BRIEFINGS IN BIOINFORMATICS. VOL 10. NO 3. 315–329
Advance Access publication March 23, 2009
doi:10.1093/bib/bbp012

An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies

Lee J. Lancashire, Christophe Lemetre and Graham R. Ball

Submitted: 14th November 2008; Received (in revised form): 13th February 2009

Abstract

Applications of genomic and proteomic technologies have seen a major increase, resulting in an explosion in the amount of highly dimensional and complex data being generated. Subsequently this has increased the effort by the bioinformatics community to develop novel computational approaches that allow for meaningful information to be extracted. This information must be of biological relevance and thus correlate to disease phenotypes of interest. Artificial neural networks are a form of machine learning from the field of artificial intelligence with proven pattern recognition capabilities and have been utilized in many areas of bioinformatics. This is due to their ability to cope with highly dimensional complex datasets such as those developed by protein mass spectrometry and DNA microarray experiments. As such, neural networks have been applied to problems such as disease classification and identification of biomarkers. This review introduces and describes the concepts related to neural networks, the advantages and caveats to their use, examples of their applications in mass spectrometry and microarray research (with a particular focus on cancer studies), and illustrations from recent literature showing where neural networks have performed well in comparison to other machine learning methods. This should form the necessary background knowledge and information enabling researchers with an interest in these methodologies, but not necessarily from a machine learning background, to apply the concepts to their own datasets, thus maximizing the information gain from these complex biological systems.

Keywords: *artificial neural networks; bioinformatics; genomics; mass spectrometry; microarray; proteomics*

INTRODUCTION

The intention of this review is to provide researchers with an understanding of the potential benefits of using artificial neural network (ANN)-based approaches within a biomedical context. They may be applied for classification, predictive modelling and biomarker identification within data sets of high complexity. The focus within this review is on transcript or gene expression data generated from DNA microarray (MA) analysis, or peptide/protein level data generated by mass spectrometry (MS). In 'Artificial neural networks' section the concepts behind ANN learning will be introduced and described detailing their advantages and disadvantages. This will include details on how robust models are generated, tested and validated using suitable

The first two authors contributed equally to this work.
Corresponding author. Lee J. Lancashire, Clinical and Experimental Pharmacology, Paterson Institute for Cancer Research, University of Manchester, Manchester M20 4BX, UK. Tel: +44-16-1446-3172; Fax: +44-16-1446-3109; E-mail: llancashire@picr.man.ac.uk

Lee J. Lancashire is a postdoctoral researcher at the Paterson Institute for Cancer Research. He received a PhD in bioinformatics from the Nottingham Trent University in 2006. His main research areas are in machine learning and bioinformatics.

Christophe Lemetre has a degree in computer sciences and bioinformatics and is currently a PhD student under the supervision of Dr Graham Ball, working on machine learning algorithms in bioinformatics.

Graham R. Ball is a reader in bioinformatics at the Nottingham Trent University. He has 16 years experience in the development of biological applications for Artificial Neural Network algorithms.

cross validation approaches. In 'Regularization' section the reader will be made aware of techniques that must be applied during the modelling process in order to obtain reliable results, a principal consideration in highly dimensional datasets. In 'Experimental methods requiring robust bioinformatics' section the MS and MA technologies will be outlined. In 'Data complications in proteomics and genomics' section, issues of high dimensional input data and the importance of reproducibility will be examined. In 'Recent applications' section, examples of publications detailing how ANNs are currently being used in genomic MA and proteomic MS studies will be summarized. 'Comparison to other machine learning methods' section provides highlighted case studies where ANNs have performed favourably in comparison to other common statistical and machine learning methodologies. 'Future trends' section briefly outlines the advanced steps necessary once a validated ANN biomarker signature has been discovered. 'Conclusions' will sum up the review. Researchers with an interest in the potential benefits that ANN approaches may bring to their laboratories should then be able to apply them to their own datasets, maximizing the information to be gained from the analysis of complex biological systems.

Background

There are a number of steps required in order to identify and validate a biomarker so that it can be used in a clinical setting [1], and despite the increasing use of high-throughput technologies such as MS and gene MAs, there remains a lack of clinically useful biomarkers emerging for diseases such as cancer. There may be several reasons for this, such as the reported lack of reproducibility of these approaches [2–5], and the sheer mass of data being generated, which is often extremely noisy, and is becoming progressively complex. This is particularly true in the field of '-omics', where for example, in the recent Affymetrix GeneChip 1.0 ST MAs (designed to target all known and predicted exons in human, mouse and rat genomes), each individual case studied contains information for approximately 1.2 million exon clusters corresponding to over 1.4 million probesets. Thus teasing out the key components from these datasets requires the use of mathematical models running on hardware capable of efficient analyses. The discovery of new biomarkers could facilitate more reliable, efficient and less subjective methods to assist the human expert in the diagnosis of disease, as well as providing new potential targets for future therapies.

With this in mind, it is clear that the identification of new biomarkers still requires a concerted, multidisciplinary effort. This necessitates the requirement for specific computational tools for data-mining, and as such remains a major challenge in bioinformatics [6]. One such tool are ANNs [7], a form of machine learning from the field of artificial intelligence utilized in many areas of bioinformatics and medicine [8] due to their ability to cope with noisy, non-linear and highly dimensional datasets, in particularly when appropriate regularization strategies are employed and when combined with appropriate feature reduction methodologies or forward selection methods such as that proposed in [9]. Using ANNs, it is possible to analyse these sophisticated datasets in identifying novel gene or protein signatures (biomarkers or fingerprints) of biological systems in an endeavour to identify specific phenotypes for diagnosis of disease, establishing a patient's clinical outcome, or even predicting a patient's response to therapy.

ARTIFICIAL NEURAL NETWORKS

ANNs are inspired by the way in which the human brain learns and processes information, with the ability to handle complex (non-linear) features within data in order to generalise and predict well for future cases. Their concept simulates the behaviour of a biological neural network; in humans, learning involves minor adjustments to the synaptic connections between neurons, in ANNs, the learning process is based on the interconnections between the processing elements that constitute the network topology.

McCulloch and Pitts first described the concept of the artificial neuron in 1943 as a mathematical function derived from simulating the basic functions of biological neurons [10]. This manuscript will focus on ANNs in their most common form, the multilayer perceptron (MLP), but other ANN-based approaches exist; for example radial basis function networks and recurrent neural networks. In the MLP, ANNs are organized into several layers, with each layer having a number of respective neurons, or processing elements, that constitute that layer (Figure 1). Simply put, the majority of ANNs have a similar topology consisting of an input layer, one or more hidden layers and an output layer. The number

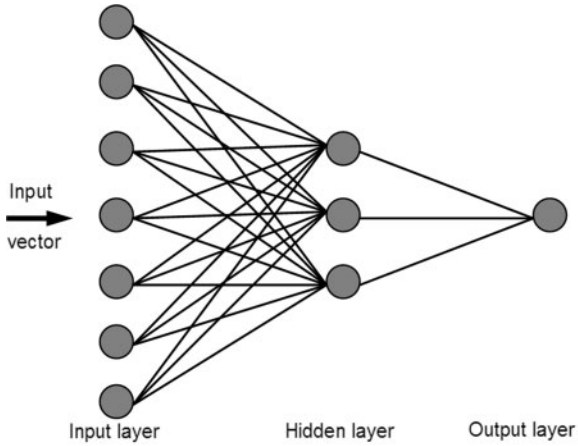

Figure 1: Architecture of a typical multi-layered perceptron artifical neural network.

