Data mining in health and medical information

Peter A. Bath

Introduction

Data mining (DM) is part of a process by which information or knowledge can be extracted from data or databases and used to inform decision-making in a variety of contexts (Michalski et al., 1997; Benoit, 2002). The process of DM includes a range of tools and methods for extracting this information and one of the main driving forces for the development of DM tools has been their use in the commercial sector for knowledge extraction and discovery in commercial/ business applications (Adriaans and Zantige, 1996; Benoit, 2002). However, DM has been developed and applied in a number of different areas and the purpose of this review is to describe and discuss one such application area, the use of DM for analysing health and medical information.

Several reviews of DM have appeared in ARIST over the last few years (Trybula, 1997; Trybula, 1999; Benoit, 2001), and this review complements these by exploring the use of DM in the specific domain of health and medical practice and research and examines the features of this domain and its particular suitability for DM in the problems specific to this area. A number of general reviews of the application of DM tools in the health and medicine have recently been published (e.g., Maojo and Sanandrés, 2000; Lavrač N, 1999a; Horn, 2001; Peňa-Reyes and Sipper, 2000; McSherry, 1999) as well as more specific reviews of the use and effectiveness of particular tools and methods in this domain, mainly artificial neural networks (Baxt, 1995; Cross et al., 1995; Dybowski and Gant, 1995; Lisboa, 2002; Liestol et al., 1994; Tu, 1996) but also machine learning methods (Lavrač, 1999b) and computer-based clinical decision support systems (Johnston et al., 1994). In addition this review considers the importance of statistics in the DM process, and numerous general

statistics texts are available and accessible to non-statisticians with an interest in analysing health and medical information (Altman, 1991; Bland, 2000; Daly and Bourke, 2000). This review provides an overview of the range of DM tools that have been applied in health and medicine and examines the issues that are affecting their development and uptake as part of routine clinical practice in this domain.

Scope and limitations of the review

Recent ARIST reviews of DM have discussed mining of structured data (Trybula, 1997), textual data (Trybula, 1999) and DM as part of the knowledge discovery process (Benoit, 2001) and these have considered applications of DM in different contexts and domains. This review will only describe those tools that have been used for mining data in health and medicine, and will discuss the issues that make these data suitable for DM and the factors affecting their use. The review will not consider developments in DM in other application areas, nor will it compare the use of DM techniques within health and medicine in relation to these areas. Although what is described for health and medical research may be equally true of other areas in which DM has been or might be applied, and there may be comparisons and contrasts in the application of DM in the medicine/health domain and other areas, discussion of these is beyond the scope of this review. The use of DM techniques in areas closely related to medicine and health, e.g., analysing genomic databases in bioinformatics etc., are outside the scope of the review and will not be discussed here. Reviews discussing the role of DM in this field have recently been published (Bertone and Gerstein, 2001; Luscombe et al., 2001; Miller 2000). The review will focus on DM tools for analysing numeric quantitative data in health and medicine and will not consider DM tools such as HINT and DEX developed to process qualitative data (Bohanec et al., 2000) or the mining of text data in this domain (Trybula, 1999; Swanson, 1987; Swanson and Smalheiser, 1999). Within the health and

medical domain the review will consider the application of DM tools in medical and health care practice and research and will not describe application of DM tools in laboratory environments; a useful, if now somewhat dated, review by Dybowski and Gant (1995) discussed the use of artificial neural networks in pathology and medical laboratories and Jones (2001) recently reviewed the use of data mining for identifying adverse events in clinical trials and adverse events databases. In undertaking the review a considerable number of research articles reporting applications of DM in the computing, health and medical literature were retrieved. Search terms were developed to cover the range of DM methods and tools that have been adopted within health and medicine, including "DM", "artificial neural networks", "machine learning", "decision trees", "rule-based", "evolutionary", "genetic algorithms", etc. The retrieved results were examined and filtered to achieve an appropriate balance of review and original empirical research publications that was representative of the techniques that have been used and the areas in which they have been applied. The scope and contents of this review are therefore intended to reflect the use of DM tools in the health and medicine. Although the methods outlined above are predominantly developed from artificial intelligence, the importance of descriptive and inferential statistics in the DM process will be discussed and the review will consider both statistical and non-statistical methods of analysing data and the relationship between them. The review commences with a discussion of the various definitions of DM and ho

Definitions of DM

Various definitions of DM and synonyms for it have emerged in recent years that are not wholly consistent with each other (Benoit, 2001), which have created some confusion and suspicion in health and medicine, and this review will attempt to clarify what DM is, and equally importantly, what it is not. Benoit (2001, p.265) in his recent ARIST review on DM 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" process (Fayyad et al., 1996). Other authors (e.g., Bellazi and Zupan, 2001) have made the distinction between DM and intelligent data analysis (IDA). Similar to Knowledge Discovery in Databases (KDD), IDA describes the complete process of data analysis, including pre-processing etc., whereas DM describes the actual techniques involved. The relationship between DM and Knowledge Discovery in Databases (KDD) has been presented in detail elsewhere (see for example, Adriaans and Zantige, 1996; Benoit, 2001). Here it is sufficient to state that DM is the knowledge extraction stage of the knowledge discovery process which includes the selection of appropriate data possibly from a variety of sources, the cleaning of these data, the merging of the data from the different sources, the coding and re-coding of the data into an appropriate format, DM itself followed by the presentation and reporting of the results of the DM activities. Data mining is therefore a central part of the knowledge discovery process and encompasses a range of techniques selected on the basis of their suitability for the particular task in hand. The process of DM in health and medical research needs to incorporate not only the analyses of data but determining appropriate research questions and interpretation of the results (Richards et al. 2001).

While this portrayal of DM is fairly clear, confusion arises through the use of various synonyms for DM as discussed by Benoit (2001) and Trybula (1999). These synonyms include "knowledge discovery" itself, which as indicated above, is the larger process of which DM is but a part. Other terms, such as "information extraction", "pattern discovery" and "pattern identification" are all potentially misleading in that they describe either the end product of the process, rather than the process itself, or part of the range of ways in which data can be mined. The discovery or identification of patterns within data can either be the goal of the DM exercise or can be just a stage

in a more complex DM process, particularly when used in health and medicine when more precise objectives are set.

Perhaps the most misleading and potentially damaging synonym used for DM is "data dredging" (Benoit, 2001; Trybula, 1999) and in the context of health and medical research a sharp distinction must be made between these two processes. In this context, "data dredging" is used to describe the process of analysing a data set to try and uncover interesting relationships between the variables or patterns within the dataset. A useful analogy is implicit in the word "dredging", which suggests laboriously, and perhaps exhaustively, trawling or sifting though something, for example sand or mud, in the hope of finding something useful, interesting and/or valuable, e.g., a gold nugget or gem. Implicit in this analogy is the idea that person undertaking this activity has no clear *a priori* idea on what they are searching for but if they search for long enough something will emerge. In the context of data dredging, the analyst has no specific aim or research question but exhaustively seeks relationships and patterns within the data. While it might be argued that the ends of this data dredging process, i.e., the identification of a nugget of information, might justify the means the problem with this approach is that spurious relationships and patterns can be identified, which arise by chance, but which may be attached undue importance (Altman, 1991). For example if a dataset containing 20 variables was analysed to try and identify any relationships using traditional statistical methods, such as Chi^2 tests or Pearson correlation coefficients, then there would be 189 tests would be carried out. If the commonly used significance level of $p \leq 0.05$ was used to determine whether the null hypotheses should be rejected then by definition 1 in 20, or in this example eight or nine test results could appear to be statistically significant, purely by chance. Even if the significance level is lowered to $p \leq 0.01$, then 1 in 100, or in this example 1-2 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 and Altman, 1995), there is controversy

concerning the precise use of these adjustments in different situations (Perneger, 1998; Bender and Lange, 1999) and data dredging is considered to be inappropriate due to its lack of clear objectives and this potential to yield spurious results.

