ORIGINAL RESEARCH

Prediction of clinical behaviour and treatment for cancers

Matthias E Futschik, Mike Sullivan, Anthony Reeve, Nikola Kasabov

Departments of Information Science and Biochemistry, University of Otago, Dunedin, New Zealand

Abstract: Prediction of clinical behaviour and treatment for cancers is based on the integration of clinical and pathological parameters. Recent reports have demonstrated that gene expression profiling provides a powerful new approach for determining disease outcome. If clinical and microarray data each contain independent information then it should be possible to combine these datasets to gain more accurate prognostic information. Here, we have used existing clinical information and microarray data to generate a combined prognostic model for outcome prediction for diffuse large B-cell lymphoma (DLBCL). A prediction accuracy of 87.5% was achieved. This constitutes a significant improvement compared to the previously most accurate prognostic model with an accuracy of 77.6%. The model introduced here may be generally applicable to the combination of various types of molecular and clinical data for improving medical decision support systems and individualising patient care.

Keywords: treatment outcome prediction, data integration, microarray, lymphoma

Introduction

Successful cancer therapy demands accurate diagnosis of the tumour type. Conventionally, the diagnosis of cancer has been based on the assessment of histological and pathological features of tissue biopsies from the patient. This method, however, depends strongly on the expertise and training of the pathologist examining the tissue specimen, so the final diagnosis may be subjective. Alternative classification procedures are, therefore, highly desirable.

The recently introduced cDNA and oligonucleotide microarrays might possibly offer such alternatives. Simultaneously measuring the expression of thousands of genes, microarrays can give detailed pictures of the molecular state of cells. Recently, microarray data have been used for outcome prediction of cancer treatments (Pomeroy et al 2002; Shipp et al 2002; van't Veer et al 2002). Although the studies showed the ability to predict disease outcome based on gene expression data, they also revealed a limited accuracy in the derived predictions. This may not be surprising, since microarrays are restricted to measuring RNA abundances. In fact, disease prognosis is a challenging task, as many factors have to be taken into account. For example, the genetic background, environment, age, the physical condition of the patient and the intensity and duration of therapy have to be considered, as well as the type, location and stage of the tumour. To exploit the full power of microarray techniques for medical applications, it is therefore necessary to integrate the data from microarrays with various clinical parameters.

We propose in this study the integration of microarray data and clinical variables using a modular hierarchical model. Separate modules are constructed for microarray and clinical data. The microarray predictor module was formed by a neural network classifier. For the clinical predictor, we converted an existing clinical prognostic model to a Bayesian classifier. The predictions of the two independent modules were combined and fused to a single prediction.

Combination of classifiers has recently attracted increased attention. Reviews of this growing area of research can be found in Kittler (1998) and Kittler et al (1998). Several models of combined classifiers have been proposed for the analysis of DNA and protein sequences (Zhang et al 1992; Xu et al 1996). To our knowledge, however, no other study has focused on the combination of clinical and microarray-based classifiers.

Data

Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoid malignancy in adults. The treatment of DLBCL usually begins with multi-agent chemotherapy and, in the case of a relapse, is often followed up by bone marrow

Correspondence: Matthias E Futschik, Departments of Information Science and Biochemistry, University of Otago, PO Box 56, Dunedin, New Zealand; tel +64 3 479 8316; fax +64 3 479 8311; email mfutschik@infoscience.otago.ac.nz

transplantation. The clinical course of chemotherapy for DLBCL is widely variable. Although most of the patients initially respond well to the treatment, the majority finally succumb to the disease. Only 40% of patients achieve a durable remission of the cancer.

Using Affymetrix chip technology, Shipp and co-workers profiled 58 DLBCL tissue samples from patients with a known chemotherapy outcome ('cured' or 'fatal') and were able to construct predictors for the outcome based on selected sets of differentially expressed genes (Shipp et al 2002). Three supervised learning methods were applied: a weighted voting algorithm (Golub et al 1999), support vector machines (SVMs) and *k*-nearest neighbour (knn) classifiers. An accuracy of 75.9% was achieved by the weighted voting algorithm, 77.6% by SVMs and 70.7% by knn classifiers. The publicly available dataset includes further clinical information in the form of the International Prediction Index (IPI) for 56 patients.

Methods

We first introduce the definition of mutual information, since it gives an indication about the ability to improve prediction by the combining of two separate prognostic models. We then describe the neural network classifier for classification of the treatment outcome based on microarray data and the use of IPI as the Bayesian classifier. Both classifiers are used as prediction modules in our modular hierarchical model.