of hidden layers and the number of neurons in each layer is dependent on the complexity of the problem, i.e. the number of input neurons. The input layer interacts with the external environment to receive the input data as a vector of predictor variables, each represented as a node. This information is passed through to the first hidden layer, and multiplied (thus modified) by a set of associated weights. These products are summed and fed through a non-linear transfer function (e.g. sigmoid, hyperbolic tangent) which scales and then produces an output, similar to the axon of the neuron. The calculation of the output for each neuron is then as follows:

$$v_k = \sum_{i=1}^{n} w_{ki} x_i$$

and

$$y_k = \Phi(v_k + v_{k_0})$$

where $x_1, x_2 \ldots x_n$ are the input signals converging to neuron k. $\omega_{k1}, \omega_{k2} \ldots \omega_{kn}$ are the weights connecting neuron k. v_k is the net input. y_k is the output of the neuron where v_{k0} is a bias term and $\phi(.)$ is the activation function commonly of the form:

$$\Phi(v) = \frac{1}{1 + e^{-v}}$$

for the sigmoid activation function and:

$$\Phi(v) = \frac{e^{v} - e^{-v}}{e^{v} + e^{-v}}$$

for the hyperbolic tangent activation function.

Ultimately this modified information reaches the node(s) in the output layer, the result of which is the output of the entire ANN, for example the predicted class for a given case, or a continuous numerical output in a regression model. In a two group classification problem, the output in the training examples is usually represented as 0 and 1, or −1 and 1. The interconnecting weights are crucial to the system and also enable a variable strength to be given to each input variable included in the model, whether it is excitatory or inhibitory.

ANN learning

ANNs must be trained to efficiently compute the gradient as to be capable of accurately modelling a set of cases and predicting their output. There are two major learning paradigms; supervised and unsupervised. Supervised learning involves providing the network with a set of cases that have values for the inputs as well as the known desired outputs. The output of the network is then compared with the true output to calculate error by assessing the network performance as learning progresses. The interconnecting weights are initially randomized (e.g. [−1, 1]) so that predictions after completion of the first training cycle are essentially random. One of the most popular forms of supervised learning is to compare the error between the true output and the predicted output and then feed this error back through the layers of the network. The weights are adjusted so that after completion of the next training cycle (or epoch) the error decreases according to:

$$\omega_{ki}(\tau) = \eta \delta_k x_i$$

Each weight update ω_{ki} at the current (τ^{th}) cycle is updated in proportion to the input value to which the weight is applied x_i, the error in the output of the unit δ_k and constants known as the learning rate η [11]. The weight change of a neuron is proportional to the influence an input had on the error during training and the learning rate is a constant which controls the size of these weight changes. The larger the learning rate, the faster learning will proceed; however too large a value may lead to non-convergence of the model. Each time a pattern is presented to the network, the weights leading to an output node are modified slightly during learning in the direction required to result in a smaller error the next time the same pattern is presented, until a target error is reached or no improvement of the error is observed. The larger the learning rate, the larger the weight changes and the faster the learning will proceed. If the learning rate is

too small, training will be slowed down, however, oscillation or non-convergence can occur if the learning rate is too large [12]. A momentum term, α, may be applied to help prevent the network becoming trapped in local minima, or being stuck along flat regions in error space. This occurs with a slight alteration to the weight update rule by making the weight update on the τ^{th} iteration depend on the update that occurred during the $(\tau - 1^{\text{th}})$ iteration:

$$\omega_{ki}(\tau) = \eta\delta_k x_i + \alpha\omega_{ki}(\tau - 1)$$

This helps to speed up the time it takes for the network to reach convergence by gradually increasing the step size of the search in regions where the gradient is not changing. As with the learning rate, effectively choosing values for these constants depends on the particular problem of interest and experimentation is important here to find optimal values. In our own experiences for MA and MS data, a learning rate of 0.1 combined with a momentum of 0.5 has proved successful [9, 13]. The target error that needs to be minimized is often determined as the total sum-of-squares based on the difference between the output and target vector as follows:

$$\varepsilon = \frac{1}{2}\sum_{j=1}^{n}(d_j - y_j)^2$$

where n is the number of cases, d_j is the target network output for case $_j$ and y_j is the network predicted output for case j. Alternative error functions also exist, such as the mean squared error, or the maximum conditional likelihood fitting, but will not be dwelt upon here. This learning process is an extension of the generalized delta rule, and is commonly known as back-propagation [14–16]. It is crucially important that the data used in training the network should be reasonably large in order to contain all the information necessary to be able to recognize which of the predictor variables are important amongst the vast amounts of noise and individual variation that is expected to cloud important information in complex '-omics' datasets. If the network outputs fail to show good discrimination on an independent test dataset, over-fitting may have occurred and training must be continued or repeated. Over-fitting can occur when the number of parameters in a model exceeds the number of cases. It is in essence a memorization of the training data (and any associated random noise) [17, 18]. In order for the network to be trained to a satisfactory level which maintains generalization for new data, it is vital to employ an appropriate regularization technique (discussed later in the review). Once learning is complete the weights are stored and can be used to predict future cases in separate test datasets. Other learning algorithms have also been proposed. These include (but are not limited to) QuickProp [19], RPROP [20] and the Levenberg–Marquardt algorithm [21, 22].

Unsupervised learning occurs when the network attempts to map the inputs to outputs without any external assistance. Therefore the network itself governs how it groups the cases based upon the input data. This is sometimes referred to as self organization, and Kohonen's self organizing maps [23] are the most popular form of neural network-based unsupervised learning. Other forms of unsupervised learning include principal components analysis, independent components analysis, hebbian learning and autoassociators. Although unsupervised learning algorithms are an active area of research, it is beyond the scope of this review to explain and review their application in detail and consequently this manuscript will focus on the use of supervised neural networks. For a more detailed discussion on unsupervised pattern recognition in high-throughput genomics and proteomics see [24].

Advantages and disadvantages of artificial neural networks

As ANNs are loosely based on the way a biological neuron is believed to organize and process information, they have many advantages in their ability to derive meaning from large complex datasets. First, they do not rely on data to be normally distributed, an assumption of classical parametric analysis methods. They are able to process data containing complex (non-linear) relationships and interactions that are often too difficult or complex to interpret by conventional linear methods. Another advantage is that they are fault tolerant, i.e. they have the ability of handling noisy or fuzzy information, whilst also being able to endure data which is incomplete or contains missing values. In addition to this (like other machine learning methods), they are capable of generalization, so they can interpret information which is different to that of the training data, thus representing a 'real-world' solution to a given problem by their ability to predict future cases or trends based on what they have previously seen. Thus, trained ANNs can be used as standalone executable systems in order to predict the class of an

unknown case of interest, and therefore have the potential application in diagnosis. Finally, there are several techniques that can be used to extract knowledge from trained ANNs, and the importance of individual variables can be easily recovered using various methods such as the analysis of interconnecting network weights [25], sensitivity analysis and rule extraction [26]. This, from a biological perspective, is perhaps one of the most useful aspects of ANN modelling. Gevrey *et al*. [27] review this subject in more depth.

