Medical Engineering & Physics 23 (2001) 217–225
www.elsevier.com/locate/medengphy

Clinical decision support systems for intensive care units: using artificial neural networks

Monique Frize [a, b,*], Colleen M. Ennett [a], Maryhelen Stevenson [c], Heather C.E. Trigg [c]

[a] Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada
[b] School of Information Technology and Engineering, University of Ottawa, Ottawa, ON, Canada
[c] University of New Brunswick, Fredericton, NB, Canada

Received 8 March 2000; received in revised form 23 February 2001; accepted 6 April 2001

Abstract

The paper provides an overview of applications of artificial neural networks (ANNs) to various medical problems, with a particular focus on the intensive care unit environment (ICU). Several technical approaches were tested to see whether they improve the ANN performance in estimating medical outcomes and resource utilization in adult ICUs. These experiments include: (1) use of the weight-elimination cost function; (2) use of 'high' and 'low' nodes for input variables; (3) verifying the effect of the total number of input variables on the results; (4) testing the impact of the value of the constant predictor on the performance of the ANNs. The developments presented intend to help medical and nursing personnel to assess patient status, assist in making a diagnosis, and facilitate the selection of a course of therapy.

Keywords: Decision-support; Artificial neural networks; Outcomes estimation; Intensive care medicine; System performance

1. Introduction

Because of their non-linear modeling capabilities, artificial neural networks (ANNs) have been widely applied to non-linear statistical modeling problems and are a natural choice for modeling large and complex databases of medical information. The goal of training an ANN is to adjust the weights of the network so as to optimize the performance of the network in estimating outcomes for a particular input space. For example, the input space can be a set of medical parameters collected at the time of patient admission to a surgical or medical intensive care unit (ICU), or the data can be collected at different points in time. The backpropagation training algorithm, a popular approach used with medical databases, adjusts the weights of an ANN to minimize a cost function. A commonly chosen cost function is the average sum of squared errors between the desired outputs and actual outputs.

It is well known that a network which has been trained as a classifier will closely approximate a Bayes classifier when the network architecture is sufficiently complex, the training set is sufficiently rich, and the training algorithm succeeds in minimizing the mean squared error [1,2]. Because a Bayes classifier is optimal in the sense that it minimizes the probability of classification error, successfully training a network using the backpropagation algorithm can result in a powerful tool [3]. In practice, however, the ANN error rate of a backpropagation-trained network is higher than the Bayes error rate. The reasons for this include the fact that there are often limitations placed on the training set size and/or network size, and the fact that the network training algorithm may settle into a local rather than a global minimum [1–3].

The literature reports several applications of ANNs to the recognition of a particular pathology. For example, Baxt used ANNs as an aid to diagnose acute coronary occlusion [4] and later for myocardial infarction [5]; Kuntz [6] designed a cascade-correlation ANN to estimate mortality and length of stay for patients with closed-head injuries; Buchman et al. [7] estimated

* Corresponding author at: Systems and Computer Engineering, 4456 MacKenzie Building, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. Tel.: +1-613-520-2600; fax: +1-613-520-2109.
E-mail address: monique_frize@carleton.ca (M. Frize).

1350-4533/01/$20.00 © 2001 Published by Elsevier Science Ltd on behalf of IPEM.
PII: S1350-4533(01)00041-8

chronicity in a surgical intensive care unit; Lau [8] discussed the principles behind the design and validation of a decision-support system for cardiovascular ICUs; Tu and Guerriere [9] reported estimations of length of stay and mortality in ICUs; and Buskard et al. and Frize et al. [10–13] added studies of estimated duration of artificial ventilation to the estimation of mortality and length of stay again in adult ICUs.

2. Background and medical context

For nearly two decades, scoring systems have been used to predict medical outcomes [14,15]. However, these have been more useful in estimating outcomes for a group of patients rather than for a single patient. New approaches should attempt to make estimates on a patient by patient basis, and this has been a main focus of the work reported here. In order to remain as close as possible to the manner in which the medical model works, an ANN model was selected which, when properly trained, provides an estimate of selected clinical outcomes, simulating a clinician's consideration of potential patient outcomes. For example, the physician may think: 'And for this particular patient, this is what I think will happen'.