The phrase DM, on the other hand, presents a different analogy in that it implies drilling down in a much more focussed approach with a clear idea of what it is that is being mined for and with a reasonable expectation, gained through prior knowledge, of retrieving something worthwhile. Data mining suggests that the analyst has a good understanding of the data that they are mining and a clear idea, e.g., through their own knowledge of the subject area and/or earlier work, of the potentially useful and important information that may be retrieved. It also implies a systematic approach to the identification of previously hidden association, patterns and relationships (Pendharkar et al., 1999). Using a DM approach might therefore involve identifying a specific research question/hypothesis, e.g., through a substantive literature review or a discussion with domain experts, and answering/testing this using an existing data source by identifying patterns/ relationships/ associations centred round 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 focussed approach reduces this risk and is scientifically justifiable. The distinction between data dredging and DM must be borne in mind when discussing the use of DM in health and medical research and while a data dredging approach can produce unreliable and potentially damaging information, especially when used to inform clinical practice and decision-making. Data mining may therefore be defined by the approach that the researcher is adopting in analysing the data as well as by the methods that are used.

The potential of DM in health and medicine

In routine health and medical care, large volumes of data are routinely generated and stored either as part of the care process, for administrative purposes or for research (Coiera, 1997; Shortliffe and Blois, 2001; Peňa-Reyes and Sipper, 200l). The large amount of data and individual data items may be of low value in their own right, but there may be valuable information contained within it that is not immediately apparent, but which may be extracted and utilised using DM approaches (Kuo et al., 2001). This availability of health/medical data and information coupled with the need to increase continually our knowledge and understanding of the biological, biochemical, pathological and psychosocial and environmental processes by which health and disease are mediated mean that the medical and health sector may be particularly suitable for DM (Shortliffe and Blois, 2001; Shortliffe and Barnett, 2001).

Medicine and health deal with complex organisms, i.e., humans and/or patients, and therefore with higher-level processes in contrast to other branches of science e.g., physics and chemistry, which deal with relatively low-level processes (Shortliffe and Blois, 2001). While 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 clinical medicine and health, and high-level descriptors are necessary to try and encapsulate the complexity of humans (Maojo et al., 2002). Therefore while traditional computing applications, e.g., routine iterative number crunching using basic numerical programs, might be appropriate for the needs of the physical sciences, these are inadequate to deal with these complexities and DM techniques have therefore been adopted and developed for this purpose (Shortliffe and Blois, 2001). Furthermore, the large and complex search spaces that are generated through the data in health and medicine may be beyond the ability of clinicians to make a decision that has one of two possible outcomes (Peňa-Reyes and Sipper, 2000).

While the collection, management, analysis and interpretation of information is a fundamental part of the processes of clinical medicine and health care, not least of all in decision-making for the categorisation, treatment and management of diseases (Shortliffe and Barnett, 2001), the capture and coding of this information for storage in databases and information systems can reduce some of its informational complexity and value. However, by analysing and interpreting the encoded data either routinely or through DM as part of the knowledge discovery can help gain insights into the high level processes that would not otherwise be possible.

The methodologies that have been developed to help understand the complexity of information that are involved in health and medicine (Shortliffe and Blois, 1991; Maojo et al., 2002,), have traditionally been based on hypothetico-deductive reasoning (Lisboa, 2002), the inference of causal relationships (Altman, 1991; Bland, 2000) and the recognition of patterns and development of heuristics among others. Health and medicine research involve the development and testing of theories that involve the human being and its components that cannot be changed (Maojo et al., 2002). Added to this is the current tension within the practice of modern clinical medicine practice and healthcare between the need for a suitable evidence base which implies that the complex processes outlined above can be adequately measured and quantified (positivism) and the more recent post-positivistic realisation that this information can only be truly informative when coupled with more qualitative information and analyses.

As health and medicine have become more data and information-intensive the amount and variety of information collected and stored has increased (Richards et al., 2001), and this information has become more accessible through the increasing use of computers in the healthcare process (van Bemmel and Musen, 1997). Traditional methods of analysing data may not be adequate to deal with the large volumes of data and to maximise the potential for secondary analyses of these data.

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 occurrence can be explained by variables of interest (potential risk factors) relative to other often unexplained, and labelled random, variance (Giuliani and Benigni, 2000). Although such studies work well when there is a "single causative agent far exceeding all the others" (Giuliani and Benigni, 2000, p.308) many diseases and conditions, particularly non-infectious diseases, may have multiple causative agents or have many risk factors, and the traditional epidemiological and statistical approaches struggle to discriminate between a range of putative risk factor or causative agents and the random variance. In other words the "signal to noise" ratio is too low to be able to elucidate causes effectively (Giuliani and 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 et al., 2000). In addition, the low signal to noise ratio that is common in health and medical data means that the potential advantages of flexible non-linear DM tools compared to statistical techniques will not be realised. However this may be overcome as advances in understanding of risk factors for disease and health outcomes improves the potential for diagnostic and prognostic models (Biganzoli et al., 2002). In order to appreciate the potential of DM approaches to analysing health and medical information, traditional inferential statistical methods will be briefly discussed together with their limitations.

Traditional statistical methods for analysing such data and their limitations.

Traditional methods of analysing health and medical data have been developed within a positivist paradigm, in which hypothetico-deductive methods to set up null hypotheses have been proposed which are then tested using inferential statistical techniques based on parametric and non-

parametric tests, such as *t*-tests, Chi^2-tests, correlation and regression (Altman, 1991; Bland 2001). Although inferential statistical methods have long been accepted for use in science, health and medicine, they have their limitations and although they provide a measure of statistical significance, this does necessarily indicate their clinical importance (Last et al., 1999).

While some of these tests, e.g., *t*-tests, Chi^2-tests and correlation involve univariate or bivariate analyses and DM techniques offer little above and beyond these, other forms of analyses such as cluster analysis and regression can involve greater numbers of variables with complex interactions, which DM tools have the potential to augment. Linear regression is used widely in health and medical research to identify the association between one (simple linear regression) or more (multiple linear regression) independent, or predictor, variables and a continuous dependent, or outcome, variable, and to predict the value of the outcomes variable for a given value of a predictor variable (Altman, 1991; Bland, 2000). Linear regression is useful in that it is relatively straightforward and uses a single coefficient within the regression model to summarise the contribution of each predictor variable (Dusseldorp and Meulman, 2001). Adaptations of linear regression have been made to permit binary outcome variables to be used as the dependent variable (simple logistic regression) and nominal variables (multinomial logistic regression) through a transformation of the dependent variable (Altman, 1991). Rather than being used to predict the value of the outcome variable for a given value of a predictor variable, logistic regression uses the coefficient to calculate the odds, termed the odds ratio, of the category of interest occurring, for each category of an independent categorical variable relative to a selected reference category, or for each increment in the magnitude of an independent continuous variable. Logistic regression is particularly useful in health and medical research as many outcomes of interest occur, or can be represented, as binary variables, e.g., the presence or absence of disease, being alive or dead, a response to treatment or not, disease recurrence, etc., (Altman, 1991). Logistic regression can also

be useful in making predictions and has therefore been widely used for assisting in clinical decision-making for diagnosis and prognosis.

"Survival" analysis is a term used to describe studies in health and medicine that account for any event (i.e., not just mortality *per se*) occurring over a period of time within a population or group of interest (Altman, 1991). The statistical methods of analysing survival are based upon comparing the distribution of survival times of different groups of patients or on the development of appropriate regression models (e.g., logistic regression and Cox regression) to analyse the effects of variables on survival (Anand et al., 1999). Parametric models for survival analysis have proved inadequate to deal with the complex relationships between predictor variables and events of interests due to their assumptions on failure time distributions and the effects of the covariates on these distributions (Biganzoli et al., 2002). The development of semi-parametric method has overcome these limitations but only allows the identification of putative risk factors. Another important issue in developing methods for analysing survival in the context of health and medicine is in ensuring that models deal adequately with right-censored data, i.e., although models may able to analyse the time to an event happening they need to be able to differentiate between the event happening and it not happening within a given time period of analysis. For example, all the patients undergoing different treatments for cancer may not die within 21 months, or some of the patients who have had one myocardial infarction may never experience another. Although logistic regression is useful in medicine and health for analysing disease risk and predicting outcomes, it fails to consider the time at which an event occurs (Altman, 1991; Bland 2000), and this information is can be particularly important in determining the importance of putative risk factors for events of interest. The Cox proportional hazards regression model (Cox, 1972), often termed Cox regression, accounts not only for whether an event has occurred or not, but the length of time which it took that event to occur, or, if the event does not happen, to a point of censorship. Analysing survival for diseases and

conditions plays an important role in clinical medicine to enable health care professionals to develop prognostic indices for people following diagnosis, either for mortality in potentially terminal illnesses, recurrence of the disease, studying outcomes for different forms of treatment or for assessing the risk of adverse health events occurring.