Like all approaches, ANNs also have their limitations. Training of ANNs can potentially be time consuming depending on the complexity of the data being modelled, and as the number of hidden layers required to capture the features of the data increases, so does the time taken for training to complete. As such, only one or two hidden layers are commonly used. Over-fitting may be a problem in ANNs, which is a memorization of the training cases causing the network to perform poorly on future cases. The one major barrier which researchers usually associate with ANNs is that it is not always apparent how they reach a solution, and because of this they have been referred to as 'black boxes' [28–31].

Further limitations originate from the data itself. Experimental data may suffer from high background variation that is difficult for computational algorithms to interpret. The challenges in terms of reproducibility of some technologies has also been investigated [18, 32–39], rendering validation with a separate cohort of samples virtually impossible. The old adage 'garbage in, garbage out' can be strongly applied to modelling with ANNs, and thus the quality of the model output is highly dependent upon the quality of the input data. If the input data is not representative of the 'real world' scenario, the model is compromised. To overcome these issues, several techniques for pre-processing the data have been proposed, and the reader is referred to [40–45] for more examples, and for a guide to considerations regarding study design see [2].

Implementing artificial neural networks

Implementing ANNs is usually performed with statistical computer software packages, or open source equivalents in R (http://www.r-project.org/index.html) and Weka (http://www.cs.waikato.ac.nz/ml/weka/). A comprehensive list of ANN software packages can be found at ftp://ftp.sas.com/pub/neural/FAQ6.html#questions.

REGULARIZATION

Commonly the main purpose of modelling is to simulate a real world system and therefore a model is judged on its ability to generalize to new data. In ANNs the risk of low generalization is mainly attributed to over-training of the model, leading to over-fitting and subsequently poor predictive performance during independent validation. Due to the fact that even a linear model would over-fit in high dimensions, ANNs must be appropriately regularized during training in order to achieve sufficiently high predictive performances. In order to address this, regularization techniques need to be applied during training. Several options for regularization exist and methods can be chosen according to the type of data or generalization performance that is required. This section will now briefly introduce some of the most common forms.

Weight decay

One of the simplest regularization methods to implement is weight decay. In weight decay, the error function includes a penalty term, for example the sum of squared weights and biases multiplied by a decay constant that controls how much the penalty term should affect the resulting error function. Since over-fitted models are more likely to contain unusually large weights, this approach aims at penalizing such large weights, in order to keep weight values smaller than they naturally otherwise would converge at, thus keeping the activation of the neurons in the linear range [7].

Resampling and early stopping

According to Ntzani and Ioannidis [46], independent validation is only conducted in about 10% of MA studies published. Given the fact that these complex datasets are likely to be non-linear in nature, one may not have prior information regarding the intricacies of the data. As such it is vital to estimate the performance of these models on new data in order to be confident that over-fitting has been avoided. It was stated earlier that the back-propagation algorithm should stop training once the network has achieved an acceptable level, however, the question remains as to what is considered to be an acceptable level, and what can be done to ensure that

the model will be capable of generalizing to additional future cases. If training is terminated solely on the basis of a set number of iterations the model is at risk of over-fitting. The most universal approach to address this problem is resampling. Typically in ANN-resampling approaches, the data is split into different subsets, where a percentage of the total sample set is used to train and optimize the ANN (the training set) and (sample size permitting) the remaining are partitioned for validation during training (the validation set) and external testing after the modelling is complete (the test set). A crude regularization technique known as the early stopping mechanism monitors the network error with respect to a validation or test dataset. This process signals to stop training either when a predetermined number of iterations have completed, or when the prediction accuracy of the model begins to worsen for the validation or test dataset, a sign of over-fitting. The weights resulting in the minimum validation or test error are then selected. Once the network has completed the learning process, it is further validated using the test data split, to give an unbiased estimation of the networks likely performance on future unseen cases. Examples of this approach can be found in [8] and [12].

Bayesian regularization

The Bayesian regularization approach involves modifying the target function (such as the sum of squared errors) in order to improve the models generalization ability. The Bayesian regularization aims to smooth the cost function by adding to it a regularization parameter based on the sum of squared weights. To reduce bias, the weights and variables of the network are assumed to follow a Gaussian distribution and are assigned prior probabilities, optimized according to the Bayesian framework of Mackay [47]. Network training then attempts to find the trade-off between minimizing the model complexity and model error, as such minimizing both the bias and variance [48]. Methods such as automatic relevance determination will identify and remove unnecessary parameters from the model since the Bayesian approach provides an estimate for the entire distribution of model parameters rather than a single optimal set of weights. Model comparison is based on highest evidence, rather than cross validation, and as such Bayesian regularization maximizes the data available as it does not require a validation set since all the training data can be used for model fitting. A review of Bayesian methods for supervised neural networks can be found in [49], and an example of its application in a microarray study can be found in [48].

Cross validation

There are a number of cross validation approaches used to give an unbiased estimation of the error rate. Examples of these will now be discussed.

First, in Monte Carlo resampling, a training, validation and test set are randomly constituted, with a predetermined number of cases in each subset. All three sets may be randomized, or alternatively the test subset may be kept constant, with the training and validation sets drawn at random a number of times, to enable comparison between models for validation data [7].

Bootstrapping has been shown to be an effective strategy for estimating the error of predictive values in neural network models, and therefore is a reliable approach in determining generalization of the network [50]. In bootstrapping, rather than repeatedly analysing subsets of data (as in the Monte Carlo approach), subsamples of the data are analysed, where many 'pseudo-replicates' are created by resampling the original data. Here, cases are drawn at random from the data set, with equal probability, in order to replicate the process of sampling multiple datasets. The 0.632 bootstrap error estimator has been preferred in small sample microarray classification [51, 52].

k-fold validation is an effective approach when the number of samples is not efficient enough to split the data into three subsets. In a widely used version of this called leave one out cross validation [53, 54], N divisions are made (where N is the total number of cases in the dataset) and in each division the network is trained on all of the samples except one, which is set aside for test purposes. This process is repeated so that all of the samples are used once for testing. Tenfold validation is commonly used when the number of samples is relatively high (e.g. >100) whilst leave-one-out methods are useful when the training set is lower (e.g. <100) or when the number of features is higher than the number of examples. This multiple cross validation helps to minimize overlap of the test set compared to resampling. For an overview of assessing the accuracy of prediction algorithms for classification problems, the interested reader is directed to [55].

EXPERIMENTAL METHODS REQUIRING ROBUST BIOINFORMATICS

The advent of these high-throughput techniques has increased the potential for identification of new biomarkers massively. These methods facilitate the comprehensive profiling of samples representing disease states. The hurdle to overcome with these technologies is now the sheer complexity of the data generated. This complexity is necessary to represent coverage (or even partial coverage given current technological limitations) of the genome or proteome. MAs are one of the methods commonly used for the high throughput sample profiling at the transcript level, whilst MS is being used to detect changes at the protein level. These technologies are therefore complementary to one another in describing biological systems, and the basic principles will be briefly outlined.

