contributed to a decline in mortality in many countries and encouraged investigation of data mining to improve diagnosis. Problems with the traditional assessment of mammographic data have included inconsistencies in interpretation, resulting in poor intra- and inter-observer agreement (reliability) (Abbass, 2002; Fogel et al., 1997). Proposed reasons for this include the poor image quality of mammographic images and human fatigue and error; this has led to the development of pattern recognition techniques to supplement radiological diagnosis (Fogel et al., 1997). The aim of such developments has been to reduce the rate of false negative diagnoses by improving sensitivity. However, given the cytotoxic side effects of chemotherapy and radiotherapy as well as the psychosocial consequences of breast surgery, it is also important to ensure that the number of false positive diagnoses is minimized and a high positive

predictive value is achieved. Additional potential benefits of developing and using automated techniques include lower costs for handling mammograms, freeing up the time of the radiologist, and improving overall efficiency and effectiveness (Fogel et al., 1997).

Wu, Giger, Doi, Vyborny, Schmidt, and Metz (1993) reported artificial neural networks that were better at analyzing mammographic data than radiologists for decision making in the diagnosis of breast cancer. However, these data had been extracted by radiologists, and the authors suggested that the real potential of neural networks was to assist the radiologists in recommending when further tests be undertaken. Setiono (1996, 2000) developed a neural network program that used pruning to extract rules and provide information on the basis for the network's decisions, thus overcoming the "black box" aspect of neural networks. Many of the cited studies used the same Wisconsin Breast Cancer data set to develop the models. Although this is useful for comparing the effectiveness of the various tools, differences may exist between such training sets and data gathered from clinical settings in which the DM might eventually be employed. Therefore, the test data may not be representative of the population to which they are being generalized, resulting in a deterioration in performance when DM tools are used in a clinical setting. This concern emphasizes the need to test DM tools on new sets of data in different settings, in addition to those in which they were developed (Lisboa, 2002).

Walker et al. (1999) described the use of the growing cell structure technique to differentiate between benign and malignant breast tumors. This technique, which was shown comparable to logistic regression, allows multidimensional data (predictor variables) to be viewed as two-dimensional color images. The particular value of this visualization is that it permits HCPs to perceive relationships between the predictor and outcome variables, as well as interactions among the predictor variables (Walker et al., 1999).

Prognosis is an important area for patient care, where the limitations of both parametric and non-parametric statistical methods have led to the development of techniques that combine traditional survival analysis methods with artificial neural networks (Anand et al., 1999; Cacciafesta et al., 2001; Liestol et al., 1994; Faraggi & Simon, 1995;

Xiang, Lapuertab, Ryutova, Buckleya, & Azena, 2000; Zupan, Demšar, Kattan, Beck, & Bratko, 1999). Although some studies have shown that data mining methods perform better than statistical models for analyzing survival (Anand et al., 1999; Zupan et al., 1999), Anand et al. (1999) found that none of the three DM tools was able to handle the censored data as well as Cox regression.

The validity of prognostic models should be tested on a sample that is independent of the training sample with respect to time, place, and patients (Wyatt & Altman, 1995). However, DM techniques are often developed, trained, and tested on sets that are drawn from the same sample of patients and are therefore not truly independent of each other (Richards et al., 2001). These models cannot be regarded as having been independently validated, but require further testing on an independent data set. Wyatt and Altman (1995) contend that all clinically relevant data should be included in any prognostic model that is developed. However, defining the data that are clinically relevant for a particular condition is not easy, as prognostic models are often developed through secondary analyses of data collected for an entirely different purpose. It may not therefore have been possible to include all clinically relevant data in the model (Richards et al., 2001).

In many cases a wide variety of clinical variables influences the prognosis for a disease and an individual. This makes predictions for individual patients problematic, although it is particularly important for those who are terminally ill. Although it is known that approximately x percent of patients survive at least y years following treatment for a particular cancer, such population-based estimates are of limited value in supporting and treating individual patients who may want to know "How long will I live?" Such predictions are especially problematic as the deviation from the mean varies greatly (Bottaci et al., 1997). Anand et al. (1999) highlighted the need for better tools for disease prognosis, especially in patients with potentially terminal diseases, in which palliation and maintaining quality of life may become the main objectives. Information on the likelihood of survival and life expectancy can greatly assist in improving the quality of life when linked to appropriate counseling and disease management (Anand et al., 1999).