Traditional 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 between variables is complex and multivariate (Biganzoli et al., 2002), and there are often interactions among the predictor variables and assessing these and their effect on the outcome variable can be complex (Dusseldorp and Meulman, 2001). In addition, the preponderance of non-linear relationships among health and medical data and the non-additive 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 other DM tools.

Logistic and Cox regression are important in generating population-based estimates of survival, for identifying putative risk factors and logistic regression can be 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 (this is explained below in relation to Table 1). It is then possible to evaluate these models by comparing the result with the actual outcome or diagnosis (Altman, 1991; Bland 2000). However, they are not used for making predictions concerning individual patients in a clinical setting (Anand et al., 1999; Botacci et al., 1997), and health care professionals tend to rely on their own knowledge, experience and judgement, which have their limitations and are prone to human error.

Decision-making by health care professionals is based on knowledge gained through initial training, updating this through continuing professional development and personal learning, and also by development of own experience (Brause, 2002). Early in their careers health care professionals have limited experience, which is particularly important for relatively new diseases/conditions or ones that are rare in occurrence. Health care professionals and managers are humans, who are better at pattern recognition tasks then basing decisions for example on statistical probabilities (Brause, 2002; Lisboa, 2002, Walker et al., 1999). Although some of these problems may be overcome, e.g., by consulting with more-experienced/ knowledgeable colleagues, decision-making, e.g., in diagnosis, may be flawed by lack of available experience with the particular condition or their ability to deal with complex data. Data mining may help overcome these problems, e.g., by identifying patterns that were not previously apparent, or by learning from data to make decisions, predictions, or prognoses and diagnoses (Downs et al., 1996).

In order to assess the suitability of traditional statistical methods of analysing data and to compare these with the performance of DM methods, appropriate means of evaluating the performance of diagnostic, prognostic and other data analytic tools.

Evaluation of methods

In order to assess the value of any data-mining tool for use in routine clinical practice, it is important to be able to evaluate its effectiveness, and compare that with other methods of analysis. For example, the correct diagnosis of diseases and being able to make an accurate prognosis is vital for managing the overall care of a patient. When developing and evaluating new methods of diagnosing conditions and making prognoses, it is necessary to compare the predicted value with the true diagnosis, or with the eventual outcome or prognosis. The predicted diagnosis or predicted prognosis can be compared with the true diagnosis and eventual outcomes respectively using a classification table as shown in table 1 (Altman and Bland, 1994a).

	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

Table 1: Table to compare the results of the true diagnosis with the results from the prediction.

It can be seen from that the true diagnosis showed that *n*= a+c individuals or cases were diagnosed as not having the condition, and of these the new method correctly diagnosed *n*= a as not having the condition (called *true negatives*). The true diagnosis showed that *n*= b+d individuals or 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) (Lavraĉ, 1999b, Altman and Bland, 1994a).

Sensitivity, sometimes called recall in information retrieval, is the measure of how many of the individuals with the condition that the test detects, in other words the proportion or percentage of true positives (Altman and Bland, 1994a). This is calculated by [sensitivity = d /(b+d)] and is expressed as either a decimal or a percentage (multiplied by 100). 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 will potentially benefit from any intervention, e.g., medication or treatment, given to people whom the test identifies. Specificity (sometimes called precision in information retrieval), on the other hand, is a measure of how many of the individuals without the condition that the test detects as not having the condition, i.e., the rate of detecting true

negatives, and is calculated by [specificity = a /(a+c)], (multiplied by 100) if expressed as a percentage. Two further measures for evaluating a method of diagnosis are the positive and negative predictive values (Altman and Bland, 1994b). The positive predictive value (ppv) is the proportion (or percentage) of individuals that the method diagnoses as having the condition that actually have the condition, and is calculated by [Positive predictive value = d /(c+d)], × 100. Conversely, negative predictive value (npv) is the proportion (percentage) of individuals that the method diagnoses as not having the condition that actually do not have the condition, and is calculated by [Negative predictive value= a /(a+b)] (× 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 and Bland, 1994c). Although by enabling the comparison of sensitivity and specificity in a single graph, this plot gives 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 and Altman, 1994c; Jefferson et al., 1995; MacNamee et al., 2002). Many DM methods are aimed at developing improved methods for making decisions, especially for diagnosis or prognosis, and evaluating the outcomes in terms of these values of sensitivity, specificity, positive and negative predictive values in this specialised area of DM will be discussed. The relative importance of these measures of effectiveness within a particular clinical or health context has an important impact on the development of tools and will be discussed in more detail later.

Data mining tools for health and medicine.

Data mining tools generally one of two forms of learning (supervised or unsupervised) for classification, making predictions and other DM activities (Peňa-Reyes and Sipper, 2000). Supervised learning is used when an DM tools is trained to recognise different classes of data by

exposing the network to a series of examples for which it has target answers (the training data set), and then testing how well it has learned from these examples by supplying it with a previously unseen set of data which it then classifies (the 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 to partition data.

Recent reviews by Lavrač (1999a, 1999b) have described and discussed different methods of machine learning available for DM in health and medical research. Machine-learning methods include three main types of DM tool including the inductive symbolic rule learning, statistical or pattern recognition methods, and artificial neural networks (Lavrač, 1999a). All of these techniques seek to improve medical diagnosis and prognosis by analysing data from previous patients, defined as a training set, and from this learning process to predict the diagnosis and/ or prognosis for new groups of patients, the test set. Lavrač (1999b) categorised DM methods into those methods that produce symbolic representations from the data they are analysing and include rule induction methods, decision trees and logic programs and those that produce a sub-symbolic representation, which include instance-based learning methods such as nearest neighbour algorithms, artificial neural networks and Bayesian classifiers. 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).

Inductive learning of symbolic rules

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

IF *Condition(s)* THEN *Conclusion*

Or

Condition(s) $\rightarrow$ *Conclusion*

in which the *Condition(s)* part includes one or more tests for values of the variables (labelled 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 variables. 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 the two.

Rule-based approaches have been used in a number of areas in health and medicine including 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 analysing meningitis data (Zhong and Dong, 2002).

Decision trees

Decision trees, also called tree-based methods, are a very popular type of DM technique and are based on a method called recursive partitioning, that has been used for solving regression and classification problems in health and medical research (Dusseldorp and Meulman, 2001; Kuo et al.,

2001). Regression trees are used to model continuous outcome variables to predict specific values for a variable of interest and classification trees are used to model categorical variables in order to predict to which group an individual or case belongs (Dusseldorp and Meulman, 2001; Kuo et al., 2001). The decision tree model can be used for descriptive purposes as well as for making predictions (Kuo et al., 2001; Ennis et al., 1998). The model is presented in the shape of a tree with 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 tumours as malignant or benign and showed that the overall accuracy of the decision tree model was better than that of the physician, as well as the sensitivity, specificity, positive and negative predictive values. Recursive partitioning has been shown to be of value in identifying interactions among variables (Carmelli et al., 1991). Carmelli et al. (1991) compared the use of recursive partitioning with Cox regression for examining the relationship between baseline biological and behavioural characteristics and mortality due to coronary heat disease and cancer over 27 years. Although both Cox regression and recursive partitioning were useful in determining factors associated with mortality, recursive partitioning enabled the identification of subgroups of individuals with particular characteristics and survival features (Carmelli et al., 1991).