Microarrays

A DNA MA consists of a solid surface, onto which DNA molecules have been chemically bonded. The purpose of MAs is to detect the presence and the abundance of labelled nucleic acids in a given biological sample, which will then hybridize to the DNA on the array, and become detectable via the label. The source of the labelled nucleic acids is the mRNA of the sample of interest, so therefore the purpose of a MA is to measure gene expression. As there may be thousands of different DNA molecules bonded to an array, it is possible to measure the expression of many thousands of genes simultaneously, leading to the potential for extremely high throughput analysis. There are two major types of MA technology used today; firstly cDNA and secondly oligonucleotide arrays, such as those marketed by Affymetrix. For a more detailed explanation of the technology, the reader is referred to [56], or more specifically [57] and [58] for cDNA and oligonucleotide MAs respectively.

Mass spectrometry

MS approaches, more specifically MALDI (matrix-assisted laser/desorption ionization) and a modification of this named SELDI (surface enhanced laser desorption/ionization) TOF (time of flight) MS are now being readily used to generate proteomic profiles of biological samples. Simply, a mass spectrometer consists of an ion source, a mass analyser to measure the mass/charge ratio (m/z) of the analytes which have been ionized (mass spectrometers do not measure mass directly, but rather the mass to charge ratio of ions formed), and finally a detector that records the number of ions at each m/z value. This generates a spectrum according to the time of flight of the ion, directly related to its mass, or a 'fingerprint' for the sample being analysed. For an overview of the method see [59]. These analyses have an inherent ability to generate profiles consisting of hundreds of thousands of points, with each point representing a protein mass, a peptide mass or a fragment of the above. This high dimensionality provides an obstacle and limits many analysis methods.

DATA COMPLICATIONS IN PROTEOMICS AND GENOMICS

Dimensionality and complexity

Biological '-omics' datasets are unusual in that there is a very large p (input variables) and relatively small n (cases). As the dimensionality of the input data space ($p \times n$) increases, it becomes exponentially more difficult to find a global optimum for the parameter space. This has been termed 'the curse of dimensionality' [60], and often leads to an input space with many irrelevant or noisy inputs, which coupled with the wide heterogeneity commonly found in biological samples, make it difficult to identify the truly important markers with predictive algorithms performing badly as a result of them modelling extraneous portions of the data space. Conventional statistical theory would indicate that for a valid representation of the population one should have a model where $n > p$, and some rules state that to have confidence in results there should be at least 10 events for each variable [61]. Clearly some form of dimensionality reduction/variable selection algorithm is required to satisfy this, because acquiring a data set containing hundreds of thousands of samples is not feasible. Ma and Huang [62] review the topic of feature selection in bioinformatics, and for a review to approaches for dimensionality reduction in biomarker studies the reader is referred to [63].

Reproducibility

Superimposed on the dimensionality issues are those of data quality. In order to identify biomarkers the data should be reproducible within samples, between sample runs and across multiple instruments (at least

instruments of the same model) [64]. This can be optimized through the use of technical and experimental replicates, where filtering and averaging of samples are methods which are commonly used to assess reproducibility and increase the confidence in the profiles for comparison. Technical replicates provide information on the variability that occurs when performing a particular assay, whilst experimental (or biological) replicates give a measure of the natural sample to sample variation. Lack of reproducibility decreases the validity of markers and makes validation and ultimately clinical use difficult [65]. Low reproducibility within the data adds to the issues of dimensionality by making the relevant features within data sparser with respect to the overall noise. Low replication and poor data quality can lead to the introduction of features not representative of disease, but of sample run, sample collection, storage and preparation. This introduces random, noisy and unimportant features within the data, further increasing the problem of data analysis.

RECENT APPLICATIONS

This section will now highlight recent applications of ANN technologies in MAs and MS. Since the majority of studies involving the use of ANNs are in tumour diagnosis, the following will focus on the field of cancer. Table 1 summarizes the majority of studies using ANNs with these technologies since 2001, and a selection of these will now be discussed in more detail.

Table 1: Cancer studies using artificial neural networks to analyse microarray and mass spectrometry data since 2001

Platform	Cancer type	Number of cases	Number of classes	References
MA	Astrocytoma	65	2	[86]
MA	Astrocytoma	60	2	[69]
SELDI-TOF	Astrocytoma	12	2	[80]
MA	Breast	58	2	[68, 71]
MA	Breast	10	2	[87]
MA	Breast	49	2	[9]
MA	Breast	78	2	[69, 88]
MA	Breast	15	2	[69]
SELDI-TOF	Breast	40	2	[89]
SELDI-TOF	Breast	82	2	[90]
MA	Colorectal	62	2	[69, 91]
SELDI-TOF	Colorectal	147	2	[82, 92]
SELDI-TOF	Colorectal	93	2	[83]
MA	Oesophageal	28	2	[93]
MA	Leukaemia	72	2	[94]
MA	Leukaemia	64	2	[95]
MA	Leukaemia	38	2	[69]
MA	Leukaemia	57	3	[69]
MALDI-TOF	Liver	132	2	[84]
SELDI-TOF	Liver	106	2	[96]
SELDI-TOF	Liver	182	2	[97]
MA	Lung	32	2	[69]
MA	Lymphoma	40	2	[72]
MA	Lymphoma	220	2	[75]
MALDI-TOF	Melanoma	100	2	[65]
SELDI-TOF	Melanoma	205	2	[85]
MA	Myeloma	105	2	[77]
MA	Neuroblastoma	56	2	[76]
MA	Ovarian	54	2	[98]
MA	Prostate	102	2	[69]
SELDI-TOF	Renal	138	2	[81]
MA	SRBCT	88	4	[66–70]

SRBCT: small round blue cell tumours.

Genomics

The seminal paper by Khan *et al.* [66] was perhaps the first major application showing the potential advantages of using ANNs for these complex datasets. Here they used principal components analysis (PCA) followed by ANNs to classify 88 round blue-cells tumours into four diagnostic categories based on cDNA MA analysis of over 6000 genes. Due to the high accuracy of the models developed the authors eluded to the potential use of ANN-based methodologies 'as an adjunct to routine histological diagnosis'. This dataset was made available for the scientific community to download and has since formed the basis for several more studies using various ANN-based algorithms in the successful classification of these samples [67–70].

In [71], Gruvberger and colleagues used PCA for dimensionality reduction followed by ANN analysis to predict the oestrogen receptor (ER) status of 58 tumours from their gene expression profiles. Here they performed a series of classifications using different sets of 100 genes and showed the ANN performance to be good discriminators on this data. As a result of using ANNs, they hypothesized that the classification was not only controlled by a few differentially expressed genes, but a more complex expression pattern existed involving a larger number of genes.

In predicting long term survival of 40 patients with large B-cell lymphoma, O'Neill and Song [72] used the data generated by Alizadeh *et al.* [73] containing 12 078 transcripts representing expression levels for 4026 genes. This was the first time ANNs were shown to have the ability to perfectly classify (100% accuracy) this type of high dimensional data, and also provided a robust solution for reducing

unknown noise and redundancies in datasets whilst maintaining correct classifications.