Another consideration is the particularly fast pace of illness in critically ill patients. This reality has spawned the development of many types of testing and monitoring technologies, rapidly evolving into complex systems [16,17]. While medical devices used in critical care units typically generate huge volumes of information in a short amount of time, much of it can be lost because physicians and nurses do not have time to read through voluminous amounts of output in a critical care setting. In addition, few devices are linked to hospital information systems and each generates its own separate output. This points to the need for investigating an integrated, rather than compartmentalized, approach to critical care (and other medical environments). This, and the need to generalize and test the tools for effectiveness and relevance in a variety of medical contexts, have been the basis on which the work reported here was engaged. The move to using temporal (time-varying) data should eventually result in a 'dynamic' system that estimates patient status in real-time.

3. Methodology

The importance of acquiring a good quality database, error-free and with a standardized approach to data collection is well recognized [18]. Another important step before analyzing the data is to ensure that outliers and obvious errors in the data are removed before proceeding to the analysis. Yale [19] states that 80 percent of the time spent to get an ANN system up and running is typically used for 'massaging' the set of training data. The adult ICU database used in the studies reported here was developed and assessed with the intention to remain as close as possible to these standards.

3.1. The adult ICU database

We had access to a medical database of over 3000 adult ICU patients, containing 98 fields of clinical and administrative information on patients admitted to the ICU at the Doctor Everett Chalmers Hospital (DECH) in Fredericton, NB, Canada. Data collection was primarily prospective, with some retrospective chart review. Up to seven medical diagnoses and multiple procedural information could be entered, with auxiliary space for free-form comments. Significant events and complications were also noted for each patient. A subset of this database with the raw APACHE II variables extracted into a new database was used for all experiments reported herewith; the size of this database being limited by the cost of the medical assistant to compile the data. The new database contained 51 input variables with the most complete profiles and excluded patients under the age of 12, which resulted in a database with 1491 cases [20]. The variable list of this database is shown in Table 1. The database was also separated into postoperative (surgical, 883 cases) and nonpostoperative (medical, 608 cases) patients for experimental purposes because these two patient sets can have drastically different characteristics.

3.2. Data pre-processing

The nonbinary-valued inputs in the two data sets described above were standardized. These variables were scaled so that zero input values represented 'normal' values of the input variables, negative inputs represented 'lower than normal' values and positive inputs represented 'higher than normal' values. The 'normal' values were selected in consultation with the physician (Dr F.G. Solven, intensivist at DECH). To obtain input data of nearly uniform magnitude, the 'normal' value of each nonbinary-valued variable was subtracted from each input value and the resulting differences were divided by three standard deviations pertaining to each input variable value over the entire data set. A zero was assigned as the 'normal' value for binary variables [20].

One remaining and most difficult problem, in our experience, is the question of how to treat variables with missing information. ANNs do not function well with variables that contain a lot of missing information. A common approach is to eliminate the cases with missing data and use the remaining data set to train and to test the network. One can also replace the missing values by the 'mean' or 'median' value for the variable. In the

Table 1
List of variables in the adult ICU database

Demographics and Administrative Information
Assigned chronic health points in APACHE II scoring
Emergency surgery prior to ICU admission
Surgery prior to admission
Patient gender
Position in data sequence[a]
Patient age (years)[a]
APACHE II (Admission Information)
Rectal temperature (°C)[a]
Mean arterial pressure (mmHg)[a]
Heart rate[a]
Respiratory rate[a]
Fraction of inspired oxygen[a]
Partial pressure of oxygen in the blood[a]
Arterial pH[a]
Serum sodium (mmol/l)[a]
Serum potassium (mmol/l)[a]
Serum creatinine (μmol/l)[a]
Hematocrit[a]
White blood cell count (total/mm^3 in 1000s)[a]
Glasgow Coma Score[a]
Admission Source
Emergency Room
4SW
4E
4W
4NE
3W
4NW
3SW
Coronary Care Unit
3E
Admission from another location
Admission Diagnosis #1
Postoperative
Acute hypercapnic respiratory failure
Trauma
Drug overdose
Ketoacidosis
Sudden cessation of heart or lungs
Other diagnosis #1
Admission Diagnosis #2
Carotid endarectomy
Nothing filled in
Abdominal aortic aneurysm repair
Motor vehicle accident
Lobectomy
Aortobifemoral bypass
Pneumonia
Acute pulmonary edema
Other diagnosis #2
Admission Diagnosis #3
Nothing filled in
Lung cancer
Postoperative
Ischemic foot
Other diagnosis #3