Artificial Neural Networks.

Although the original research into ANNs started in the 1950s it is only relatively recently that they have emerged as a useful and effective set of tools for tackling a range of DM problems, including pattern recognition, prediction of outcomes, classification and partitioning of multivariate data (Haykin, 1999; Bath and Philp, 1998). They have been applied in a variety of domains (Dayhoff, 1990; Trybula 1999; Benoit 2001), including health and medicine (Cross et al., 1995; Baxt, 1995; Dybowski and Gant, 1995; Brause, 2002). ANNs are so-called because they have structures and

processes that are modelled on the architecture and learning processes in biological nervous systems. ANNs have the potential to extract information that is complementary, rather than an alternative, to that obtained using statistical methods. ANNS differ from such methods in being adaptive, i.e., the data are presented to the ANN iteratively, during which the network "learns" and then revises the predictions or classifications it has made. During these iterations the network is trained and is able to "recognise" patterns in the data and as a result of the training the ANN can make predictions or classifications (Lipmann, 1987).

ANNs use both supervised and unsupervised learning to mine data. ANNs employing unsupervised learning, e.g., Kohonen self-organising maps, are able to analyse multi-dimensional data sets in order to discover the natural patterns, or clusters and sub-clusters, that exist within the data (Lipmann, 1987; Kohonen, 1995). 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, e.g., 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 modelling data derived from health and medicine (Lavrač, 1999b).

Artificial neural networks have been used in a wide variety of applications in clinical medicine, including diagnosis, risk assessment, analysing medical images and wave forms, treatment selection and predicting outcomes and drug activities and responses to medication in clinical pharmacology (cited in and in Lavrač, 1999b). Artificial neural networks have been used for diagnosing a wide range of health and medical problems including myocardial infarction (heart attack) (Baxt, 1991; Baxt and Skora, 1995; Ennis et al., 1998), different forms of cancer (Pendharkar et al., 1999),

detecting ischemia (Papaloukas et al., 2002), appendicitis, back pain, dementia, psychiatric emergencies, pulmonary embolism, sexually transmitted diseases, skin diseases and temporal ateritis (cited in 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 very serious and potentially fatal (Baxt, 1995). Clinicians therefore tend to diagnose to avoid the risk of missing diagnosis of myocardial infarction and 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, to predict outcomes following surgery in intensive care units and orthopaedic rehabilitation units (Grigsby et al., 1994), recovery from prostate, breast and ovarian cancer (Downs et al., 1996), cardiopulmonary resuscitation and liver transplantation (Doyle et al., 1994) and rehospitalization following stroke (Ottenbacher et al., 2001). Neural networks have also been used extensively for analysing survival data (Biganzoli et al., 1998; Biganzoli et al., 2002; Cacciafesta 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 et al., 2001).

ANNs have a number of advantages over statistical techniques that make them particularly suitable for mining health and medical data. ANNs are non-parametric and therefore do not make assumptions about the underlying distributions of the data that statistical methods make (Lippmann, 1987). ANNs therefore may be more robust and perform better when data are not normally distributed or where there is a non-linear relationship between predictor variables and an outcome variable. Artificial neural networks are able to analyse the higher–order relationships frequently present in health and 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 why their use has been regarded with suspicion and mistrust within the medical and statistical communities. Downs et al. (1996, p.411) discussed the need to supplement using neural networks with the extraction of symbolic rules to "provide explanatory facilities for the network's 'reasoning'" and developed symbolic rules to try and explain the reasoning behind the decision-making by the neural network, and a number of techniques have been developed which permit this (Andrews et al., 1995).

A further problem with ANNs is that their performance on test data set is often worse than that achieved through the training set (Brause, 2002) due to the network over-training and adapting to any biases in the training set. Although on solution to this is to use 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, or when the network is used on truly independent data set, e.g., in clinical environment (Brause, 2002). Cross et al. (1995) commented that on the less rigorous development of artificial neural networks compared to that for conventional statistical tests and that large scale clinical trials may be needed to evaluate their use statistically before they are accepted as a diagnostic tool. Further limitations of DM tools will be discussed later.

Evolutionary DM tools

As the name implies 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, 1998; Peňa-Reyes and Sipper, 2000). Evolutionary tools are

methods of searching through the high dimensional space of possible solutions to a given problem to find an optimal solution and are particularly suited to use in DM in health and medicine given the preponderance of variables and multivariate relationships discussed previously.

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 make-up 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. Reproduction may be asexual in which only one individual of a species is involved, or sexual in which two members of the species are involved. Sexual reproduction involves the development of an entirely new genotype by recombination of the genetic material of the parents. This process is supplemented by mutation in which small changes to the genetic material are introduced at random. The offspring from sexual reproduction then undergo the process of 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 using heuristics (Peňa-Reyes and Sipper, 2000). The features or attributes of each individual are encoded via genes on a chromosome and associated with each chromosome is a fitness function, which measures its

suitability to the environment or problem situation. The population then 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 cross-over to create new genotypes, which are also subject to mutation. The offspring from this process then join the population and each has fitness function associated with its genotype. 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 form 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 cross-over ensures that new combinations of genetic material are introduced and "move towards promising new areas of the search space" (Peňa-Reyes and Sipper, 2000, p.23). Mutation prevents the process from converging in local optima that do not represent optimal solutions and the new individuals then enter the environment and the next generation commences. Thus, similar to natural evolution, over time and 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.

One of the advantages of evolutionary computational tools over more traditional search methods is that they are able to combine a search of all the available search space with the capacity to search the most promising areas (Peňa-Reyes and Sipper, 2000). The results of the searches in these spaces can then be combined via cross-over 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 on the search space and make fewer assumptions about it (Peňa-Reyes and Sipper, 2000). The key considerations when using evolutionary DM tools include not only how to encode the features of possible solutions into genes but also how to measure the fitness of the individuals and chromosomes. These two issues are dependent on the specific problem and have to be tailored to the particular needs of that problem (Peňa-Reyes and Sipper, 2000), and although the processes of selection, cross-over and mutation are relatively problem-independent, it is likely that these issues will need to be considered in adapting tools to particular problems.

Genetic algorithms

Several different types of evolutionary DM tools exist although the there is a deal of similarity among these types and they are all based on the principles and process of evolution. The most commonly–used type of evolutionary tools are genetic algorithms, which represent the genome (genotype) of the individual (phenotype) using a fixed-length binary string (Peňa-Reyes and Sipper, 2000). Although genetic algorithms 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 (termed redundancy), in order to achieve a good solution (Peňa-Reyes and Sipper, 2000). Using genetic algorithms, the number of individuals (population) is kept constant and during each generation these are decoded and their fitness is evaluated and the fittest are selected for reproduction.

Genetic algorithms have been used for analysing sleep patterns (Baumgart-Schmitt et al., 1998), diagnosis of female urinary incontinence and breast cancer (cited in Peňa-Reyes and Sipper, 2000), development of prognostic systems for colorectal cancer (Anand et al., 1999), selection of features for recognizing skin tumors (Handels et al., 1999), prediction of depression after mania (Jefferson et al., 1998b), predicting outcomes after surgery, predicting survival after lung cancer (Jefferson et al.,

1998a), improving response to warfarin (Naranyan and Lucas, 1993), survival after skin cancer and estimation of tumor stage and lymph node status in patients with colorectal adenocarcinoma (cited in Peňa-Reyes and Sipper, 2000).

Genetic programming and other evolutionary methods

Work by Koza (1990a and b) developed and extended the idea of evolutionary computational tools such as genetic algorithms by using genetic programming. While the basic evolutionary principles of genetic algorithms and genetic programming are similar, the features by which these tools carry out their tasks are fundamentally different and are discussed by Peňa-Reyes and Sipper (2000). Genetic programming encodes possible solutions to problems as computer programs rather than as binary strings and to achieve this they use parse trees and functional programming languages, unlike genetic algorithms, 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 subtrees beneath these points are exchanged between the parents (Peňa-Reyes and Sipper, 2000). Genetic programming tools have been less widely adopted as a data-mining tool in health and medical research than genetic algorithms but have been used to identify causal relationships in a database containing information on children with limb fractures and to identify relationships in a database containing information on spinal deformation (Ngan et al., 1999), to classify brain tumours into meningioma and non-meningioma classes (Gray et al., 1998), learning rules from a fractures data base (Wong et al., 2000) and for the diagnosis of chest pain (Bojarczuk et al., 2000).