Using the data made accessible by Rosenwald *et al*. [74], Ando and co-workers [75] described the use of fuzzy neural networks as an approach to variable selection in the expression profiling of 220 diffuse large B-cell lymphoma patients in an effort to predict survival from 7384 genes. Here, using just four genes, ANNs were shown to predict outcome with a classification accuracy of 73%. The analysis in the original manuscript achieved a lower accuracy using more genes in a Cox model. Moreover, the authors showed that by increasing the number of genes in their model to 35 (many of which were clinically relevant to the prognosis of lymphoma), the accuracy increased to 91%. They were able to extract informative rules from their models, with a view to using these approaches in future approaches focused on personalized medicine.

Wei *et al*. [76] used cDNA MAs to analyse 56 tumour samples from patients with neuroblastoma. Total 37 920 data points for each of the samples remained to be analysed after the removal of poor quality data. Due to this complexity, the authors chose to utilize the power of ANNs in order to develop a predictor of survival. Using all of the data in a model, high accuracies were achieved (88%). What is more, they proposed an ANN base gene minimization strategy and identified a signature of 19 genes, some of which had previous affiliations as prognostic markers. This subset of 19 genes had the ability to correctly classify 98% of the patients and further partition the patients into subgroups according to survival status. They concluded that ANN-based approaches such as this would allow therapies to be tailored in a patient specific manner according to their gene expression profiles.

Using ANNs to analyse a 7129 gene expression dataset derived from 74 patients diagnosed with multiple myeloma and 31 normal bone marrow cases, Narayanan *et al*. [77] showed how genes that were consistently positive or negatively expressed could be identified from large datasets. They achieved this by using the interconnecting weights of the trained ANN model, and demonstrated how ANNs could be utilized as a powerful method for dimensionality reduction by identifying 39 genes with 100% generalization on unseen cases. Many of these genes had been previously linked to cancer. Furthermore, the authors described how symbolic knowledge can be extracted from these trained ANN models in order to create simple rules. For example, if gene x is present then myeloma, and if gene y is absent then normal. This made clear the potential for the use of ANNs in a clinical setting.

In one of our own studies [9], we presented a novel stepwise algorithm using ANNs so that optimal predictive gene signatures can be identified from highly complex, noisy and heterogeneous datasets. Using the dataset published by West *et al*. [78] we identified gene subsets highly predictive for ER status and lymph node status in 49 breast cancer cases analysed by MA containing 7129 gene transcript intensities per patient. As with other studies using ANNs, many of these genes had previously been associated with cancer. When the models were applied to a completely separate 88-patient cohort dataset made available by Huang *et al*. [79], accuracies of 88% and 83% were seen for predicting ER and lymph node status respectively. This manuscript also showed how ANNs could be used in the interrogation of predictive biomarkers to provide an insight into how the increased or decreased expression affects the class of interest, enabling rules for molecular classification to be derived.

Proteomics

One of the first major applications of ANNs for the analysis of MS data was in the classification of astrocytoma by Ball *et al* [80]. They showed the early promise of utilizing SELDI-TOF MS technology combined with intensive computer algorithms for protein expression screening in cancer patients. Here ANNs were used to screen ~100 000 data points generated by SELDI-TOF MS, and by scrutinizing the interconnecting network weights, the authors were able to assign a relative importance value to each ion in terms of its contribution to the classification. The top 50 ions were identified, which could be grouped into several sub-groups according to their mass. Furthermore, an additive approach was performed in order to find the optimal combination of ions in terms of predictive ability. This led to the identification of two ions that in combination were able to predict tumour grade with an accuracy of 94%.

Rogers *et al*. [81] also used SELDI-TOF MS in their study on urinary proteins in renal cancer. Here, ANNs were utilized in an effort to detect early onset of disease, and identify indicative biomarkers. Following pre-processing using peak identification, ANN models were built and trained using several

types of controls (healthy controls and benign cases combined with healthy controls). Both peak presence/absence (categorical), as well as actual peak intensities (continuous) were used, with the latter shown to be more efficient. This highlighted the importance and extra information gain that is achieved using actual intensity data to capture the heterogeneity in biological systems rather than peak presence/absence. Moreover, this study utilizing ANNs achieved superior results to the urinary protein assays that were available at the time for bladder cancer.

With a current lack of reliable biomarkers for colorectal cancer, Chen *et al.* [82] proposed the use of proteomics combined with ANN analysis for the discovery of key proteins able to distinguish colorectal cancer patients from a healthy population. To achieve this, MS profiles were generated by SELDI-TOF MS for an age and gender matched cohort of 55 colorectal cancer cases and 92 healthy controls. Initially analysis by cluster analysis showed 54 peaks of interest, culminating in the identification of four candidate biomarkers significantly elevated in colorectal cancer patients. These four ions were then used in an ANN model to build a classifier and discriminate healthy controls from cancer. Here, this approach was shown to outperform discriminant analysis and achieve a sensitivity of 91% and specificity of 93%.

Similarly, Ward *et al.* [83] were also interested in data mining SELDI-TOF MS data for reliable biomarkers of colorectal cancer. They performed proteomic profiling on 62 colorectal cancer patients and 31 non-cancer controls. First, feature selection by *t*-test was conducted, with statistically significant differentially expressed peaks selected for ANN training. The final ANN model included seven peaks and was able to classify with high sensitivity (95%) and specificity (91%), and outperformed CEA (a marker of proven benefit in prognosis and benefit) in discriminating colorectal cancer.

In an effort to improve the prognosis of breast cancer patients through early diagnosis, Hu *et al.* [54] also used SELDI-TOF-MS to explore for reliable tumour markers in serum. They performed screening of the serum proteome in 49 breast cancer patients, 51 patients with benign breast diseases and 33 healthy controls. Total 253 mass peaks were identified using discriminant analysis in classifying between breast cancer and benign, and also between breast cancer and benign plus controls. Using a stepwise approach to assess the predictive ability for each peak, an ANN was able to narrow down the number of markers of interest to just four peaks. These were able to accurately predict the outcome of cancer with a sensitivity of 76% and specificity of 90% for the blind test set. This four-peak model did not result in a statistically significant reduction predictive performance compared to the 253 peak model, and therefore the four-peak model was shown to be more parsimonious in discriminating cancer patients from healthy controls.

Luk *et al.* [84] focused their work on hepatocellular carcinoma biomarkers, investigating differentially expressed proteins between tumour and adjacent healthy liver tissue. Here, proteomic profiling was performed using MALDI-TOF MS and 2D gel electrophoresis followed by analysis by ANNs and decision trees. Both techniques proved to be excellent discriminators of the two phenotypes, with ANNs superior in both training and validation data.

Mian *et al.* [85] were interested in profiling the serum proteome in the classification of early and late stage melanoma, and also predicting disease progression. Here, screening of the patients' proteome was performed with MALDI-TOF MS, showing an interesting signal with significantly higher intensity in 25% of the stage IV samples. ANN modelling in the lower mass range of the spectrum was shown to accurately classify between disease stages and also between progressors and non-progressors. Interestingly, when predicting disease progression, this ANN approach was shown to outperform S100-β, a widely utilised correlate of tumour burden in melanoma.