[a] Identifies nonbinary-valued variables. (Note: fraction of inspired oxygen is a one-sided continuous variable, therefore, the high/low node experiments did not require an additional node for this variable.)

work reported here, we decided to replace the missing values by the 'normal' value for the variable, which is a third, well-accepted approach. The thinking behind this decision is that, in a set of ICU patients, the mean may be biased towards some particular pathology or outcome, whereas 'normal' values are expected to have the least impact on the outputs. An exception to this technique was made when almost all critical information was missing. In fact, there were only 12 cases missing a significant amount of data, so these records were eliminated before proceeding to the experimental stage. Another reason to select this approach was based on the physician's knowledge that, frequently, missing data occur in medical databases when a test is not done for that patient or the information was deemed to have little importance with respect to the outcomes of interest for that particular patient. Moreover, it was felt that adding a 'normal' value would not disrupt the results for 'abnormal' values since the latter are expected to have the largest impact on poor medical outcomes. This process successfully allowed the ANN to use information contained in incomplete records, thus providing a larger number of records to train and test the networks [20].

3.3. Artificial neural network (ANN) designs to estimate outcomes

3.3.1. Architecture

First, the database was divided into its two main parts since the patient types in each were quite distinct: postoperative patients admitted to the ICU (called POSTOP, 883 patients) and those who were not admitted after a surgery (called NONPOSTOP, 608 patients). For each group (POSTOP and NONPOSTOP), the data were divided into a training set and a test set using two-thirds of the data set to train the ANN and the remaining one-third to test its performance. Using Matlab's Neural Network Toolbox [21], a feedforward ANN was trained using the backpropagation algorithm to estimate the following medical outcomes: mortality, duration of artificial ventilation, and length of stay in the adult ICU [10].

Several architectures were designed, but the best results were obtained with a simple network with one hidden layer (i.e. a three-layer network: input layer, hidden layer, output layer) [20]. This type of network was chosen because of its relative ease of implementation and success in completing various classification tasks, as demonstrated by Haykin [22] and Widrow et al. [23]. Although preliminary results were promising, the ANNs exhibited the behavior characteristic of network memorization (or 'overfitting') [10]. Rather than just reducing the input network size (and therefore network complexity) by eliminating what might potentially be useful input information, we chose to use a technique called 'weight-elimination' to overcome the overfitting problem [24].

The following ANN network parameters were adjusted to optimize the network's performance: learning rate and its adaptive parameters, momentum, weight-elimination scale factor and its adaptive parameters, weight-decay constant, and error ratio. It was also possible to adjust the number of hidden layers and hidden nodes for each ANN experiment.

3.4. Four techniques to improve ANN performance

3.4.1. Impact of the weight-elimination cost function

The weight-elimination cost function includes a penalty term (in addition to the average squared error) that serves to reduce the weights of the least important variables to zero or near zero, thereby removing their influence from the network. A series of experiments for the POSTOP database were run using ANNs with and without the weight-elimination cost function, and results were compared for the particular outcome: 'duration of artificial ventilation less than or equal to 8 hours' or 'more than 8 hours'. In this work, Trigg [20] verified the accuracy of the code by testing its predictive powers on sunspot data collected by Tong in 1983 [25]. The ANN experiments without weight-elimination simply used the sum of squared errors cost function.