Challenges and Solutions for Data Mining in Health and Medicine: Technical and Human Issues

Moving from the description of DM tools and their application in health/medicine, this section examines the technical and human challenges to acceptance and adoption of DM (Lisboa, 2002) and suggestions of how these challenges may be met. Mistrust and suspicion of DM tools can be reduced by acknowledging and presenting their limitations clearly, avoiding exaggeration of their potential. Several authors have suggested how the development of DM and decision support tools based on DM might gain wider acceptance (Kononenko, Bratko, & Kukar, 1998; Lisboa, 2002).

Data Quality

Some technical challenges are common to statistical and DM methods. These include appropriate design of studies that develop and test DM tools, the need to represent data in an appropriate format (Isken & Rajagopalan, 2002), and the importance of ensuring that the data are of a high quality (e.g., in relation to missing data and consistency of data collection and recording). The statistical aspects of underlying data and models must be considered (Biganzoli et al., 2002); and it is important that descriptive statistics of mined data are available, as well as data that are analyzed statistically. Although many studies in health and medicine have used descriptive and inferential statistics without the apparent need for data mining tools, these tools cannot be developed in isolation from traditional statistical methods.

Lisboa (2002) discussed the need to clarify a study's purpose and to specify in advance expected benefits. The data mining tools in use are not necessarily the most advanced available, or it may be that what is preferred is not the best (Tu, 1996). The performance of DM tools may be enhanced by using more advanced types of GAs or artificial neural networks (Anand et al., 1999).

Data may be collected for a purpose other than that for which they are being analyzed and therefore not be clinically relevant for the diagnosis or prognosis for which they are being used (Richards et al., 2001, Wyatt

& Altman, 1995). Missing data, a particular a problem in medical data-bases, often arise through incomplete data being recorded or human error in recording/transcription (Brause, 2001; Richards et al., 2001). Problems with missing data can be improved by removing variables and/or cases that have a high proportion of missing values, although this approach may introduce bias because cases with large amounts of missing data may not be representative of the sample or may be associated with the outcome of interest. Replacing missing data with statistical descriptors, such as the mean value for a variable, is generally acceptable if done with care, but may introduce bias into the data (Altman, 1991).

Validity of Data Mining Methods

It is important to ensure that other biases are not allowed to influence the results when developing and testing DM tools. The correct classification must be concealed from domain experts until studies are completed so that the DM methods can be credited for the associations that are reported (Richards et al., 2001). However, the main objective of such studies should be to develop models that are clinically useful and of potential benefit to patients so that once models and tools have been validated, combining the domain knowledge of clinical experts with sophisticated analytic techniques may help to further improve performance. Richards et al. (2001) and Wyatt and Altman (1995) have stressed the need for training, validating, and testing of DM tools to be carried out on independent data and systems before implementation in real settings. Good practice should be followed in designing models, particularly to ensure that over-fitting is controlled (Lisboa, 2002), and that appropriate methods are available for variable selection (Tu, 1996). Bias can also arise from the minority class problem (MacNamee et al., 2002), in which the majority of cases in a data set belong to one class and the other class is significantly under-represented, resulting in a model being very good at identifying the former class but relatively poor at identifying the latter.

Usability of Data Mining Tools

DM diagnostic/prognostic tools can also increase the complexity of decision making for HCPs (Kononenko et al., 1998). Thus, tools should

be simple to use with user-friendly interfaces. Knowing *how* a model improves accuracy in decision making is as important as whether it does. HCPs must understand how any model works to be able to take responsibility for the results it produces (Lisboa, 2002). This means understanding not only basic mathematical principles underlying the models (Koh & Leong, 2001), but also how the models reached particular decisions—the inside of the "black box" discussed previously. Although the accuracy/performance of DM tools may be greater than statistical analysis, the lack of information about how they arrive at a decision may not be clear because of the "black box" and because of the complexity of the architecture (Setiono, 1996). Even though considerable progress has been made in developing sub-symbolic DM tools that are able to extract rules to explain how they reached their decisions (Andrews et al., 1995), these have not yet been widely adopted for use in health/medicine.

Lisboa (2002) commented on the increase in DM methods that allow visualization of the data and their potential to assist in a decision-making process. The Growing Cell Structure technique demonstrates the value of visualization (Walker et al., 1999). Humans are better at analyzing and interpreting data that are presented visually rather than numerically (Lisboa, 2002; Walker, 1999), consequently, DM models that present a visual image of how a decision was made may gain greater acceptance among HCPs. Involving HCPs in the design of user-friendly interfaces to DM systems will also help overcome resistance to their use.