Mutual information

The dependence of two random variables x and y with probability distributions $P(x)$ and $Q(y)$ can be measured by the 'mutual information' I. It is defined as the relative entropy between the joint distribution $R(x,y)$ and the product distribution $P(x)Q(y)$:

$$I(x,y) = H(P) + H(Q) - H(R) = H(P) - H(P|Q)$$
$$= \Sigma_{x,y} R(x,y) \log_2 (R(x,y) / [P(x)Q(y)]) \quad (1)$$

where $H(P)$ is the entropy of distribution P, $H(Q)$ is the entropy of Q, $H(R)$ is the entropy of R and $H(P|Q)$ is the entropy of P given Q, respectively. The entropy H is here defined as Shannon entropy ie $H(P) = \Sigma_x P(x) \log_2 (P(x))$.

The mutual information I indicates the reduction of uncertainty about one variable given the other. If the variables are statistically independent, the joint probability $R(x,y)$ is equal to the product probability $P(x)Q(y)$. In this case, the mutual information is minimal ($I = 0$). This also holds in the reverse direction; if the mutual information is zero, the variables are statistically independent (Cover and Thomas 1991).

IPI as Bayesian classifier

The hitherto standard model for the outcome prognosis for DLBCL, the International Prediction Index (IPI), uses clinical data to derive its prediction (Shipp et al 1993). The IPI is based on a clinical study of patients, who were treated using a combination therapy. A step-down regression was used to select several risk factors that remained independently significant for overall survival: age, tumour stage, number of extranodal sites, performance status and serum lactate dehydrogenase concentration (LDH). Assessing these risk factors, patients can be assigned to four different risk groups with different five-year survival rates: 73% for 'low' risk group with zero to one risk factor present; 51% for 'low–intermediate' risk group with two risk factors present; 42% for 'intermediate–high' risk group with three risk factors present; and 26% for 'high' risk group with four or all risk factors present. The IPI is currently used to stratify patients for intensified therapy.

To use the IPI as a predictor, we have to formulate the decision rule for the classification procedure. Given the IPI value as a single feature and the two classes 'cured' and 'fatal', we state the conditional probabilities $P(\text{class} = i \,|\, \text{IPI} = j)$ of the class i (eg 'cured') given the IPI risk group j (eg 'low') based on the observed samples if we want to classify a new sample. Using these class probabilities, we can apply the Bayesian decision rule:

Decide for class 'cured'
if $P(\text{'cured'} \,|\, \text{IPI}) > P(\text{'fatal'} \,|\, \text{IPI})$
otherwise decide for class 'fatal'.

This rule simply means that we classify an unseen sample based on the majority of observed samples in the corresponding IPI category. Using this rule, we minimise the overall Bayesian risk (giving equal weight to all misclassifications) and convert the IPI to a Bayesian classifier.

Evolving fuzzy neural networks

For the microarray-based predictor, we used evolving fuzzy neural network (EFuNN) classifiers recently introduced for adaptive supervised learning (Kasabov 2001). EFuNNs use a five-layer structure where nodes and connections are created as data examples are presented (Figure 1). The first layer serves for fuzzification of the input, so that the activation values in the second layer are the fuzzy

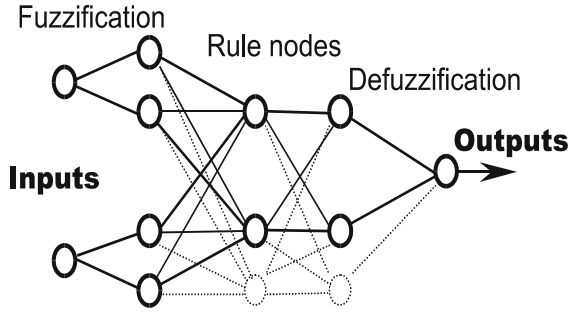

Figure 1 Outline of a simplified five-layered EFuNN (evolving fuzzy neural network) structure.

representation of the input. Nodes in the third layer represent clusters of examples in the training set. By adding new nodes or adjusting existing nodes, homogeneous clusters can be formed; ie all samples in a cluster belong to one class. This is implemented in the EFuNN algorithm by unsupervised learning of the connection between the second and third layer, and supervised learning of the connections between the third and fourth layer, respectively.