3.4.2. Impact of 'high'/'low' node approach for the input variables

The weight-elimination approach was further combined with a novel technique whereby data were presented to a pair of 'high' and 'low' nodes, depending on the value of the parameter, again using the POSTOP database. This means that each of the 14 nonbinary-valued variables was presented to two nodes (a high node and a low node) rather than to one node (i.e. in these experiments there were 65 input nodes, as opposed to 51 in the first set of experiments). To implement the high/low node technique, the standardized values of nonbinary-valued variables were assigned as follows: (1) if the value of the variable was zero or greater, it was presented to the high input node for that variable, and the corresponding low input was assigned a value of zero; (2) if the value of the variable was negative, its absolute value was assigned to the low input node and the corresponding high input node was set to zero. It was hoped that this technique would facilitate the independent interpretation of higher- or lower-than-normal values of input parameters in predicting medical outcomes rather than simply 'abnormal' values. For example, a fever (i.e. a higher-than-normal value) clinically presents different challenges than an abnormally low body temperature [20].

3.4.3. Reduced network complexity (number of input variables)

Here, the research question was: how will reducing the number of input parameters affect the analysis time (the number of epochs needed to reach the highest correct classification rate) and the classification rate itself, when compared to using the full number of variables (51) in the original database? This question was tested using the same outcome 'duration of artificial ventilation: less than or equal to 8 hours; or more than 8 hours' for POSTOP patients. Two input data sets were constructed with a different number of input variables. The first data set contained the original 51 input variables as listed in Table 1. The second data set was constructed from the six variables that attained the largest weights after the application of the weight-elimination cost function. The six parameters that remained after weight elimination were: heart rate, respiratory rate, fraction of inspired oxygen, partial pressure of oxygen in blood, arterial pH, and Glasgow Coma Score [26–28].

Note that as Trigg's results [20] with this particular database only showed a marginal benefit of using three-layer networks (a 1% improvement of the classification rate), two-layer networks were constructed to do this comparative analysis.

3.4.4. Impact of the constant predictor value

Our research group further expanded this research into an analysis of how the constant predictor affects the performance of our ANNs [26–28]. A constant predictor is a statistical benchmark where all cases are classified as belonging to the class with the highest a priori probability. In this series of experiments, we investigated how well the ANNs classified cases into two output classes as the representation of the dominant class approached 100%. Six different dichotomous situations involving the number of hours of mechanical ventilation were investigated: less than or equal to 4 hours, 12, 24, 36 and 336 hours, and between 24 and 336 hours. Also investigated were estimates of the length of ICU stay: 0 days, less than or equal to 1, 4, 5, and 14 days. A commonly estimated medical outcome is mortality (or 'survival rate'), therefore, this output variable was also investigated. Each of the above outcomes under investigation had different outcome distributions with the dominant class ranging from 50.8 to 98.1%.

4. Results and discussion

4.1. Measures of performance

The network performance was evaluated based on the correct classification rate of the test set (i.e. the number of correctly classified cases divided by the total number of cases) and the area under the receiver operating characteristic (ROC) curve. The number of epochs required to reach the best test set classification rate was noted to provide a measure of the convergence speed of the training algorithm. The results are compared to the constant predictor and the minimum distance classifier. These classifiers gauge the difficulty of the classification problem, and provide a lower bound for the network's achievable performance. The standard error for the reported classification rates and areas under the ROC curves was approximated by measuring the maximum variation in the results observed when each network was trained from a set of five different initial weight conditions and estimating the appropriate value as half the maximum variation observed.

The cost of misclassification is an important point to consider. For example, predicting that a patient will not survive surgery (when the patient actually lives) has a different associated cost than foretelling survival, when in actual fact the patient will die. In our case, we are predicting the duration of artificial ventilation for patients in the ICU. Misclassification may upset the management of equipment usage in the unit; however, it would not have a significant impact (negative or positive) on the patient. This decision tool is designed to aid the clinician in estimating the duration of ventilation that the patient requires, which is useful for consultations with the patient and his/her family, as well as for resource management.

4.1.1. Impact of the weight-elimination cost function

The weight-elimination cost function was tested with the outcome 'duration of artificial ventilation less than or equal to 8 hours' or 'more than 8 hours' for the POSTOP patients. Table 2 shows that the weight-elimination ANNs achieved a correct classification rate that was approximately 1.7% better than that of the no weight-elimination networks for the two-layer ANNs, and approximately 1.3% better for the three-layer networks. Weight elimination also eradicated the problem of overfitting previously mentioned. The ROC results show that the networks discriminated well between the two patient sets; however, the bounds of their standard errors overlap slightly (0.9182±0.0213 and 0.9301±0.0195 for the two- and three-layer networks, respectively).