Several authors have identified the need to establish an appropriate evidentiary base for the use of DM tools in medical/health practice, especially in respect of tools for diagnosis and prognosis (Cross et al., 1995; Johnston et al., 1994). Lisboa (2002) and Cross et al. (1995) discussed the need to compare the performance of DM tools with conventional methods before the utility of such techniques could be evaluated fully. Johnston et al. (1994) identified the need to evaluate computer-based decision-support systems not only in relation to reliability, acceptability, and accuracy, but also with respect to improving the clinical behavior and performance of HCPs, and ultimately patient well-being and treatment outcomes. The accepted gold standard for evaluating healthcare interventions, the randomized controlled trial (RCT), may not always be practical or feasible for evaluating computer-based decision-support systems developed using DM. Nevertheless, investment in evaluating the

effectiveness and efficiency of such systems is necessary to maximize the potential benefits and minimize the potential for harm or waste that may arise (Johnston et al., 1994). Lisboa (2002) highlighted the need to evaluate DM tools through multi-center RCTs and to establish an appropriate evidentiary base for the use of DM tools (Anand et al., 1999; Brause, 2001; Lisboa, 2002).

Downs et al. (1996) highlighted the tension between the need for symbolic rules discovered during the DM process to be acceptable to domain experts and the need to demonstrate that the method provides new knowledge or understanding in the domain area. Having a means of demonstrating how a system arrives at its decision is critical for both symbolic and sub-symbolic methods. Certainly, the ability of neural networks to detect previously unknown lower-order relationships, which can then be tested using statistical models, can help them gain acceptance among medical/health professionals. This can increase the perceived trustworthiness of DM tools when interactions among the data are discovered that cannot be verified using statistical methods (Lisboa, 2002). An additional problem is that DM tools may identify patterns not accepted or not in accordance with current knowledge (Richards et al., 2001; Wyatt & Altman, 1995), which may limit their acceptance among HCPs.

Data Mining and Statistical Methods

DM is useful for generating hypotheses for further testing as in identifying associations or relationships between variables/data that are then tested using conventional statistical techniques (Richards et al., 2001). There is a need not only to show how DM methods can complement statistical techniques in analyzing health/medical data, but also to emphasize the added value that DM methods can bring to the knowledge discovery process. Understanding the similarities and differences between DM and statistical methods highlights the contribution that each makes in improving our understanding of the processes underlying health and illness. For instance, although both Cox regression and tree-structured survival analysis allow the identification of risk factors for adverse health events, Cox regression can provide an estimate of the strength of these risk factors and tree-structured analysis helps to identify high-risk groups with particular features in common (Carmelli et

al., 1991). Comparing the performance of different DM and statistical approaches also allows different information to be extracted from the data. For example, Lee, Liao, and Embrechts (2000) compared a variety of techniques including correlation analysis, discriminant analysis, data visualization, and artificial neural networks to analyze data from a heart disease database. They were able to identify people at risk of heart disease, detect risk factors for heart disease, and establish multivariate relationships among the predictor variables. This provides further evidence of the need to use statistical methods alongside non-statistical tools.

It is particularly important to understand the objectives of studies in trying to improve prognostic and diagnostic performance (Lisboa, 2002). The ultimate aim of 100 percent accuracy is rarely achieved, and the relative importance of sensitivity, specificity, and positive and negative predictive values within the context of clinical care must be considered. For certain diseases, high sensitivity is critical because of the serious, and potentially fatal, consequences for an individual of not diagnosing an actual case (false negatives), or to ensure that a correct diagnosis is obtained as soon as possible so that treatment can commence at an early stage in the disease (Fogel et al., 1997; Fogel et al., 1995). For other diseases, however, the imperative may be to ensure that the specificity is very high in order to minimize the number of people who are wrongly diagnosed as having the disease and receiving unnecessary treatments (Downs et al., 1996). Diagnosing all positive cases may be important in improving survival rates and reducing co-morbidities, but reducing false positives may also be important so that patients are not given medications with toxic side effects (and high costs) unnecessarily, and so HCPs can maximize time with true cases (Abbass, 2002).

User Acceptance of Data Mining

DM is an important part of the knowledge management process within healthcare organizations (Bellazi & Zupan, 2001; Liebowitz, 2001a). Data mining relies on the explicit knowledge present in the available health/medical literature that is used by clinical researchers, clinicians, methodologists, and information specialists to help identify appropriate research questions. The tacit knowledge of clinicians, HCPs,

and managers is also required to develop and understand the data and to evaluate/assess and interpret the results. The explicit knowledge of clinicians and HCPs may also be embodied in specific DM methods (e.g., Bayesian networks and fuzzy systems) for analyzing data (Bellazi & Zupan, 2001). This highlights the importance of multidisciplinary collaboration between health/medical professionals and information analysts in using DM (Kuo et al., 2001) to overcome the suspicions of the former and any over-confidence among the latter (Biganzoli et al., 2002; Kuo et al., 2001). In the same way that healthcare professionals build trust in each other through sharing information in decision making, they need to develop trust in their decision-making tools (Abbass, 2002).