Evolution strategies and evolution programming, two other methods of evolutionary computation, have had some relatively little use in mining health and medical data and are described by Peňa-

Reyes and Sipper (2000). Their use has been restricted to for analysing sleep patterns (Baumgart-Schmitt et al., 1998), detecting breast cancer using histologic data (Fogel et al., 1995) and radiographic features (Fogel et al., 1997) and optimising electrical parameters for therapeutic stimulation of the carotid sinus nerves (Peters et al., 1989).

Combined approaches

Evolutionary computing techniques have been used in combination with other tools for mining health and medical data. Genetic algorithms have been combined with several statistical and non-statistical methods as a way of optimising the variables for inclusion in models. Several groups of researchers have combined genetic algorithms with neural networks for detecting and diagnosing breast cancer (Abbass, 2002; Fogel et al., 1995), predicting response to warfarin (Naranyan and Lucas, 1997), predicting outcomes following surgery (Jefferson et al., 1997), predicting haemorrhagic blood loss (Jefferson et al., 1998a), predicting depression following mania (Jefferson et al., 1998b) and for predicting falls and identifying risk factors for falls in older people (Bath et al., 2000). Fogel et al. (1995) used evolutionary artificial neural networks for analysing 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. One of the problems in the use of artificial neural networks is that they can get stuck in local optima, and although increasing number of nodes and weights associated with them can help overcome this problem this is computationally more 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). Genetic algorithms have been used in combination with Bayesian networks to predict survival following malignant skin melanoma (Sierra and Larrañaga, 1998). Ngan et al. (1999) also used genetic programming in combination with Bayesian networks to

identify rules for limb fracture patterns and for classifying and treating scoliosis. Holmes et al. (2000) combined a genetic algorithm with a rule-based system for epidemiologic surveillance. Peňa-Reyes and Sipper (1999) combined genetic algorithms with a fuzzy system for the diagnosis of breast cancer. Although these studies represent attempts to combine evolutionary computing techniques with DM tools there has been little work combining evolutionary computing methods with statistical methods to optimise the variables used in predictive models (Jefferson, 2001) and there is 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 have been particularly used for diagnosis and prognosis of diseases and, in this section, their application in the diagnosis of breast cancer and for prognosis are discussed.

Breast cancer is one disease that has attracted a deal of interest from data miners, particularly in relation to diagnosis. Reasons for this include its high incidence and relatively high mortality associated with it, the importance of early diagnosis, and, as Abbass (2002, p.265) suggests, because of the very high "economic and social values" associated with it. Problems with the traditional assessment of mammographic data have included inconsistencies in interpretation resulting in poor intra- and inter-observer disagreement (Abbass, 2002; Fogel et al., 1997). The proposed reasons for this have been poor image quality of mammographic images and human fatigue and error, and have led to the development of search for pattern recognition techniques to supplement the diagnosis by the radiologist (Fogel et al., 1997). The aim of such developments has been to reduce the rate of false negative diagnoses to improve the sensitivity. However, given the cytotoxic side effects of chemotherapy and radiotherapy and psychosocial consequences of breast surgery it is also important to ensure that the number of false positive diagnoses is minimised, i.e.,

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

Wu et al. (1993) reported artificial neural networks that were better at analysing mammographic data than radiologists for decision-making in relation to the diagnosis of breast cancer. However these data had been extracted by radiologists, and the authors recommended that the real potential of neural networks was to assist the radiologists in recommending further tests to be undertaken. Setiono (1996, 2000) developed an accurate neural network program that was used pruning to extract rules to provide information on the basis on which the network had made its decisions and overcome the "black box" element of neural networks. Many of the cited studies used the same Wisconsin Breast Cancer data set for developing the models. While this is useful for comparing the effectiveness of different tools developed at different times, it emphasises the need to test DM tools on new sets of data in different settings, in addition to the ones in which they were developed (Lisboa, 2002).

Walker et al. (1999) described the use growing cell structure technique to differentiate between benign and malignant breast tumours. This technique, which was shown to have a similar performance to logistic regression, allows the multidimensional data (the predictor variables) to be viewed as two-dimensional colour images. The particular value of the this visualisation was that it permits health care professionals to perceive relations between the predictor and outcome variables, as well as interactions among the predictor variables (Walker et al., 1999).

Prognosis has already been highlighted as an important area for patient care and 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 and Simon, 1995; Xiang et al., 2000; Zupan et al., 1999). Although some studies have shown that data mining methods perform better than statistical models for analysing survival (Anand et al., 1999; Zupan et al., 1999), the study by Anand et al. (1999) showed that none of the three DM tools was able to handle the censored data as well as Cox regression dealt with them.

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

In many diseases a wide variety of clinical variables influence the prognosis for a disease and an individual, and making predictions for individual patients remains problematic, but is particularly important among patients diagnosed with a potentially terminal illness. Although it is known that approximately *x*% 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, many of whom may want to know “How long will I live?”, especially as the deviation from the

mean varies greatly among such patients (Bottaci et al., 1997). Anand et al. (1999) highlighted the need for better tools for prognosis of the disease especially in those patients with potentially terminal diseases, in which palliation and maintaining quality of life may become the main objective. Information on the likelihood of survival and expected life expectancy can greatly assist in improving the quality of life of such patients by providing appropriate counselling and disease management (Anand et al., 1999).

Challenges for DM in health and medicine.

Having described a number of DM tools and discussed their application in the domain of health and medicine the challenges that such tools face will be discussed together and suggestions of how these may be tackled. Mistrust and suspicion of DM tools can be overcome to some extent by acknowledging and presenting clearly the limitations of DM tools and avoiding exaggeration of their potential and a number of authors have made recommendations for the development of DM models and decision support tools based on DM tools in order that they may gain wider acceptance (Kononenko et al., 1998; Lisboa, 2002).

At one level there are a number of "technical" challenges that DM tools have to addressed in order to gain wider acceptance among health and medical professionals and statisticians (Lisboa, 2002) and at another level, there are more "human" challenges that need to be addressed. Some of these more technical issues, such as the appropriate design of studies that develop and test DM tools and the need to represent data in an appropriate format (Isken et al., 2002) and to ensure that the data are of a high quality (e.g., in relation to missing data, consistency of data collection and recording), are common to statistical and DM methods. The statistical aspects of underlying data and models may not being given appropriate consideration (Biganzoli et al., 2002) and it is important that descriptive statistics are available of data that are being mined as well as data that are being tested statistically.

While many studies in health and medicine have made use of descriptive and inferential statistics without the apparent need for data mining tools, data mining tools cannot be developed in isolation of traditional statistical methods.

Lisboa (2002) discussed the need to clarify the purpose of studies and to specify in advance what is expected to be of value in future studies. Data mining tools being used are not necessarily the most advanced available or it can be difficult to determine that the chosen model is the best possible (Tu, 1996). The performance of DM tools could be enhanced by using more advanced types of genetic algorithms, artificial neural networks, etc. (Anand et al., 1999).

Data may be collected for a purpose other than that for which they are being analysed and therefore not clinically relevant for the diagnosis or prognosis for which they are being used (Richards et al., 2001, Wyatt and Altman, 1995). Missing data are particularly a problem in medical databases and often arise through incomplete data being recorded or human error in recording/ transcription (Richards et al., 2001; Brause, 2002). Missing data can be dealt with by removing variables and/or cases that have a high proportion of missing values to minimise the amount of missing data, although this approach may introduce bias to the remaining data because cases (individuals) with large amounts of missing data may not be representative of the sample have particular associations with the outcome of interest. Replacing missing data with statistical descriptors, e.g., the mean value for a variable, is generally acceptable if done with care, but may introduce bias to the data (Altman, 1991).