For the sake of completeness, Table 2 also reports that the weight-elimination networks exceeded the performance of the constant predictor and minimum distance classifier (improvements of 19.4% and 4.4% for the two-layer networks, and 20.7% and 5.7% for the three-layer ANNs, respectively). These results show that using the weight-elimination cost function can improve the classification performance of these ANNs [20].

4.1.2. Impact of 'high'/'low' node approach for the input variables

Fourteen continuous-valued physiological variables were separated into high and low nodes as described in the methodology section, according to their values relative to the physiological normal values. Table 3 compares results obtained with using high/low nodes for the ANNs with the weight-elimination cost function, and with ANNs again using weight elimination but with the regular data representation technique (i.e. all values of the variable are presented to the same node, whether they are high or low). Table 3 shows a slight decline in the classification rate for ANNs using the high/low node format compared to the regular data presentation technique (approximately 1% and 0.3% for two- and three-layer networks, respectively). However, these networks still attained higher classification rates than either the constant predictor or the minimum distance classifier (18.4% and 2.4% for the two-layer ANNS, and 20.4% and 4.4% for the three-layer networks, respectively). The poorer performance compared to the ANNs with the regular data presentation approach could be due to the increased number of input variables with the high/low networks. Moreover, the ANNs using high/low nodes were more complex given that the three-layer ANN required eight hidden nodes compared to the regular weight-elimination network which only used two nodes in the hidden layer [20].

Table 2
Performance of ANNs using weight elimination and no weight elimination compared to the constant predictor (CP) and the minimum distance classifier (MDC) for the POSTOP patient database

ANN architecture	Max test set CCR[a] (%)	CP (%)	Performance improvement over CP (%)	MDC (%)	Performance improvement over MDC (%)	ROC[b] curves
Best two-layer network						
With weight elimination (51:1)	90.5±1.20	71.1	19.4	86.1	4.4	0.9182±0.0213
Without weight elimination (51:1)	88.8±0.70	71.1	17.7	86.1	2.7	0.9165±0.0194
Best three-layer network						
With weight elimination (51:2:1)	91.8±1.15	71.1	20.7	86.1	5.7	0.9301±0.0195
Without weight elimination (51:2:1)	90.5±0.90	71.1	19.4	86.1	4.4	0.9212±0.0195

[a] CCR=correct classification rate.
[b] ROC=receiver operating characteristic.

Table 3
Performance of weight-elimination ANNs using regular data presentation or high/low nodes data presentation compared to the constant predictor (CP) and the minimum distance classifier (MDC) for POSTOP patient cases

ANN architecture	No. of input variables	Max test set CCR[a] (%)	CP (%)	Performance improvement over CP (%)	MDC (%)	Performance improvement over MDC (%)	Area under ROC[b] curves
Best two-layer network							
With weight elimination	51	90.5±1.20	71.1	19.4	86.1	4.4	0.9182±0.0213
With weight elimination and high/low nodes	65	89.5±0.85	71.1	18.4	87.1	2.4	0.9204±0.0207
Best three-layer network							
With weight elimination (2 hidden nodes)	51	91.8±1.15	71.1	20.7	86.1	5.7	0.9301±0.0195
With weight elimination and high/low nodes (8 hidden nodes)	65	91.5±1.00	71.1	20.4	87.1	4.4	0.9480±0.0146

[a] CCR=correct classification rate.
[b] ROC=receiver operating characteristic.

4.1.3. Reduced network complexity (number of input variables)

When the number of input variables was reduced from 51 to six, the ANN's classification performance improved. Table 4 shows the results from the two simulations. The constant predictor of the original database was 71.1%. The highest correct classification rate for this test set was 88.8%, an improvement of 17.7% over the constant predictor. The results stabilized after approximately 394 epochs. For the data set with only six input variables, the original assumption was that the results may not be as good as for the other case. We hypothesized that there might be unknown interactions between variables that would not be present when using only a partial list. However, the constant predictor for this data set was 72.4% (the difference is due to case sampling when dividing the training and test sets) and after only 130 epochs, a classification rate of 90.5% was obtained [26].