Despite all the research and success of DM tools, no tool or automated process arising from DM has been adopted for use on a routine basis (Abbass, 2002). HCPs may mistrust technology so the complementary nature of DM tools must be emphasized: DM as adjuncts to decision making by HCPs rather than replacements (Abbass, 2002). For HCPs to trust DM tools, they need to understand not only their performance, but also their limitations (Cross et al., 1995). Clinical judgment and experience must be combined for careful interpretation of the results (Botacci et al., 1997), and it should be made clear that data mining tools are "just another source of possibly useful information" (Kononenko et al., 1998, p. 403) that healthcare professionals may use in decision making and providing care for patients.

DM tools need to be evaluated from a patient's perspective (Sullivan & Mitchell, 1995), and should demonstrate an overall improvement in patient outcomes if they are to achieve wider acceptance (Lisboa, 2002). Although studies have demonstrated the effectiveness of DM techniques in terms of diagnostic and prognostic accuracy, little research has shown an improvement in patient health and well-being.

A final, but by no means the least, important concern in health and medicine is ethics. Ethical considerations are particularly important in health/medicine because patients are often in a vulnerable position when receiving care or treatment. It is important, therefore, that DM tools are developed ethically, with the ultimate well-being of patients and the public in mind.

Conclusions

Selected DM and statistical techniques used in health/medicine have been examined and the factors affecting the development of DM in this domain have been discussed. A number of technical and human issues have been identified, including the importance of ensuring that data are of high quality, validating results obtained through DM, evaluating the performance of DM tools, involving the collaboration and trust of HCPs in the development process, and demonstrating the benefits of using DM.

Although our understanding of the complex processes underlying health and illness is improving, the available data are becoming more numerous and complex, creating increasing demands for more effective ways to process these data and answer clinically relevant questions. Data mining can help overcome some of the problems of statistical methods in analyzing medical/health data, and can complement these methods for diagnosis, prognosis, decision making, and generating hypotheses so that the strengths of different techniques can be maximized and their weaknesses minimized. DM tools should be user-friendly and designed to be used by HCPs with the ultimate goal of improving patient health and well-being. The development of DM applications requires investment of time and resources (Koh & Leong, 2001), but perhaps most essential is recognizing that it is part of a process that involves the multidisciplinary and open-minded collaboration of HCPs and information professionals.

References

Abbass, H. A. (2002). An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial Intelligence in Medicine, 25*(3), 265–281.

Adriaans, P., & Zantige, D. (1996). *Data mining*. Harlow, U.K.: Addison-Wesley.

Alberg, A. J., Singh, S., May, J. W., & Helzlsouer, K. J. (2000). Epidemiology, prevention, and early detection of breast cancer. *Current Opinion in Oncology, 12*(6), 515–520.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman Hall/CRC.

Altman, D. G., & Bland, M. (1994a). Statistics notes: Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal, 308,* 1552.

Altman, D. G., & Bland, M. (1994b). Statistics notes: Diagnostic tests 2: Predictive values. *British Medical Journal, 309,* 102.

Altman, D. G., & Bland, M. (1994c). Statistics notes: Diagnostic tests 3: Receiver operating characteristic plots. *British Medical Journal, 309,* 188.

Anand, S. S., Smith, A. E., Hamilton. P. W., Anand, J. S., Hughes, J. G., & Bartels, P. H. (1999). An evaluation of intelligent prognostic systems for colorectal cancer. *Artificial Intelligence in Medicine, 15*(2), 193–214.

Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems, 8*(6), 373–389.

Bath, P. A., Morgan, K., Pendleton, N., Clague, J., Horan, M., & Lucas, S. (2000). A new approach to risk determination: Prediction of new falls among community-dwelling older people using a genetic algorithm neural network (GANN). *Journal of Gerontology (medical science). 55A,* M17–21.

Bath, P., & Philp, I., (1998). A hierarchical classification of dependency amongst older people using artificial neural networks. *Health Care in Later Life, 3*(1), 59–69.

Baumgart-Schmitt, R., Herrmann, W. M., & Eilers, R. (1998). On the use of neural network techniques to analyze sleep ECG data. *Neuropsychobiology, 37,* 49–58.

Baxt, W. G. (1991). Use of an artificial neural network for the diagnosis of myocardial infarction. *Annals of Internal Medicine, 115,* 843–848.

Baxt, W. G. (1995). Application of artificial neural networks to clinical medicine. *Lancet, 346,* 1135–1138.