COMPARISON TO OTHER MACHINE LEARNING METHODS

There have been a number of studies comparing ANNs with other statistical and machine learning approaches to data analysis. Some of these will now be briefly reviewed, outlining how ANNs have performed compared to other statistical and machine learning methods when applied to biological data. This brief discussion will include but will not be focused singly on MA and MA methodology benchmarking studies, as few have been published.

Dreiseitl *et al.* [99] compared the ability of *K*NN, logistic regression, ANNs, decision trees and SVMs in classification of skin lesion data. The authors found

logistic regression, ANNs and SVMs to give almost identical results, with *k*-nearest neighbours and decision trees performing the worst. Interestingly, even the worst of the five methods (decision trees) achieved sensitivity and specificity values comparable to human experts indicating these approaches may be of use to assist human decisions in the medical arena.

Sargent [100] carried out a review on 28 cases comparing ANNs with other statistical approaches when applied to medium and large data sets with more than 200 cases. ANNs outperformed regression in 36% of the studies, and was outperformed in 14% of the studies, with the results being similar in the remaining cases.

Pal *et al.* [67] used an ANN-based approach in categorizing subgroups of cancer from microarray data. They identified a smaller number of biomarkers when compared to other machine learning tool such as SVMs whilst performing equally well, suggesting ANNs found a more parsimonious solution. The study performed in [101] was a direct comparison of SVMs and ANNs in the detection of mammographic CAD. Overall, the authors found a similar performance between the two techniques, with ANNs slightly outperforming SVMs in detection and diagnosis in the test set of data.

Song and co-workers [102] compared various machine learning techniques to more classical statistical approaches in the prediction of outcome in two datasets. They used ANNs (single and multi-layered), logistic regression, least squares linear separation and support vector machines (SVMs) to determine the risk of death in a population of patients with cardiac problems. They found the multi layered ANN to be consistently better than the other approaches, suggesting that the ability of the ANN to model non-linear data was providing additional information regarding the datasets leading to higher predictive capabilities [103].

Eftekhar and colleagues [104] made a comparison between ANNs and logistic regression models to study patients with head injury trauma. It was reported that ANNs significantly outperformed the logistic models in discrimination and calibration (goodness of fit) in 77.8% of cases but underperformed in 68% of cases when comparing model accuracies.

In the study by Hu *et al.* [54] the authors compared their ANN model with other commonly used machine learning techniques such as SVMs and decision trees. They showed ANNs to be more reliable than the other methods in the discrimination of cancer patients from normal controls from mass spectrometry data.

Shen and Tan [105] used different coding strategies and feature selection methods in comparing SVMs to other machine learning methods on two cancer microarray datasets. Here, ANNs achieved similar results to SVMs and outperformed K-nearest neighbour and C4.5 decision tree approaches.

Another direct comparison between ANNs and SVMs was performed by Romero and Toppo on a variety of benchmark datasets [106]. Overall, ANNs obtained similar accuracies to SVMs and the two approaches remained competitive across the different datasets.

In 2008, Peterson and co-workers [69] performed a comparison of a large number of machine learning methods (including ANNs, SVMs, K-nearest neighbour, linear discriminant analysis and logistic regression) in the classification of DNA microarrays in cancer research. One of the main findings here was that at the greatest level of sample size ANNs out-performed all other methods resulting in the greatest area under the curve.

Judson *et al.* [107] performed a comparison of six machine learning approaches in complex simulated datasets. They showed that, particularly when using a large number of features, ANNs and SVMs were always the top performers, whereas recursive partitioning and regression trees and K-nearest neighbours were always the poorest.

In a study investigating heart rate variability before a Paroxysmal atrial fibrillation event using ANNs and SVMs [108], Chesnokov showed ANNs provided better results in terms of sensitivity, specificity and positive predictive value compared to SVM which became biased towards positive cases.

Muselli and co-workers [109] proposed an ANN-based method for gene selection microarray data. In both the artificial and the real gene expression data, they showed that SVMs exhibited poor performance compared to the ANN-based method.

FUTURE TRENDS

As with the development of a novel therapeutic agent, model systems representing novel biomarker expression signatures (be it gene or protein expression) must be validated carefully and extensively in a

medical setting. Trained models of these biomarker signatures need to be incorporated into simple software solutions so that medical practitioners who are unsavvy in machine learning techniques can simply enter the biomarker profiles from their patients and receive an instant prediction with an acceptable degree of confidence. If it can be shown that the application of such models leads to an improvement in medical care towards the holy grail of cost effective 'personalized medicine', then these ANN software applications may be more widely acceptable and made more readily available to assist patient care in a larger number of hospitals and clinics.

CONCLUSIONS

Rapidly advancing technologies in genomics and proteomics have increased the complexity of data being generated, and with that the requirement for robust data mining approaches in order to analyse and extract panels of biomarkers from biological systems. This review introduces one such approach, artificial neural networks, as a robust tool able to digest these datasets and identify the key components (biomarkers), thus providing an increased understanding of the biological system being modelled whilst also pointing out potential therapeutic targets for focusing future research. Representative works in this field and comparisons with other popular statistical and machine learning techniques are highlighted to provide the interested reader with the sufficient background information required so that they can utilize the potential power of these approaches in the modelling of their own complex datasets.

Key Points

- Artificial neural networks are introduced; background theory, advantages and disadvantages are described.
- Modelling considerations when analysing high dimensional data are discussed, specifically the importance of regularization.
- Complex datasets generated by microarray and mass spectrometry experiments are outlined.
- Recent applications of artificial neural networks in analysing microarray and mass spectrometry data for predictive modelling and biomarker identification are reviewed.
- Example studies comparing artificial neural networks to other modelling approaches are highlighted.

Acknowledgements

The John and Lucille van Geest Foundation.

References

1. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006;**24**:971–83.
2. Simon R, Radmacher MD, Dobbin K. Design of studies using DNA microarrays. *Genet Epidemiol* 2002;**23**: 21–36.
3. Check E. Proteomics and cancer: running before we can walk? *Nature* 2004;**429**:496–7.
4. Dickie GL, Fleminger S, Altman DG, *et al*. Statistics notes. *Br Med J* 1994;**309**:539.
5. Garber K. Debate rages over proteomic patterns. *J Natl Cancer Inst* 2004;**96**:816–8.
6. Baldi P, Brunak S. *Bioinformatics: The Machine Learning Approach*, 2nd revised edn. Cambridge, MA: MIT Press, 2001.
7. Bishop C. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, 1995.
8. Lisboa PJ, Taktak AF. The use of artificial neural networks in decision support in cancer: A systematic review. *Neural Netw* 2006;**19**(4):408–15.
9. Lancashire LJ, Rees RC, Ball GR. Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach, *Artif Intell Med* 2008;**43**(2):99–111.
10. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol* 1990;**52**: 99–115; discussion 173–97.
11. Mitchell TM. *Machine Learning*. USA: McGraw-Hill Education, 1997.
12. Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. *J Microbiol Methods* 2000;**43**:3–31.
13. Lancashire LJ, Mian S, Ellis IO, *et al*. Current developments in the analysis of proteomic data: artificial neural network data mining techniques for the identification of proteomic biomarkers related to breast cancer. *Current Proteomics* 2005; **2**:15–29.
14. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; **323**:533–6.
15. Rumelhart DE, McClelland JL. *Parallel Distribution Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 1986.
16. Werbos P. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Science*. Cambridge, MA: Harvard University, 1974.
17. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004;**4**: 309–14.
18. Simon R, Radmacher MD, Dobbin K, *et al*. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003;**95**:14–8.
19. Fahlman SE. Faster-learning variations on back-propagation: An empirical study. In Sejnowski TJ, Hinton GE, Touretzky DS (eds). *1988 Connectionist Models Summer School*. San Mateo, CA: Morgan Kaufmann, 1988.