Comparing the results of these experiments, the simplest database with only six input variables which used only inputs whose weights did not go to zero in Trigg's weight-elimination experiments, produced the highest correct classification rate (90.5%) after just 130 epochs [26]. The results of these experiments indicate that reducing the complexity of this system increased its generalization ability to allow for a better correct classification of the test patterns based on the information provided by the training set in the fewest epochs.

4.1.4. Impact of the constant predictor value

As the distribution of the outcome variables increases from 50.8 to 98.1%, becoming skewed towards one possible outcome, the ANN is less able to significantly exceed the classification performance of the constant predictor. Fig. 1 illustrates the relationship between the correct classification rate of the ANN and that of the constant predictor for the POSTOP database; Fig. 2 shows the same variables for the NONPOSTOP database. These figures illustrate how the classification rate of the ANN and of the constant predictor converge to a theoretical limit for the superior performance of the ANN. This occurs as the division between the two desired output classes becomes highly skewed towards 100%. In cases where the number of sample patterns for a particular case are quite small, after the first few passes through the ANN everything becomes classified as belonging to the largest class—in essence, the ANN classifies like a constant predictor. This point is the lower limit for acceptable ANN performance. From this information one can deduce the minimum number of training patterns required for the ANN to identify rare outcomes.

Using linear regression of the classification rate for

Table 4
Comparison of ANN test results

Database	No. of variables	Max CCR[a] (%)	ASE[b] at Max CCR	Approx. no. epochs
All variables (demographics, APACHE II variables, admission source, admission diagnoses)	51	88.8	0.38	394
Variables with largest weights after weight elimination	6	90.5	0.32	130

[a] CCR=correct classification rate.
[b] ASE=average squared error.

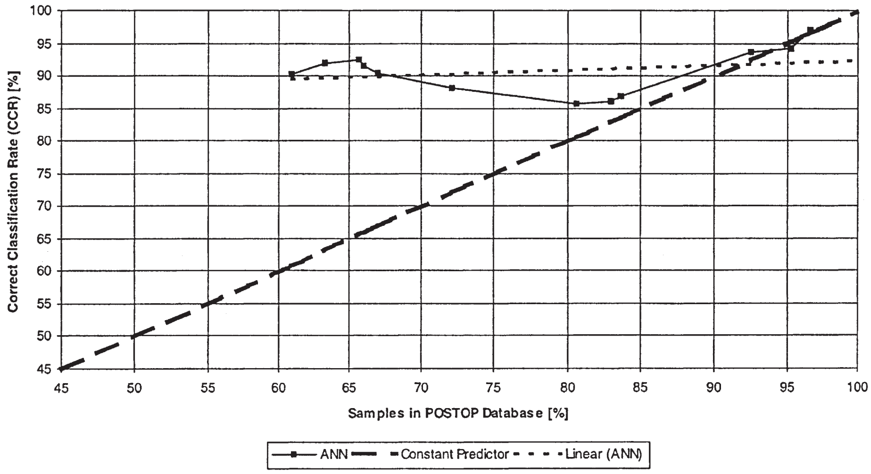

Fig. 1. Comparison of correct classification rate (CCR) for constant predictor and ANN using POSTOP database.