Ensuring that other forms of bias are not allowed to influence the results when developing and testing DM tools is important and concealing the correct classification from domain experts until the studies are completed so that the DM methods can be truly acknowledged responsible for the associations that were discovered and 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 then combining the domain knowledge of clinical experts with sophisticated analytic techniques may help to improve performance further. Richards et al. (2001) and Wyatt (1995) among others have stressed the need for training and testing of DM tools to be carried out on independent data sets, and the problems associated with this have already been discussed, as well as the need for appropriate training, validating and testing of data and systems before implementation in real settings. Lisboa (2002) also recommended that good practice be followed in designing models, particularly with respect to ensuring that over-fitting is controlled 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 method (statistical or non-statistical) being very good at identifying the former class but relatively poor at identifying the latter class.

A problem with the development of diagnostic and prognostic tools through DM is that it increases the complexity of decision-making for health care professionals (Kononenko et al., 1998), so that such tools need to be made as simple to use as possible with user-friendly interfaces. Lisboa (2002) commented that knowing how a model improves accuracy in decision-making is as important as whether it improves accuracy and emphasised the need for health care professionals to understand how any model works for them to be able to take responsibility for the results it produces. This does not only mean understanding basic mathematical principles underlying the models (Koh and Leong, 2001), but how the models reached particular decisions, the opening of the "black box" that has been previously discussed. While the accuracy/performance of DM tools may be greater than that of traditional methods of analysis the lack on 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).

Although there has been considerable progress in developing sub-symbolic DM tools that are able to extract rules to provide an explanation of how they reached their decision-making (Andrews et al., 1995), these have not yet been widely adopted for use in health and medicine, and this requires further progress.

A number of authors have identified the need to establish an appropriate evidence base for the use of DM tools in medical and health practice, especially when being used for developing tools for diagnosis, prognosis, etc. (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 fully evaluated. Johnston et al. (1994) identified the need to evaluate systematically computer-based decision support systems not only in relation to reliability, acceptability and accuracy, but also with respect to improving the clinical behaviour and performance of health care professionals, and ultimately to improve patient well-being and patient outcomes. While Johnston et al. (1994) acknowledge that the accepted gold standard for evaluating healthcare interventions, the randomised controlled trial (RCT), may not always be practical or feasible for evaluating computer-based decision support systems that have been developed through the use of DM techniques, nevertheless investment in evaluating the effectiveness and efficiency of such systems is necessary to maximise the potential benefits and minimise the potential for harm or waste that may arise. Lisboa (2002) highlighted the need to evaluate of DM tools through multi-centre RCTs and to establish an appropriate evidence base for the use of DM tools (Brause, 2001; Lisboa, 2002; Anand et al., 1999).

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 in this respect for both symbolic and sub-symbolic methods. Certainly the ability of neural networks to detect lower order relationships previously unknown but which can then be tested using statistical models can help gain their acceptance among medical and health professionals, and increase their trust when interactions among the data are discovered that that cannot be verified using statistical methods (Lisboa, 2002). An additional problem is that DM tools may identify and report patterns not accepted and not in line with current understanding (Richards et al., 2001, Wyatt, 1995), which may limit their acceptance among health care professionals.

Data mining have been shown to be useful for generating hypotheses for further testing e.g., to identify associations or relationships between variables/data that are then tested using conventional statistical techniques (Richards et al., 2001). There is a need to focus on the one hand on the way in which DM methods can complement the use of statistical techniques in analysing health and medical data but also to emphasise the added value that DM methods can bring in the knowledge discovery process. Understanding and appreciating the similarities and differences between DM tools and statistical methods, and valuing the unique contribution that each makes in improving our understanding of the processes underlying health and illness, e.g., while both Cox regression and tree-structured survival analysis both 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 et al. (2000) compared a variety of techniques including correlation analysis, discriminant analysis, data visualisation and artificial neural networks to analyses data from a heart disease database, which meant that it was possible to identify people at risk of heart disease, risk factors for heart disease and establishing

multivariate relationships among the predictor variables. Therefore, the maximising the potential of data mining may require the use of statistical alongside non-statistical methods.

Lisboa (2002) commented on the need to ensure that the purpose of studies is clear at the outset. It is particularly important to understand the objectives in trying to improve the performance of prognosis and diagnosis. Although obtaining 100% accuracy may be the seen as the overall aim, this is rarely achieved and the relative importance of sensitivity, specificity positive and negative predictive values within the context of clinical care on the relative importance of these evaluation measures, as was discussed in the diagnosis of breast cancer. 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 early as possible so that treatment can commence at an early stage in the disease and improve the outcomes for patients (Fogel et al., 1997; Fogel et al., 1995). For other diseases, however the imperative may be to ensure that the specificity is very high to minimise the number of people who are wrongly diagnosed as having the disease and receiving unnecessary treatments (Downs et al., 1996). On the one hand diagnosing all positive cases may be important to improve survival rates reduce co-morbidities. Reducing false positives may be important so that patients are not given drugs and medication with potential toxic effects (and high costs) unnecessarily and allowing health care professionals to spend maximal time with true cases (Abbass, 2002).

Bellazi and Zupan (2001) and Liebowitz (2001a) recently discussed the overlap between knowledge management and DM, suggesting that DM is an important part of the knowledge management process within health care organisations. Data mining relies on the explicit knowledge present in the available health and medical literature that is used by clinical researchers, clinicians, methodologists and information specialists to help identify appropriate research questions. The

implicit knowledge of clinicians, health care professionals and health service managers is also required for helping to develop and understanding of the data and for evaluating/assessing and interpreting the results. The explicit knowledge of clinicians, health care professionals may also be embodied into specific DM methods, e.g., Bayesian networks and fuzzy systems, for analysing the data (Bellazi and Zupan, 2001). This highlights another important aspect of the use of DM in the context of medicine and health in that it requires the multidisciplinary collaboration between health and medical professionals and information analysts (Kuo et al., 2001).

Despite all research and success of DM tools no tools or automated process arising from DM has been adopted on a routine basis (Abbass, 2002), Abbass (2002) has proposed several possible explanations in that the aim of such systems might be perceived to be to replace the health care professional, as well as people's mistrust and suspicions of technology. This illustrates the need to emphasise the complementary nature of DM tools, as an adjunct to decision-making by health and medical professionals rather than to replace them (Abbass, 2002). Botacci et al. (1997) emphasised the need to combine to use the clinical judgement and experience for careful interpretation of the results, and it must be made clear that any data mining tools are "just another source of possibly useful information" (Kononenko et al., 1998, p. 403) that the health care professional may use in decision-making with and providing care for patients.

The point has already been made that data need to be represented in an appropriate format and that health care professionals should be able to interpret the results and understand how DM models reached their decisions. Lisboa (2002) commented on the increase in DM methods that allow visualisation of the data and their potential to assist in the decision-making process and the Growing Cell Structure technique demonstrates the value of visualisation (Walker et al., 2002). Human beings are better at analysing and interpreting data that are presented visually rather than

numerically (Walker, 1999; Lisboa, 2002) so that DM models that are able to present a visual image of the way in which a decision was made may gain greater acceptability among health care professionals. Health care professionals have to trust DM tools, and therefore they need not only to understand their performance in terms of accuracy etc., but also in terms of their limitations and understand that they are there to aid decision-making by health care professionals not to replace it (Cross et al., 1995). In the same way that healthcare professionals build up trust in each other through sharing information, decision-making etc., they need to develop trust in the decision-making tools (Abbass, 2002). This re-emphasises the need for health care professionals, statisticians and data miners to collaborate together to improve models and methods of tackling the complex issues of analysing health and medical data and overcome the suspicions of the former and any over-confidence among the latter (Biganzoli et al., 2002; Kuo et al., 2001).