Baxt, W. G., & Skora, J. (1996). Prospective validation of artificial neural networks trained to identify acute myocardial infarction. *Lancet, 280*(3), 229–231.

Bellazzi, R., & Zupan, B. (2001). Intelligent data analysis [Special issue]. *Methods of Information in Medicine, 5,* 362–364.

Bender, R., & Lange, S. (1999). Multiple test procedures other than Bonferroni's deserve wider use. *British Medical Journal, 318,* 600a–600.

Benoît, G. (2002). Data mining. *Annual Review of Information Science and Technology, 36,* 265–310.

Bertone, P., & Gerstein, M. (2001). Integrative data mining: The new direction in bioinformatics. *IEEE Engineering in Medicine & Biology Magazine, 20*(4), 33–40.

Biganzoli, E., Boracchi, P., Mariani, L., & Marubini, E. (1998). Feed forward neural networks for the analysis of censored survival data: A partial logistic regression approach. *Statistics in Medicine, 17,* 1169–1186.

Biganzoli, E., Boracchi, P., & Marubini, E. (2002). A general framework for neural network models on censored survival data, *Neural Networks, 15*(2), 209–218.

Bland, M. (2000). *An introduction to medical statistics* (3rd ed.). Oxford, U.K.: Oxford Medical Publications.

Bland, M. J., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *British Medical Journal, 310,* 170.

Bohanec, M., Zupan B., & Rajkovič, V. (2000). Applications of qualitative multi-attribute decision models in health care. *International Journal of Medical Informatics, 58–59,* 191–205.

Bojarczuk, C. C., Lopes, H. S., & Freitas, A. A. (2000). Genetic programming for knowledge discovery in chest-pain diagnosis. *IEEE Engineering in Medicine & Biology Magazine, 19*(4), 38–44.

Bottaci, L., Drew, P. J., Hartley, J. E., Hadfield M. B., Farouk, R., Lee, P. W. R., Macintyre, I. M. C., Duthie, G. S., & Monson, J. R. T. (1997). Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet, 350*(9076), 469–472.

Brause, R. W. (2001). Medical analysis and diagnosis by neural networks. *Lecture Notes in Computer Science 2199,* 1–13.

Cacciafesta, M., Campana, F., Piccirillo, G., Cicconetti, P., Trani, I., Leonetti-Luparini, R., Marigliani, V., & Verico, P. (2001). Neural network analysis in predicting 2-year survival in elderly people: A new mathematical-statistical approach. *Archives of Gerontology and Geriatrics, 32*(1), 35–44.

Carmelli, D., Halpern, J., Swan, G. E., Dame, A., McElroy, M., Gelb, A. B., & Rosenman, R. H. (1991). 27-year mortality in the western collaborative group study: Construction of risk groups by recursive partitioning. *Journal of Clinical Epidemiology, 44*(12), 1341–1351.

Chae, Y. M., Ho, S. H., Cho, K. W., Lee, D. H., & Ji, S. H. (2001). Data mining approach to policy analysis in a health insurance domain. *International Journal of Medical Informatics, 62,* 103–111.

Coiera, E. (1997). *Guide to medical informatics, the Internet and telemedicine.* London: Arnold.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society B*, 4, 232–236.

Cross, S. S., Harrison, R. F., & Kennedy, R. L. (1995). Introduction to neural networks. *Lancet, 346,* 1075–1079.

Daly, L. E., & Bourke, G. J. (2000). *Interpretation and uses of medical statistics* (5th ed.). Oxford, U.K.: Blackwell Science.

Dayhoff, J. E. (1990). *Neural network architectures: An introduction.* New York: Van Nostrand Reinhold.

Downs, J., Harrison R. F., Kennedy, R. L., & Cross, S. S. (1996). Application of the fuzzy ARTMAP neural network model to medical pattern classification tasks. *Artificial Intelligence in Medicine, 8*(4), 403–428.

Doyle, H. R., Dvorchik, I., Mitchell, S., Marino, I. R., Ebert, F. H., McMichael, J. & Fung, J. J. (1994). Predicting outcomes after liver transplantation: A connectionist approach. *Annals of Surgery, 219*(4), 408–415.

Dusseldorp, E., & Meulman J. J. (2001). Prediction in medicine by integrating regression trees into regression analysis with optimal scaling. *Methods of Information in Medicine, 40,* 403–409.

Dybowski, R., & Gant, V. (1995). Artificial neural networks in pathology and medical laboratories. *Lancet, 346,* 1203–1207.

Ennis, M., Hinton, G., Naylor, D., Revow, M., & Tibshirani, R. (1998). A comparison of statistical learning methods on the GUSTO database. *Statistics in Medicine, 17,* 2501–2508.