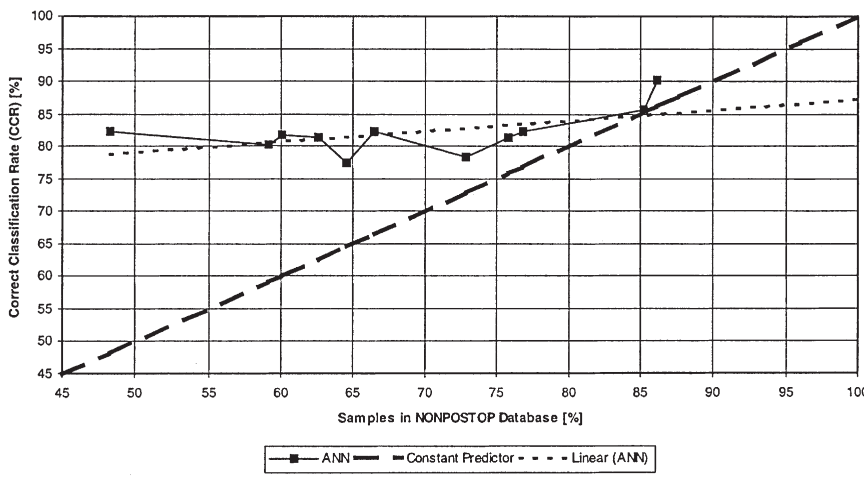

Fig. 2. Comparison of correct classification rate (CCR) for constant predictor and ANN using NONPOSTOP database.

the ANN, we identified this limit [26–28]. The point at which the linear regression line crosses the constant predictor projections is the theoretical limit for the ANN's performance abilities. After this point, as the division between the output classes becomes more skewed, the ANN starts classifying like a constant predictor or its classification performance is worse than a constant predictor due to misclassification of patient cases. From Figs. 1 and 2, the dominant output class may represent at most 92.0% of the POSTOP database, and 84.5% for the NONPOSTOP database under consideration. These limitations cannot necessarily be directly applied to other databases (medical or otherwise) because the ANN relies heavily on the relationships between the input parameters. However, this information could be used as a guideline for verifying the usefulness of ANNs as a predictor with a variety of databases and statistical distributions.

The results of these simulations imply that the ANN had more difficulty classifying the NONPOSTOP patients than the POSTOP patient cases. A possible explanation is that the extreme diversity of the circumstances surrounding the patients in the NONPOSTOP subdatabase makes them more difficult to classify or that more such cases are needed to improve the performance.

5. Conclusion and future work

The new ANN experiments yielded interesting results, allowing a large reduction in the complexity of the system while maintaining high correct classification rates. The results are valid for the adult ICU databases used in these experiments. Other databases are currently being tested to see whether this approach is valid in a variety of contexts. On-going work is applying the same techniques to neonatal intensive care patients (NICU) and to cardiac surgery patients [29–31].

The work reported here attempted to identify how the performance of ANNs could be improved. We conclude that the weight-elimination cost function not only improved the correct classification rate for the adult ICU database, it overcame the network memorization problem. On the other hand, the high/low node approach, in this case, did not improve the performance of the ANNs. This technique should, however, be applied to other databases before considering it ineffective. The third approach has a great potential. Reducing the complexity of the database by eliminating variables that have little impact on the outcome will make the system easier to implement in a clinical environment. For example, entering six variables into a database to assist with clinical decision making is less time consuming compared to systems requiring the full 51 variables used in the first experiments. The smaller database size and reduced complexity will facilitate speedy outcome estimations (due to a lower computational demand on the system) for physicians and nurses when the prototype is used in a clinical setting for this particular data set. Finally, the impact of the constant predictor value on ANN performance may be used as a guideline when looking into appropriate approaches for outcome estimation.

In future work, the missing data problem will be readdressed. In the work described here the missing data were replaced with 'normal' values and the results were quite good. The missing value question must be further tested with a variety of databases and medical environments.

Acknowledgements

This work was completed with the assistance of MRC Grant CGAA-45088 and NSERC Grant 202972-97. The research group is also grateful to Dr F.G. Solven for providing the intensive care patient database from the Dr E. Chalmers Hospital in Fredericton, NB, Canada. Thanks are also due to Helena Ho who performed some of the experiments described in this paper.