Sullivan and Mitchell (1995) discussed the need to evaluate the use of tools from a patient perspective and Lisboa (2002) commented the difference between assessing whether a DM model improves on the performance of health care professionals and assessing whether there is any overall improvement in patient outcomes through the use of tools developed using DM. Although studies have demonstrated the effectiveness of DM techniques in terms of diagnostic or prognostic accuracy, little research has shown an improvement in patient health and well-being.

A final, but by no means the least, important consideration in health and medicine is that of ethics. Although ethical considerations are of importance in other disciplines they come under particular close scrutiny in health and medicine because of its involvement with people, who are often in a vulnerable position when being treated for a condition. It is important therefore that any developments in DM tools are conducted ethically, with ultimate well being of patients and the public in mind, not only in reaching an end, but also in developing the means to achieve this.

Conclusions

In this review selected DM and statistical techniques that have been used in medicine and health have been examined and their strengths and weaknesses have been discussed. Our understanding of the complex processes underlying health and illness is increasing and the available data are becoming more numerous and it is becoming possible to integrate and store ever larger volumes in data warehouses. As this happens, the demands for appropriate ways of processing these data and answering clinically relevant questions will increase as well. The limitations of current ways of analysing medical and health data will become more apparent and the search for new and alternative methods will intensify. Data mining has the potential to play a part in this. However, to achieve greater acceptance and use in clinical settings on a routine basis DM must be seen as a systematic process with clear, precise and realistic objectives. The greatest opportunity for DM is that it becomes widely recognised as complementary to traditional methods of analysing data in health and medicine, and that it can be used alongside, and together with, descriptive and inferential statistical methods in the knowledge discovery process, so that the strengths of different techniques can be maximised and their weaknesses can be minimised. The development of DM applications requires investment of time and resources (Koh and Leong, 2001), but perhaps what is most essential is that it is part of a process that involves the multidisciplinary and open-minded collaboration of medical and health care professionals, statisticians and information professionals.

Bibliography

Abbass HA (2002) An evolutionary artificial neural networks approach for breast cancer diagnosis. Artificial Intelligence in Medicine. 25(3):265-281.

Adriaans P and Zantige D (1996) Data Mining. Addison Wesley. Harlow, UK.

Altman DG (1991) Practical Statistics for Medical Research. Chapman Hall/CRC. London.

Altman DG and Bland M (1994a) Statistics Notes: Diagnostic tests 1: sensitivity and specificity. British Medical Journal; 308: 1552.

Altman DG and Bland M (1994b) Statistics Notes: Diagnostic tests 2: predictive values. British Medical Journal; 309: 102.

Altman DG and Bland M (1994c) Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots. British Medical Journal; 309: 188.

Anand SS, Smith AE, Hamilton PW, Anand JS, Hughes JG and Bartels PH. (1999) An evaluation of intelligent prognostic systems for colorectal cancer. Artificial Intelligence in Medicine. 15(2):193-214.

Andrews R, Diederich J and Tickle AB (1995) Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowledge-Based Systems; 8(6):373-389.

Bath PA, Morgan K, Pendleton N, Clague J, Horan M and 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 and 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 WM and Eilers R (1998) On the use of neural network techniques to analyze sleep ECG data. Neuropsychobiology; 37:49-58.

Baxt WG (1991) Use of an artificial neural network for the diagnosis of myocardial infarction. Annals of Internal Medicine; 115:843-848.

Baxt WG (1995) Application of artificial neural networks to clinical medicine. Lancet 346, 1135-1138.

Baxt WG and Skora J (1995) Prospective validation of artificial neural networks trained to identify acute myocardial infarction. Lancet.

Bellazzi R and 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.

Benoit 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 and 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 and 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. Oxford Medical Publications. Third Edition. Oxford.

Bland JM and Altman DG. (1995) Multiple significance tests: the Bonferroni method. British Medical Journal; 310: 170.

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

Bojarczuk CC. Lopes HS. Freitas AA. Genetic programming for knowledge discovery in chest-pain diagnosis. IEEE Engineering in Medicine & Biology Magazine. 19(4):38-44, 2000 Jul-Aug.

Bottaci L, Drew PJ, Hartley JE, Hadfield MB, Farouk R, Lee PWR, Macintyre IMC, Duthie GS and Monson JRT (1997) Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. Lancet Vol. 350, Issue 9076:469-472.

Brause RW (2001) Medical analysis and diagnosis by neural networks. In Proceedings of ISMDA 2001, Lecture Notes in Computer Science 2199: pp.1-13. Editors: Crespo J, Maojo V and Martin F. Springer-Verlag. Berlin.

Brickley MR, Shepherd JP and Armstrong RA (1998) Neural networks: a technique for development of decision support systems in dentistry. Journal of Dentistry 26(4):305-309.

Cacciafesta M, Campana F, Trani I, Annichiarico R, Leonetti-Luparini R, Gianni W, Cicconetti P and Piccirillo G (2000) A neural network study of the correlation between metabolic cardiovascular diseases and disability in elderly people. Archives of Gerontology and Geriatrics 31(3): 257-266.

Cacciafesta M, Campana F, Piccirillo G, Cicconetti P, Trani I, Leonetti-Luparini R, Marigliani V and 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 GE, Dame A, McElroy M, Gelb AB and Rosenman RH (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 YM, Ho SH, Cho KW, Lee DH and Ji SH (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. Arnold, London and Oxford University Press, New York.

Cox DR (1972) Regression models and life tables. Journal of the Royal Statistical Society B; 4:232-236.

Cross SS, Harrison RF and Kennedy RL (1995) Introduction to neural networks. Lancet; 346:1075-1079.

Daly LE and Bourke GJ (2000) Interpretation and Uses of Medical Statistics. Fifth Edition. Blackwell Science, Oxford.

Dayhoff, J. E. (1990) Neural Network Architectures An Introduction. New York Van Nostrand Reinhold.

Doddi S, Marathe A, Ravi SS and Torney DC (2001) Discovery of association rules in medical data. Medical Informatics; 26(1):25-33.

Downs J, Harrison RF, Kennedy RL and Cross SS (1996) Application of the fuzzy ARTMAP neural network model to medical pattern classification tasks. Artificial Intelligence in Medicine; 8(4):403-428.

Doyle HR, Dvorchik I, Mitchell S, Marino IR, Ebert FH, McMichael J and Fung JJ (1994) Predicting outcomes after liver transplantation a connectionist approach. Annals of Surgery; 219(4):408-415.

Dusseldorp E and Meulman JJ (2001) Prediction in medicine by integrating regression trees into regression analysis with optimal scaling. Methods of Information in Medicine; 40:403-409.

Dybowski R and Gant V (1995) Artificial neural networks in pathology and medical laboratories. Lancet; 346:1203-1207.

Ennis M, Hinton G, Naylor D, Revow M and Tibshirani R. (1998) A comparison if statistical learning methods on the GUSTO database. Statistics in Medicine; 17:2501-2508.

Faraggi D and Simon R (1995) A neural network model for survival data. Statistics in Medicine, 14:73-82.

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

Floyd CE, Lo JY, Yun AJ, Sullivan DC and Kornguth PJ (1994) Prediction of breast cancer malignancy using an artificial neural network. Cancer; 74(11):2944-2948.

Fogel DB, Wasson EC and Boughton EM (1995) Evolving neural networks for detecting breast cancer. Cancer Letters; 96(1):49-53.

Fogel DB, Wasson EC Boughton EM and Porto VW (1997) A step toward computer-assisted mammography using evolutionary programming and neural networks, Cancer Letters; 119(1):93-97

Friedman GD (1994) Primer of epidemiology. Fourth edition. McGraw-Hill. New York.

Giuliani A and Benigni R (2000) Principal components analysis for descriptive epidemiology. In Medical Data Analysis. Lecture Notes in Artificial Intelligence. Ed. Brause RW and Hanisch E. Springer, Berlin. Vol.:1933:308-313.

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Publishing Company, New York.

Gray HF. Maxwell RJ. 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):217-24.