Examples will be classified according to their distance from the nearest rule nodes. The fourth layer of neurons performs defuzzification, so that the fifth layer represents the real values for the output variables.

Results

After constructing the microarray-based EFuNN classifier, we tested the hypothesis that clinical and microarray data contain independent information by applying set and information theory. Both types of data are then combined in a modular hierarchical model.

Training of the microarray-based classifier

To construct a microarray-based classifier, we selected and ordered genes according to their signal-to-noise ratio, as defined by:

$$S=(X_1-X_2)/(\sigma_1+\sigma_2) \quad (2)$$

with $X_{1,2}$ and $\sigma_{1,2}$ as the mean and the standard deviation of the expression values of a gene in class 1 ('cured') or 2 ('fatal'). The training of EFuNNs was performed as follows: for different EFuNN parameter settings, we started with a low number ($n=3$) of genes as network input; gradually increasing the number of genes generally led to an increase in classification accuracy; we continued to add genes to the network input until a maximum accuracy was achieved; further increasing the number of genes yielded a decrease in accuracy, as additional genes contributed noise to the classification process.

The classification performance was recorded applying the same leave-one-out method as used by Shipp and colleagues (Shipp et al 2002). For the 56 samples for which the IPI classification was present, our best EFuNN classifier achieved a maximum accuracy of 78.5% for 17 genes as input. The clinical prognostic model described above achieved an accuracy of 73.2% on the same dataset.

Complementary information of two classifiers

An improvement of the outcome prediction by combining the two classifiers constructed can only be achieved if their predictions are at least partially complementary. Basic set theory showed that the predictions agree in only 37 of 56 cases (Figure 2). This indicates that the predictions are independent.

The independence of two classifiers can also be assessed by examining the mutual information. To use equation (1), we first calculate the probabilities P, Q and R, with variable x as the EFuNN prediction and y as the IPI-based predictions. The variables x and y are defined here as 1 if the classifier predicts 'cured' and 0 if the classifier predicts 'fatal', respectively. For the EFuNN classifier, this yielded $P(x=1)=0.55$ and $P(x=0)=0.45$. For the IPI-based classifier, this yielded $Q(y=1)=0.66$ and $Q(y=0)=0.34$. The joint probabilities were as follows: $R(x=1, y=1)=0.42$, $R(x=1, y=0)=0.13$, $R(x=0, y=1)=0.23$ and $R(x=0, y=0)=0.21$.

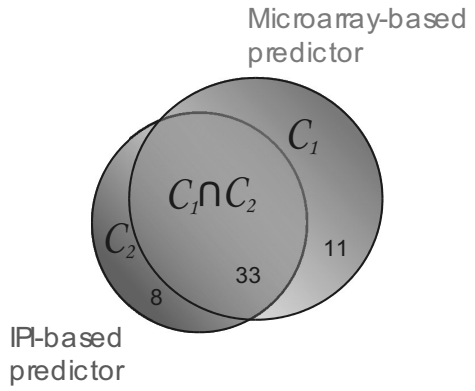

Figure 2 Venn diagram of overlapping sets of samples correctly predicted by the microarray-based predictor C_1 and by the IPI-based clinical predictor C_2. The predictions are complementary in 19 of 56 cases: 44 samples are correctly classified by the microarray-based predictor, while 41 samples are correctly classified by the clinical predictor. Altogether, 33 samples are classified correctly by both predictors. Of the remaining samples, 11 are classified correctly only by the microarray-based predictor (and not by the clinical predictor) and 8 are classified correctly only by the clinical predictor (and not by the microarray-based predictor). Overall, 52 samples are predicted correctly by at least one predictor, setting an upper threshold (92.9%) for the accuracy of the combined model.

Based on these calculated probabilities, and using equation (1), a mutual information of 0.05 was derived for the predictions by the EFuNN and IPI-based classifiers. Since the mutual information is almost zero, the microarray- and IPI-based prognostic models can be considered as statistically independent classifiers (see Mutual Information section). This is a rather surprising result, as several risk factors are considered surrogates of underlying molecular mechanisms (Shipp et al 1993) and should therefore be highly correlated with gene expression in the tumour.

Hierarchical modular system

To exploit the complementary nature of the two prognostic models, we combined both predictors in a hierarchical modular prognostic model (Figure 3). It consists of an EFuNN and an IPI predictor in the first layer. For each sample, these modules predict independently whether the sample belongs to either tissue class 'cured' or tissue class 'fatal'. Each of these predictions has a defined strength representing the confidence of the predictor in its decision.