Faraggi, D., & Simon, R. (1995). A neural network model for survival data. *Statistics in Medicine, 14,* 73–82.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27–34.

Floyd, C. E., Lo, J. Y., Yun, A. J., Sullivan, D. C., & Kornguth, P. J. (1994). Prediction of breast cancer malignancy using an artificial neural network. *Cancer, 74*(11), 2944–2948.

Fogel, D. B., Wasson, E. C., & Boughton, E. M. (1995). Evolving neural networks for detecting breast cancer. *Cancer Letters, 96*(1), 49–53.

Fogel, D. B., Wasson, E. C., Boughton, E. M., & Porto, V. W. (1997). A step toward computer-assisted mammography using evolutionary programming and neural networks. *Cancer Letters, 119*(1), 93–97.

Friedman, G. D. (1994). *Primer of epidemiology* (4th ed.). New York: McGraw-Hill.

Giuliani, A., & Benigni, R. (2000). Principal components analysis for descriptive epidemiology. *Lecture Notes in Artificial Intelligence, 1933,* 308–313.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning.* New York: Addison-Wesley.

Gray, H. F., Maxwell, R. J., Martinez-Perez, I., Arus, C., & Cerdan, S. (1998). Genetic programming for classification and feature selection: Analysis of 1H nuclear magnetic resonance spectra from human brain tumour biopsies. *NMR in Biomedicine, 11*(4–5), 217–224.

Grigsby, J., Kooken, R., & Hershberger, J. (1994). Simulated neural networks to predict outcomes, costs, and length of stay among orthopedic rehabilitation patients. *Archives of Physical and Medical Rehabilitation, 75,* 1077–1081.

Handels, H., Rob, T., Kruesch, J., Wolff, H. H., & Pöppl, S. J. (1999). Feature selection for optimized skin tumor recognition using genetic algorithms. *Artificial Intelligence in Medicine, 16,* 283–289.

Haykin, S. S. (1999). *Neural networks : A comprehensive foundation* (2nd ed.). Upper Saddle River, NJ.: Prentice Hall International.

Holmes, J. H., Durbin, D. R., & Winston, F. K. (2000). The learning classifier system: An evolutionary computation approach to knowledge discovery in epidemiologic surveillance. *Artificial Intelligence in Medicine, 19,* 53–74.

Horn, W. (2001). AI in medicine on its way from knowledge-intensive systems to data-intensive systems. *Artificial Intelligence in Medicine, 23,* 5–12.

Isken, M. W., & Rajagopalan, B. (2002). Data mining to support simulation modelling of patient flow in hospitals. *Journal of Medical Systems, 26*(2), 179–197.

Jefferson, M. (2001). Outcome prediction in medicine with genetic algorithm neural networks. Unpublished doctoral dissertation, University of Manchester.

Jefferson, M. F., Pendleton, N., Lucas, S. B., & Horan, M. A. (1995). Neural networks. *Lancet, 346,* 1712.

Jefferson, M. F., Pendleton, N., Lucas, S. B., & Horan, M. A. (1997). Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma. *Cancer, 79*(7), 1338–1342.

Jefferson, M. F., Pendleton, N., Lucas, C. P., Lucas S. B., & Horan, M. A. (1998a). Evolution of artificial neural network architecture: Prediction of depression after mania. *Methods of Information in Medicine, 37,* 220–225.

Jefferson, M. F., Pendleton, N., Mohamed, S., Kirkman, E., Little, R. A., Lucas, S. B., & Horan, M. A. (1998b). Prediction of hemorrhagic blood loss with a genetic algorithm neural network. *Journal of Applied Physiology, 84,* 357–361.

Johnston, M. E., Langton, K. B., Hayes, R. B., & Mathieu, A. (1994). Effects of computer-based clinical decision support systems on clinician performance and patient outcome: A critical appraisal of research. *Annals of Internal Medicine, 120,* 135–142.

Jones, J. K. (2001). The role of data mining technology in the identification of signals of possible adverse drug reactions: Values and limitations. *Current Therapeutic Research, 62*(9), 664–673.

Koh, H. C., & Leong, S. K. (2001). Data mining applications in the context of case mix. *Annals of the Academy of Medicine, 30*(4), 41–49.

Kohonen, T. (1995). *Self-organizing maps.* Berlin: Springer Verlag.

Kononenko, I., Bratko, I., & Kukar, M. (1998). Application of machine learning in medical diagnosis. In R. S. Michalsko, I. Bratko & M. Kubat (Eds.). *Machine learning and data mining: Methods and applications* (pp. 389–408). New York: John Wiley.