Grigsby J, Kooken R and 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, Roβ T, Kruesch J, Wolff and Pöppl SJ (1999) Feature selection for optimised skin tumor recognition using genetic algorithms. Artificial Intelligence in Medicine; 16:283-289.

Haykin, S. S. (1999) Neural networks : a comprehensive foundation. 2nd edition. Prentice Hall International. Upper Saddle River, N.J. and London. 0132733501.

Holmes JH, Durbin DR and Winston FK (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 MW, Rajagopalan, B Data mining to support simulation modelling of patient flow in hospitals. Journal of Medical Systems; 26(2):179-197.

Jefferson, MF, Pendleton, N, Lucas SB, Horan MA (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, MF, Pendleton, N, Mohamed, S, Kirkman, E, Little, RA, Lucas SB, Horan MA (1998a). Prediction of hemorrhagic blood loss with a genetic algorithm neural network. Journal of Applied Physiology; 84: 357-361.

Jefferson, MF, Pendleton, N, Lucas CP, Lucas SB and Horan MA (1998b) Evolution of artificial neural network architecture: prediction of depression after mania. Methods of Information in Medicine; 37:220-225.

Jefferson, MF, Pendleton, N, Lucas SB and Horan MA (1995). Neural networks. Lancet; 346:1712.

Jefferson, M (2001) Outcome prediction in medicine with genetic algorithm neural networks. PhD thesis. University of Manchester.

Johnston ME, Langton KB, Hayes RB and 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 JK (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 HC and Leong SK (2001) Data mining applications in the context of casemix. Annals of the Academy of Medicine; 30(4)41-49.

Kohonen T. (1995) Self-Organizing Maps Springer Verlag. Berlin.

Kononenko I, Bratko I and Kukar M (1998) Application of machine learning in medical diagnosis. In: Machine Learning and data mining. Methods and applications. Edited by Michalsko RS, Bratko I and Kubat M. pp.389-308. John Wiley and Sons Ltd. New York.

Koza JR. (1990a) Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems. Stanford University Computer Science Department technical report. STAN-CS-90-1314.

Koza JR. (1990b) Genetically breeding populations of computer programs to solve problems in artificial intelligence. In Proceedings of the Second International Conference on Tools for AI. Herndon, Virginia, November 6-9, 1990. Los Alamitos, CA: IEEE Computer Society Press. Pages 819-827.

Kuo W-J, Chang R-F, Chen D-R and Lee CC (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 and Kandel A (1999) Applying fuzzy hypothesis testing to medical data. In "New directions in rough sets, data mining, and granular-soft computing". Editors: Zhong N, Skowron A and Ohsuga S. Springer. Berlin.

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. Proceedings of AIMDM'99, Lecture Notes in Artificial Intelligence 1620:47-62.

Lee I-N, Liao S-C and 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 PK and Andersen U (1994) Survival analysis and neural nets. Statistics in Medicine, 13:1189-1200.

Lin F, Chou S, Pan S and 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 4-22.

Lisboa PJG. (2002) A review of evidence of health benefit from artificial neural networks in medical intervention, Neural Networks 15(1):11-39.

Luscombe NM. Greenbaum D. Gerstein M. (2001) What is bioinformatics? A proposed definition and overview of the field. Methods of Information in Medicine. 40(4):346-58.

MacNamee B, Cunningham P, Byrne S and Corrigan OI (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 and Billhardt H (2002) Theory, Abstraction and design in Medical Informatics. Methods of Information in Medicine 41:44-50.

Maojo V and Sanandrés J (2000) A survey of data mining techniques. In Medical Data Analysis. Lecture Notes in Artificial Intelligence. Ed. Brause RW and Hanisch E. Springer, Berlin. Vol.:1933:17-21.

McSherry D (1999) Dynamic and static approaches to clinical data mining. Artificial Intelligence in Medicine; 16:97-115.

Michalski RS, Bratko I and Kubat M. (1997) Machine learning and data mining. Methods and applications. John Wiley and Sons Ltd. New York.

Miller PL. (2000) Opportunities at the intersection of bioinformatics and health informatics: a case study. Journal of the American Medical Informatics Association. 7(5):431-438.

Naranyan MN and Lucas SB (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 PS, Wong ML, Lam W, Leung KS and Cheng JCY (1999) Medical data mining using evolutionary computation, Artificial Intelligence in Medicine. 16(1):73-96.

Ottenbacher KJ, Smith PM, Illig SB, Linn T, Fiedler RC and Granger CV (2001) Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. Journal of Clinical Epidemiology; 54:1159-1165.

Owens DK and Sox HC (2001) Medical Decision-making: probabilistic medical reasoning. In Medical Informatics Computer Applications in Health Care and Biomedicine. Editors Shortliffe EH and Perreault LE. 2nd Edition. Springer. New York. Pp. 76-131.

Papaloukas C, Fotiadis DI, Likas A and Michalis LK (2002) An ischemia detection method based on neural networks. Artificial Intelligence in Medicine; 24:167-178.

Peňa-Reyes CA and Sipper M (1999) A fuzzy-genetic approach to breast cancer diagnosis. Artificial Intelligence in Medicine 17:131-155.

Peňa-Reyes CA and Sipper M (2000) Evolutionary computation in medicine: an overview. Artificial Intelligence in Medicine 19:1-23.

Pendharkar PC, Rodger JA, Yaverbaum GJ, Herman N and 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 TK, Koralewski H-E and Zerbst EW (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 VJ, Sönksen PH, Carey S and Weng C. Data mining for indicators of early mortality in a database of clinical records. Artificial Intelligence in Medicine 22: 215-231.

Saunders P, Mathers J, Parry J and Stevens A (2001) Identifying ‘non-medical’ datasets to monitor community health and well-being. Journal of Public Health Medicine; 23(2):103-108.

Schwarzer G, Vach W and 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 8:37-51.

Shortliffe EH and Blois MS (2001) The Computer meets biology and medicine: Emergence of a Discipline. In Medical Informatics Computer Applications in Health Care and Biomedicine. Editors Shortliffe EH and Perreault LE. 2nd Edition. Springer. New York. Pp. 3-40.

Shortliffe EH and Barnett GO (2001) Medical data: their acquisition, storage. and use. In Medical Informatics Computer Applications in Health Care and Biomedicine. Editors Shortliffe EH and Perreault LE. 2nd Edition. Springer. New York. Pp.41-75.

Sierra B and 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.

Sullivan F, Mitchell E. (1995) Has general practitioner computing made a difference to patient care? A systematic review of published reports. British Medical Journal;311:848-52.

Swanson, DR (1987) Two medical literatures that are logically but not bibliographically connected. Journal of the American Society for Information Science; 38(4):228-233.

Swanson, DR and Smalheiser, NR (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 JV (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 Bemmel JH and Musen MA (1997) Handbook of Medical Informatics. Springer Verlag. Heidelberg, Germany.

Vinterbo S and Ohono-Machado L (2000) A genetic algorithm approach to multi-disorder diagnosis. Artificial Intelligence in Medicine; 18:117-132.

Walker AJ, Cross SS and Harrison RF (1999) Visualisation of biomedical datasets by use of growing cell structure networks: a novel diagnostic classification technique. Lancet; 354:1518-1521.

Wong ML, Leung KS and Cheng JCY (2000) Discovering knowledge from noisy databases using genetic programming. Journal of the American Society for Information Science 51(9):870-881.

Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA and Metz CE (1993) Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology; 187(1):81-87.

Wyatt JC and Altman DG. (1995) Commentary: prognostic models; clinically useful or quickly forgotten? British Medical Journal;311:1539-1541.

Xiang A, Lapuertab P, Ryutova A, Buckleya J and 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 and Dong J (2002) Mining interesting rules in meningitis data by cooperatively using GDT-RS and RSBR. In Lecture Notes in Artificial Intelligence; 2336:405-416. Springer-Verlag. Berlin.

Zupan B, Demšar J, Kattan MW, Beck JR and Bratko I (1999) Machine learning for survival analysis: a case study on recurrence of prostate cancer. In Lecture Notes in Artificial Intelligence; 1620:346-355.