In the case of EFuNNs, the prediction strengths derive from the activation values of the rule nodes. The nearer a sample lies to a node belonging to a class, the higher is the activation value of this node and thus the prediction strength for the class. In brief, the EFuNN prediction is based on the location of the sample in the gene space.

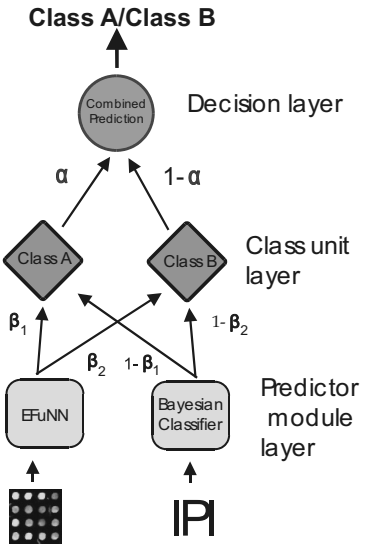

Figure 3 Three-layered hierarchical model for combination of clinical and microarray-based predictors: the first layer ('predictor module layer') consists of independently trained predictor modules; the second layer ('class unit layer') integrates the weighted predictions of both predictors for classes A and B ('cured' and 'fatal'); the third layer ('decision layer') produces the final prediction based on the weighted sum of the outputs of class units.

The strength of each IPI prediction is the corresponding conditional class probability and represents the homogeneity of the IPI category. As in the case of the EFuNN module, the prediction strengths for both classes will propagate to the second layer.

The 'class unit layer' combines the prediction of the two modules and consists of units for the classes 'cured' and 'fatal'. Predictions of the EFuNN module were weighted by a factor β_1, predictions of the IPI module were weighted by a factor $(1 - \beta_1)$. The factor β_1 indicates how much we rely, for the class 'cured', on the predictions of the EFuNN module compared to the IPI module. The weighting for the class unit 'fatal' is performed independently (β_2), since microarray-based and IPI-based predictor modules may have different accuracies for each class.

Outputs of the class units were weighted with factor α and $(1-\alpha)$ before being fed into the decision unit to adjust for a possible bias towards one class. The outcome prediction for the sample is produced in the 'decision layer' by comparing the outputs of the class units.

Different methods can be used for the optimisation of the model parameters α, β_1 and β_2 such as error backpropagation and expectation-maximisation with constraints. For this study, however, it seems more informative to achieve a global picture of the dependence of the performance on the weighting parameter. We therefore present here the results using an exhaustive grid search (for α, β_1, β_2 in [0,1]) over all possible combinations of the parameters within a leave-one-out procedure. The training of the microarray-based EFuNN predictor is performed in parallel. The IPI-based Bayesian classifier is static in this study, since it derives from a previously determined prognostic model.

The maximal accuracy of the hierarchical modular model was 87.5%, which is significantly higher than the accuracies achieved by the single predictor modules. The dependence of the accuracy on model parameters is visualised in Figure 4. If we are weighting the outputs of the class units equally, the accuracy reaches a maximum for $\beta_1, \beta_2 \approx 0.7$. This means, to improve the accuracy by combination, more weight has to be given to the microarray-based predictions. However, it also demonstrates that the incorporation of clinical information helps to improve the accuracy.

Areas of expertise

Regions in the domain space for which a classifier achieves a very high accuracy are frequently called 'areas of expertise'. For an optimal combination of classifiers, it is

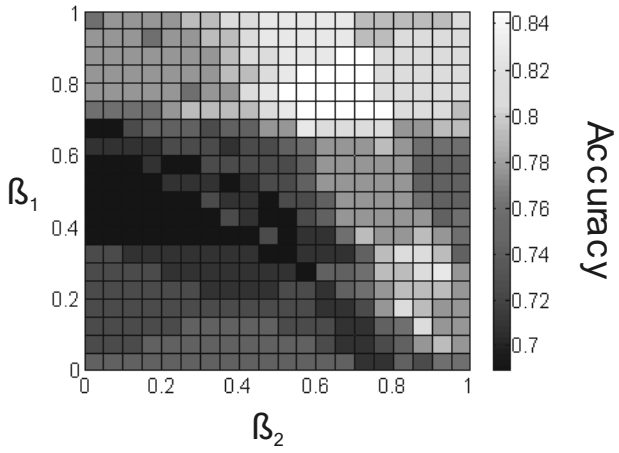

Figure 4 Accuracy of prediction with respect to weight parameters β_1 and β_2 for $\alpha = 0.5$. Maximum accuracy is achieved for parameter settings of $\beta_1, \beta_2 \approx 0.7$. Destructive interference is observed for $\beta_1, \beta_2 \approx 0.5$.

essential to find these areas for each of the classifiers under consideration.