References

[1] Haykin S. Neural networks: a comprehensive foundation. New York: MacMillan, 1994.
[2] Hinton GE. How neural networks learn from experience. Sci Am 1992;267(9):144–51.
[3] Richard MD, Lippmann RP. Neural network classifiers estimate Bayesian a posteriori probabilities. Neural Comp 1991;3(4):461–83.
[4] Baxt WG. Use of an artificial neural network for data analysis in clinical decision-making: the diagnosis of acute coronary occlusion. Neural Comp 1990;2(4):480–9.
[5] Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991;115:843–8.
[6] Kuntz GR, Hammond J. Predicting outcomes after closed-head injury (CHI) using a cascade-correlation artificial neural network. Crit Care Med 1994;22(1):A25.
[7] Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ. A comparison of statistical and connectionists models for the prediction of chronicity in a surgical intensive care unit. Crit Care Med 1994;22(5):750–62.
[8] Lau F. Development and validation of a decision-support system for cardiovascular intensive care. Can Med Inf 1994;May/June:28–9.
[9] Tu JV, Guerriere MRJ. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Comput Biomed Res 1993;26:220–9.
[10] Buskard T, Stevenson M, Frize M, Solven FG. Estimation of ventilation, length of stay, and mortality using artificial neural networks. Proc Can Conf Elec Comput Eng 1994;726–9.
[11] Frize M, Buskard T, Stevenson M, Solven FG, Dunfield V. Preliminary results of the development of an artificial neural network for the intensive care. Proc IEEE-EMBS 1993;296.
[12] Frize M, Solven FG, Stevenson M, Nickerson B, Buskard T, Taylor K. Computer-assisted decision-support systems for patient management in an intensive care unit. Medinfo 1995;8 Pt. 2:1009–12.
[13] Frize M, Solven F, Stevenson M, Nickerson BG, McGowan HCE. Information technologies approach and development for various medical applications. Proc Can Conf Elec Comput Eng 1996;365–8.
[14] Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE—Acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 1981;9:591–7.
[15] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centres. Ann Intern Med 1986;104:410–8.
[16] Dawant BM, Uckun S, Manders EJ, Lindstrom DP. The SIMON Project: model-based signal acquisition, analysis and interpretation in intelligent patient monitoring. IEEE Eng Med Biol Mag 1993;12(4):82–91.
[17] Misiano DR, Meyerhoff ME, Collison ME. Current and future directions in the technology relating to bedside testing of critically ill patients. Chest 1990;97(5 Suppl):204S–14S.
[18] Gardner RM, Shabot MM. Computerized ICU data management: pitfalls and promises. Int J Clin Monit Comput 1990;7:99–105.
[19] Yale K. Preparing the right data diet for training neural networks. IEEE Spectrum 1997;Mar:64–6.
[20] Trigg HCE. An investigation of methods to enhance the performance of artificial neural networks used to estimate medical outcomes. Master's thesis, Electrical Engineering Department, University of New Brunswick, 1997.
[21] Demuth H, Beale M. Matlab: Neural Network Toolbox User's Guide. Natick, MA: The Math Works Inc, 1997.
[22] Haykin S. Neural networks expand SP's horizons. IEEE Signal Process Mag 1996;13(2).
[23] Widrow B, Rumelhart DE, Lehr MA. Neural networks: applications in industry, business, and science. Commun ACM 1994;37(3):93–105.
[24] Weigend AS, Rumelhart DE, Huberman BA. Generalization by weight-elimination with application to forecasting. In: Lippmann RP, Moody J, Touretzky DS, editors. San Mateo: Morgan Kaufmann, 1991:875–82.
[25] Tong H. Threshold Models in Non-Linear Time Series Analysis. Springer, 1983.
[26] Ennett CM, Frize M. Investigation into the strengths and limitations of artificial neural networks: an application to an adult ICU patient database. In: Proceedings of the AMIA Symposium, 1998:998.
[27] Frize M, Wang L, Ennett CM, Nickerson B, Solven FG, Stevenson M. New advances and validation of knowledge management tools for critical care using classifier techniques. In: Proceedings of the AMIA Symposium, 1998:553–8.

[28] Ho H. A study of the performance of artificial neural networks to estimate outcomes in the intensive care unit. Undergraduate senior thesis, University of Ottawa, Ottawa, ON, 1998.
[29] Tong Y, Frize M, Walker R. Estimating ventilation using artificial neural networks in intensive care units. Proc BMES/EMBS 1999;.
[30] Walker R, Frize M, Tong Y. Data analysis using artificial neural networks and computer-aided decision-making in the neonatal intensive care unit. Paediatric Academy Society 1999 Annual Meeting. Paediatr Res 1999;45:231.
[31] Ennett CM, Frize M, Shaw RE. Methodologies for predicting coronary surgery outcomes. Proc. BMES/EMBS 1999.