20. Riedmiller M, Braun H. Direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: The *IEEE International Conference on Neural Networks,* 1993. pp. 586–91.
21. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quart Appl Math* 1944;**2**: 164–8.
22. Marquardt DW. An algorithm for the least-squares estimation of nonlinear parameters. *SIAM J Appl Math* 1963;**11**: 431–41.
23. Kohonen T. *Self-Organization and Associative Memory*. Berlin: Springer, 1989.
24. Boutros PC, Okey AB. Unsupervised pattern recognition: an introduction to the whys and wherefores of clustering microarray data. *Brief Bioinform* 2005;**6**:331–43.
25. Olden JD, Joy MK, Death RG. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol Modell* 2004;**178**: 389–97.
26. Silva I, Cortez P, Santos MF, *et al*. Rating organ failure via adverse events using data mining in the intensive care unit. *Artif Intell Med* 2008;**43**:179–93.
27. Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol Modell* 2003;**160**:249–64.
28. Smith AE, Nugent CD, McClean SI. Evaluation of inherent performance of intelligent medical decision support systems: utilising neural networks as an example. *Artif Intell Med* 2003;**27**:1–27.
29. Tung WL, Quek C, Cheng P. GenSo-EWS: a novel neural-fuzzy based early warning system for predicting bank failures. *Neural Netw* 2004;**17**:567–87.
30. Wall R, Cunningham P, Walsh P, *et al*. Explaining the output of ensembles in medical decision support on a case by case basis. *Artif Intell Med* 2003;**28**:191–206.
31. Duh MS, Walker AM, Ayanian JZ. Epidemiologic interpretation of artificial neural networks. *Am J Epidemiol* 1998; **147**:1112–22.
32. Chiorino G, Mello Grand M, Scatolini M, *et al*. From single gene to integrative molecular concept MAPS: pitfalls and potentials of microarray technology. *J Biol Regul Homeost Agents* 2008;**22**:7–16.
33. Shi L, Jones WD, Jensen RV, *et al*. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics* 2008;**9**(Suppl. 9):S10.
34. Shi L, Perkins RG, Fang H, *et al*. Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. *Curr Opin Biotechnol* 2008;**19**:10–8.
35. Kiehntopf M, Siegmund R, Deufel T. Use of SELDI-TOF mass spectrometry for identification of new biomarkers: potential and limitations. *Clin Chem Lab Med* 2007;**45**: 1435–49.
36. Callesen AK, Christensen R, Madsen JS, *et al*. Reproducibility of serum protein profiling by systematic assessment using solid-phase extraction and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2008;**22**:291–300.
37. Callesen AK, Vach W, Jorgensen PE, *et al*. Reproducibility of mass spectrometry based protein profiles for diagnosis of breast cancer across clinical studies: a systematic review. *J Proteome Res* 2008;**7**:1395–402.
38. Schiffer E, Mischak H, Theodorescu D, *et al*. Challenges of using mass spectrometry as a bladder cancer biomarker discovery platform. *World J Urol* 2008;**26**:67–74.
39. Tiss A, Smith C, Camuzeaux S, *et al*. Serum peptide profiling using MALDI mass spectrometry: avoiding the pitfalls of coated magnetic beads using well-established ZipTip technology. *Proteomics* 2007; **7**(Suppl. 1):77–89.
40. Wang J. Computational biology of genome expression and regulation—a review of microarray bioinformatics. *J Environ Pathol Toxicol Oncol* 2008;**27**:157–79.
41. Barla A, Jurman G, Riccadonna S, *et al*. Machine learning methods for predictive proteomics. *Brief Bioinform* 2008;**9**: 119–28.
42. Grant GR, Manduchi E, Stoeckert CJ, Jr. Analysis and management of microarray gene expression data. *Curr Protoc Mol Biol* 2007; Chapter 19: Unit 19.6.
43. Phan JH, Quo CF, Wang MD. Functional genomics and proteomics in the clinical neurosciences: data mining and bioinformatics. *Prog Brain Res* 2006;**158**:83–108.
44. Wong JW, Cagney G, Cartwright HM. SpecAlign – processing and alignment of mass spectra datasets. *Bioinformatics* 2005;**21**:2088–90.
45. Wong JW, Durante C, Cartwright HM. Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets. *Anal Chem* 2005;**77**:5655–61.
46. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003;**362**:1439–44.
47. Mackay DJC. A practical Bayesian framework for back-propagation networks. *Neural Comput* 1992;**4**:448–72.
48. Kelemen A, Liang WL. Bayesian regularized neural network for multiple gene expression pattern classification. *Proc Int Joint Conf Neural Networks* 2003;**1–4**:654–9.
49. MacKay DJC. Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks. *Network: Comput Neural Syst* 1995;**6**: 469–505.
50. Tibshirani R. A comparison of some error estimates for neural network models. *Neural Comput* 1996;**8**:152–63.
51. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci USA* 2002;**99**:6562–6.
52. Braga-Neto UM, Dougherty ER. Is cross-validation valid for small-sample microarray classification? *Bioinformatics* 2004;**20**:374–80.
53. Braga-Neto U, Dougherty E. Exact performance of error estimators for discrete classifiers. *Pattern Recogn* 2005;**38**: 1799–814.
54. Hu Y, Zhang S, Yu J, *et al*. SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast* 2005;**14**:255.
55. Baldi P, Brunak S, Chauvin Y, *et al*. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 2000;**16**:412–24.
56. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 1999;**21**:33–7.