Stratification of the samples into clinical subgroups according to their IPI value demonstrated that the two predictor modules differed in their areas of expertise. The IPI risk groups 'low', 'low–intermediate' and 'intermediate–high' were weighted towards the microarray-based predictor, while the 'high' risk group was weighted towards the clinical predictor for the final outcome prediction of the combined model. Interestingly, the SVM used by Shipp et al (2002), as well as our neural network method, classified the samples with IPI 'high' incorrectly. This indicates that the molecular basis of DLBCL for this group of patients might differ from that of patients in other IPI risk groups.

Discussion and conclusions

The outcome of complex diseases, such as cancer, is likely to depend on various factors. For accurate outcome prediction, information about the patient as well as the molecular state of the disease may have to be considered. Microarrays have offered comprehensive new views of many diseases; however, they are limited to detecting changes in the abundance of mRNAs in cells. Almost certainly, medical decision support systems will require wider approaches incorporating different types of patient and disease-related information.

Our study demonstrates that the integration of microarray data with previously established clinical parameters can considerably improve disease outcome prediction. Calculating the mutual information as a measure of statistical independence, we could show that the EFuNN predictor trained on microarray data is statistically independent from the IPI predictor. This independence enabled us to improve overall prediction accuracy by a combined modular prediction system, which can be regarded as a first step towards full integration of microarray data into clinical decision systems.

The modular system is not restricted to the data analysed here, but is generic. The inclusion of other types of clinical and molecular data as well as predictor modules is possible and may be favourable for individualising patient care.

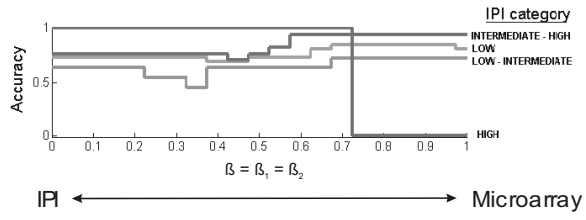

Figure 5 Accuracy of prediction by the combined prognostic model with respect to weighting parameters β_1, β_2 (for $\alpha = 0.5$). Stratification into the different IPI categories shows the dominance of the predictor modules for the final decision of the combined model. For samples with IPI 'low' and 'high', the incorporation of clinical information is beneficial.

Acknowledgements

Matthias Futschik was supported by a PhD scholarship and bridging grant from the University of Otago. We would like to thank the reviewers for their constructive comments and Bronwyn Carlisle for proofreading. Commercial use of the techniques described and applied in this paper is subject to obtaining a licence from Pacific Edge Biotechnology Ltd (http://www.peblnz.com).

References

Cover TM, Thomas JA. 1991. Elements of information theory. New York: J Wiley.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA et al. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286:531–7.

Kasabov N. 2001. On-line learning, reasoning, rule extraction and aggregation in locally optimised evolving fuzzy neural networks. *Neurocomputing*, 41:25–45.

Kittler J. 1998. Combining classifiers: a theoretical framework. *Pattern Anal Appl*, 1:18–27.

Kittler J, Hatef M, Duin RPW, Matas J. 1998. On combining classifiers. *IEEE Trans Pattern Anal Machine Intell*, 20:226–39.

Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C et al. 2002. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415:436–42.

Shipp MA, Harrington DP, Anderson JR, Armitage JO, Bonadonna G, Brittinger G, Cabanillas F, Canellos GP, Coiffier B, Connors JM et al. 1993. A predictive model for aggressive non-Hodgkin's lymphoma, The International Non-Hodgkin's Lymphoma Prognostic Project. *N Engl J Med*, 329:987–94.

Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, Gaasenbeek M, Angelo M, Reich M, Pinkus GS et al. 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med*, 8(1):68–74.

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–6.

Xu Y, Mural RJ, Einstein JR, Shah MB, Uberbacher EC. 1996. Grail: a multi-agent neural network system for gene identification. *Proc IEEE*, 84:1544–52.

Zhang X, Mesirov JP, Waltz DL. 1992. Hybrid system for protein secondary structure prediction. *J Mol Biol*, 225:1049–63.