Koza, J. R. (1990a). *Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems* (STAN-CS-90-1314). Stanford. CA: Stanford University Computer Science Department.

Koza, J. R. (1990b). Genetically breeding populations of computer programs to solve problems in artificial intelligence. *Proceedings of the Second International Conference on Tools for AI,* 819–827.

Kuo, W. J., Chang, R. F., Chen, D. R., & Lee, C. C. (2001). Data mining with decision trees for diagnosis of breast tumour in medical ultrasonic images. *Breast Cancer Research and Treatment, 66,* 51–57.

Last, M., Schenker, A., & Kandel, A. (1999). Applying fuzzy hypothesis testing to medical data. In N. Zhong, A. Skowron & S. Ohsuga (Eds.), *New directions in rough sets, data mining, and granular-soft computing* (pp. 221–229). Berlin: Springer.

Lavrač, N. (1999a). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine, 16,* 3–23.

Lavrač, N. (1999b). Machine learning for data mining in medicine. *Lecture Notes in Artificial Intelligence, 1620,* 47–62.

Lee, I. N., Liao, S. C., & Embrechts, M. (2000). Data mining techniques applied to medical information. *Medical Informatics, 25*(2), 81–102.

Liebowitz, J. (2001a). Knowledge management and its link to artificial intelligence. *Expert Systems with Applications, 20,* 1–6.

Liebowitz, J. (2001b). If you are a dog lover, build expert systems; if you are a cat lover, build neural networks. *Expert Systems with Applications, 21,* 63.

Liestol, K., Andersen, P. K., & Andersen, U. (1994). Survival analysis and neural nets. *Statistics in Medicine, 13,* 1189–1200.

Lin, F., Chou, S., Pan, S., & Chen, Y. (2001). Mining time dependency patterns in clinical pathways. *International Journal of Medical Informatics, 62,* 11–25.

Lipmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine, 4,* 4–22.

Lisboa, P. J. G. (2002). A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Networks, 15*(1), 11–39.

Luscombe, N. M., Greenbaum, D., Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine, 40*(4), 346–358.

MacNamee, B., Cunningham, P., Byrne, S., & Corrigan, O. I. (2002). The problem of bias in training data in regression problems in medical decision support. *Artificial Intelligence in Medicine, 24,* 51–70.

Maojo, V., Martin, F., Crespo, J., & Billhardt, H. (2002). Theory, abstraction and design in medical informatics. *Methods of Information in Medicine, 41,* 44–50.

Maojo, V., & Sanandrés, J. (2000). A survey of data mining techniques. *Lecture Notes in Artificial Intelligence, 1933,* 17–21.

McSherry, D. (1999). Dynamic and static approaches to clinical data mining. *Artificial Intelligence in Medicine, 16,* 97–115.

Michalski, R. S., Bratko, I., & Kubat, M. (1997). *Machine learning and data mining: Methods and applications.* New York: John Wiley.

Miller, P. L. (2000). Opportunities at the intersection of bioinformatics and health informatics: A case study. *Journal of the American Medical Informatics Association, 7*(5), 431–438.

Mills, J. L. (1993). Data torturing. *New England Journal of Medicine, 329,* 1196–1199.

Naranyan, M. N., & Lucas, S. B. (1993). A genetic algorithm to improve a neural network performance to predict a patient's response to Warfarin. *Methods of Information in Medicine, 32,* 55–58.

Ngan, P. S., Wong, M. L., Lam, W., Leung, K. S., & Cheng, J. C. Y. (1999). Medical data mining using evolutionary computation. *Artificial Intelligence in Medicine, 16*(1), 73–96.

Ottenbacher, K. J., Smith, P. M., Illig, S. B., Linn, T., Fiedler, R. C., & Granger, C. V. (2001). Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. *Journal of Clinical Epidemiology, 54,* 1159–1165.

Papaloukas, C., Fotiadis, D. I., Likas, A., & Michalis, L. K. (2002). An ischemia detection method based on neural networks. *Artificial Intelligence in Medicine, 24,* 167–178.

Peña-Reyes, C. A., & Sipper, M. (1999). A fuzzy-genetic approach to breast cancer diagnosis. *Artificial Intelligence in Medicine, 17,* 131–155.

Peña-Reyes, C.A., & Sipper, M. (2000). Evolutionary computation in medicine: An overview. *Artificial Intelligence in Medicine, 19,* 1–23.

Pendharkar, P. C., Rodger, J. A., Yaverbaum, G. J., Herman, N., & Benner, M. (1999). Association, statistical, mathematical and neural approaches for mining breast cancer patterns. *Expert Systems with Applications, 17,* 223–232.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal, 316,* 1236–1238.