57. Schena M, Shalon D, Davis RW, *et al*. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;**270**:467–70.
58. Lipshutz RJ, Fodor SP, Gingeras TR, *et al*. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; **21**:20–4.
59. Petricoin EF, Ardekani AM, Hitt BA, *et al*. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;**359**:572–7.
60. Bellman RE. *Adaptive Control Processes. Princeton, NJ:* Princeton University Press, 1961.
61. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med* 2003;**138**:644–50.
62. Ma S, Huang J. Penalized feature selection and classification in bioinformatics. *Brief Bioinform* 2008;**9**:392–403.
63. Hilario M, Kalousis A. Approaches to dimensionality reduction in proteomic biomarker studies. *Brief Bioinform* 2008;**9**:102–18.
64. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* 2004;**3**:367–78.
65. Matharoo-Ball B, Ratcliffe L, Lancashire L, *et al*. Diagnostic biomarkers differentiating metastatic melanoma patients from healthy controls identified by an integrated MALDI-TOF mass spectrometry/bioinformatics approach. *Proteomics Clin Appl* 2007;**1**:605–20.
66. Khan J, Wei JS, Ringner M, *et al*. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;**7**: 673–9.
67. Pal NR, Aguan K, Sharma A, *et al*. Discovering biomarkers from gene expression data for predicting cancer subgroups using neural networks and relational fuzzy clustering. *BMC Bioinformatics* 2007;**8**:5.
68. Peterson C, Ringner M. Analyzing tumor gene expression profiles. *Artif Intell Med* 2003;**28**:59–74.
69. Peterson LE, Coleman MA. Machine learning-based receiver operating characteristic (ROC) curves for crisp and fuzzy classification of DNA microarrays in cancer research. *Int J Approx Reason* 2008;**47**:17–36.
70. Xuan J, Wang Y, Dong Y, *et al*. Gene selection for multiclass prediction by weighted fisher criterion. *EURASIP J Bioinform Syst Biol* 2007;**2007**:Article No. 64628.
71. Gruvberger S, Ringner M, Chen Y, *et al*. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001;**61**:5979–84.
72. O'Neill MC, Song L. Neural network analysis of lymphoma microarray data: prognosis and diagnosis near-perfect. *BMC Bioinformatics* 2003;**4**:13.
73. Alizadeh AA, Eisen MB, Davis RE, *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;**403**:503–11.
74. Rosenwald A, Wright G, Chan WC, *et al*. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;**346**: 1937–47.
75. Ando T, Suguro M, Kobayashi T, *et al*. Multiple fuzzy neural network system for outcome prediction and classification of 220 lymphoma patients on the basis of molecular profiling. *Cancer Sci* 2003;**94**:906–13.
76. Wei JS, Greer BT, Westermann F, *et al*. Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. *Cancer Res* 2004;**64**:6883–91.
77. Narayanan A, Keedwell EC, Gamalielsson J, *et al*. Single-layer artificial neural networks for gene expression analysis. *Neurocomputing* 2004;**61**:217–40.
78. West M, Blanchette C, Dressman H, *et al*. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci USA* 2001;**98**: 11462–7.
79. Huang E, Cheng SH, Dressman H, *et al*. Gene expression predictors of breast cancer outcomes. *Lancet* 2003;**361**: 1590–6.
80. Ball G, Mian S, Holding F, *et al*. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics* 2002;**18**: 395–404.
81. Rogers MA, Clarke P, Noble J, *et al*. Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. *Cancer Res* 2003;**63**:6971–83.
82. Chen YD, Zheng S, Yu JK, *et al*. Artificial neural networks analysis of surface-enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population. *Clin Cancer Res* 2004; **10**:8380–5.
83. Ward DG, Suggett N, Cheng Y, *et al*. Identification of serum biomarkers for colon cancer by proteomic analysis. *Br J Cancer* 2006;**94**:1898–905.
84. Luk JM, Lam BY, Lee NP, *et al*. Artificial neural networks and decision tree model analysis of liver cancer proteomes. *Biochem Biophys Res Commun* 2007;**361**:68–73.
85. Mian S, Ugurel S, Parkinson E, *et al*. Serum proteomic fingerprinting discriminates between clinical stages and predicts disease progression in melanoma patients. *J Clin Oncol* 2005;**23**:5088–93.
86. Petalidis LP, Oulas A, Backlund M, *et al*. Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol Cancer Ther* 2008;**7**:1013–24.
87. Ellis M, Davis N, Coop A, *et al*. Development and validation of a method for using breast core needle biopsies for gene expression microarray analyses. *Clin Cancer Res* 2002;**8**:1155–66.
88. Blazodonakis ME, Zervakis M. Support vector machines and neural networks as marker selectors in cancer gene analysis. *Studies in Computational Intelligence* 2008;**109**: 237–58.
89. Mian S, Ball G, Hornbuckle J, *et al*. A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to Paclitaxel and Doxorubicin under in vitro conditions. *Proteomics* 2003;**3**: 1725–37.
90. Hu Y, Zhang SZ, Yu JK, *et al*. Diagnostic application of serum protein pattern and artificial neural network software in breast cancer. *Ai Zheng* 2005;**24**:67–71.

91. Kim K-J, Cho S-B. Prediction of colon cancer using an evolutionary neural network. *Neurocomputing* 2004;**61**: 361–79.
92. Chen YD, Zheng S, Yu JK, *et al*. Application of serum protein pattern model in diagnosis of colorectal cancer. *Zhonghua Zhong Liu Za Zhi* 2004;**26**:417–20.
93. Kan T, Shimada Y, Sato F, *et al*. Prediction of lymph node metastasis with use of artificial neural networks based on gene expression profiles in esophageal squamous cell carcinoma. *Ann Surg Oncol* 2004;**11**:1070–8.
94. Bicciato S, Pandin M, Didone G, *et al*. Pattern identification and classification in gene expression data using an autoassociative neural network model. *Biotechnol Bioeng* 2003;**81**:594–606.
95. Catchpoole D, Lail A, Dachuan G, *et al*. Gene expression profiles that segregate patients with childhood acute lymphoblastic leukaemia: An independent validation study identifies that endoglin associates with patient outcome. *Leukemia Res* 2007;**31**:1741–7.
96. Wang J-X, Zhang B, Yu J-K, *et al*. Using ANN and serum protein pattern models in liver cancer diagnosis. *Natl Med J China* 2005;**85**:189–92.
97. Ward DG, Cheng Y, N'Kontchou G, *et al*. Changes in the serum proteome associated with the development of hepatocellular carcinoma in hepatitis C-related cirrhosis. *Br J Cancer* 2006;**94**:287–92.
98. Tan TZ, Quek C, Ng GS, *et al*. Ovarian cancer diagnosis with complementary learning fuzzy neural network. *Artif Intell Med* 2008;**43**:207–222.
99. Dreiseitl S, Ohno-Machado L, Kittler H, *et al*. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J Biomed Inform* 2001;**34**:28–36.
100. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer* 2001;**91**:1636–42.
101. Garcia-Orellana CJ, Gallardo-Caballero R, Macias-Macias M, *et al*. SVM and neural networks comparison in mammographic CAD. *Conf Proc IEEE Eng Med Biol Soc* 2007;**2007**:3204–7.
102. Song X, Mitnitski A, Cox J, *et al*. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Medinfo* 2004;**11**: 736–40.
103. Chernushevich IV, Loboda AV, Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* 2001;**36**:849–65.
104. Eftekhar B, Mohammad K, Ardebili HE, *et al*. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med Inform Decis Mak* 2005;**5**:3.
105. Shen L, Tan EC. Reducing multiclass cancer classification to binary by output coding and SVM. *Comput Biol Chem* 2006;**30**:63–71.
106. Romero E, Toppo D. Comparing support vector machines and feedforward neural networks with similar hidden-layer weights. *IEEE Trans Neural Netw* 2007;**18**: 959–63.
107. Judson R, Elloumi F, Setzer RW, *et al*. A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. *BMC Bioinformatics* 2008;**9**:241.
108. Chesnokov YV. Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artif Intell Med* 2008;**43**:151–65.
109. Muselli M, Costacurta M, Ruffino F. Evaluating switching neural networks through artificial and real gene expression data. *Artif Intell Med* 2008; Sep 10. [Epub ahead of print]: doi: 10.1016/j.artmed.2008.08.002.