Peters, T. K., Koralewski, H. E., & Zerbst, E. W. (1989). The evolution strategy: A search strategy used in the individual optimisation of electrical parameters for therapeutic carotid sinus nerve stimulation. *IEEE Transactions on Biomedical Engineering, 36*(7), 668–675.

Richards, G., Rayward-Smith, V. J., Sönksen, P. H., Carey, S., & Weng, C. (2001). Data mining for indicators of early mortality in a database of clinical records. *Artificial Intelligence in Medicine, 22,* 215–231.

Sarle, W. S. (2002). How are NNs related to statistical methods? Retrieved November 28, 2002, from http://www.faqs.org/faqs/ai-faq/neural-nets/part1/section-15.html

Schwarzer, G., Vach, W., & Schumacher, M. (2000). On the misuses of artificial neural network for prognostic and diagnostic classification in oncology. *Statistics in Medicine, 19,* 451–561.

Setiono, R. (1996). Extracting rules from pruned neural networks for breast cancer diagnosis. *Artificial Intelligence in Medicine, 8,* 37–51.

Setiono, R. (2000). Generating concise and accurate classification rules for breast cancer diagnosis. *Artificial Intelligence in Medicine, 18,* 205–219.

Shortliffe, E. H., & Barnett, G. O. (2001). Medical data: Their acquisition, storage and use. In E. H. Shortliffe & L. E. Perreault (Eds.), *Medical informatics computer applications in health care and biomedicine* (2nd ed.) (pp. 41–75). New York: Springer.

Shortliffe, E. H., & Blois, M. S. (2001). The computer meets biology and medicine: Emergence of a discipline. In E. H. Shortliffe & L. E. Perreault (Eds.), *Medical informatics computer applications in health care and biomedicine* (2nd ed.) (pp. 3–40). New York: Springer.

Sierra, B., & Larrañaga, P. (1998). Predicting survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms: An empirical comparison between different approaches. *Artificial Intelligence in Medicine, 14,* 215–230.

Smith, A., & Nelson, M. (1999). Data warehouses and clinical data warehouses. In M. J. Ball, J. V. Douglas, & D. E. Garets (Eds.), *Strategies and technologies for healthcare information* (pp. 17–31). New York: Springer.

Sullivan, F., & Mitchell, E. (1995). Has general practitioner computing made a difference to patient care? A systematic review of published reports. *British Medical Journal, 311,* 848–852.

Swanson, D. R. (1987). Two medical literatures that are logically but not bibliographically connected. *Journal of the American Society for Information Science, 38,* 228–233.

Swanson, D. R. & Smalheiser, N. R. (1999). Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. *Library Trends, 48,* 48–59.

Trybula, W. J. (1997). Data mining and knowledge discovery. *Annual Review of Information Science and Technology, 32,* 197–229.

Trybula, W. J. (1999). Text mining. *Annual Review of Information Science and Technology, 34,* 385–420.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology, 49*(11), 1225–1231.

van Rijsbergen, C. J. (1979). *Information retrieval* (2nd ed.). London: Butterworths. Retrieved November 28, 2002, from http://www.dcs.gla.ac.uk/Keith/Preface.html

Walker, A. J., Cross, S. S., & Harrison, R. F. (1999). Visualisation of biomedical datasets by use of growing cell structure networks: A novel diagnostic classification technique. *Lancet, 354,* 1518–1521.

Wong, M. L., Leung, K. S., & Cheng, J. C. Y. (2000). Discovering knowledge from noisy databases using genetic programming. *Journal of the American Society for Information Science, 51,* 870–881.

Wu, Y., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., & Metz, C. E. (1993). Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. *Radiology, 187*(1), 81–87.

Wyatt, J. C., & Altman, D. G. (1995). Commentary: Prognostic models; clinically useful or quickly forgotten? *British Medical Journal, 311,* 1539–1541.

Xiang, A., Lapuertab, P., Ryutova, A., Buckleya, J., & Azena, S. (2000). Comparison of the performance of neural network methods and Cox regression for censored survival data. *Computational Statistics & Data Analysis, 34*(2), 243–257.

Zhong, N., & Dong, J. (2002). Mining interesting rules in meningitis data by cooperatively using GDT-RS and RSBR. *Lecture Notes in Artificial Intelligence, 2336,* 405–416.

Zupan, B., Demšar, J., Kattan, M. W., Beck, J. R., & Bratko, I. (1999). Machine learning for survival analysis: A case study on recurrence of prostate cancer. *Lecture Notes in Artificial Intelligence, 1620,* 346–355.