Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Identifying risk factors for two complication types for neonatal intensive care patients (NICU)

Monique Frize, *Senior Member IEEE*, RC Walker and Doaa Ibrahim

Abstract— **This paper discusses the results of applying artificial neural networks to predicting complication for neonatal intensive care patients. Risk factors that lead to necrotizing entero-colitis or broncho-pulmonary dysplasia were identified. Future work will expand this work to other outcomes and add probability information to the estimations.**

I. Introduction

Past research has produced models to predict mortality, length of stay (LOS), and artificial ventilation duration for neonatal intensive care unit (NICU) patients. Some models use severity of illness scoring systems frequently determined by logistic regression analysis. Others apply artificial neural networks to estimate these clinical and resource utilization outcomes. Examples of scoring systems validated for NICU patients are: NTISS (Neonatal Therapeutic Intervention Scoring System) correlates with mortality, LOS, and total hospital charges for survivors in the USA [1]. Other scores contain demographic and physiologic variables instead of therapeutic interventions. 'CRIB' (Clinical Risk Index for Babies) and 'SNAP' (Score for Neonatal Acute Physiology). CRIB, were validated only for infants under 1500 grams or 31 weeks gestational age, correlates with risk of mortality for this cohort of infants and with morbidity, but less well with length of stay (LOS) [2]. CRIB is practical and accounts for the major portion of neonatal mortality and morbidity, but it has several limitations [3]. SNAP requires a much larger number of variables than CRIB, but has been validated for a wider spectrum of neonates [4]. It correlates well with mortality and LOS in NICUs, particularly for infants under 2500 grams [5]. The Canadian Neonatal Network (CNN), with the original SNAP Network and others has developed a simplified version of SNAP using six variables (nine in its "Perinatal Extension" version); this has similar validity to the original SNAP score and has been termed "SNAP-II" [6].

Although illness severity and therapeutic intensity scores have been useful as research tools, they also have important limitations. For example, none has been shown to be of sufficient accuracy in predicting the outcome of individual patients for use in supporting clinical or ethical decisions. Moreover, all scores have limitations in the outcomes that they predict and they do not appear to have any utility in improving diagnostic accuracy or reducing medical errors. Richardson et al. [7] reviewed 30 neonatal scoring systems and found that few had been validated on large, concurrent samples of newborns. Scoring systems have been compared [8]-[9] to help choose such tools, but it is reported that many clinicians remain skeptical about using scoring models in actual patient care [10]-[11]. Scoring systems appear to be little better than clinical judgment in predicting probability of death as it approaches 50%.

The work of our research group focuses on refining and developing clinical decision support tools (CDSSs) to assist physicians in individual patient decision-making [12]. We have used artificial neural networks (ANNs) with the feedforward backpropagation algorithm and hyperbolic tangent transfer function with one hidden layer and an optimal number of nodes to estimate outcomes for a single new patient, based on the experience acquired with a large database of similar patients. ANNs have the potential to model complex interactions between variables and the advantage over conventional statistics is that they can be trained to predict outcomes on any database, even with multiple input parameters and multiple possible outcomes. Another advantage is that they can estimate outcomes for a single patient, whereas statistical tools typically estimate outcomes for a group of patients. Our past work has created models to estimate mortality and duration of ventilation [13], and LOS for both adult and neonatal intensive care patients. The new work involves analysing the data with ANNs to estimate two potential complications while infants are patients of the NICU.

II. Methodology

A. The data

The patient data collected by the Canadian Neonatal Network (CNN) from January 8, 1996 to October 31, 1997 contains all admissions to seventeen NICUs from across Canada during that period (all but one were tertiary-level perinatal centres). This data accounts for 75% of all tertiary-level NICU beds in Canada. The original database contained 20488 admissions during the 22-month collection period (19507 infants when readmissions and transfers are excluded). The experiments carried out for this work focused on the admission data collected within the first 12 hours in the NICU. Patients who were in the NICU for less than 24 hours were excluded to avoid the impact of any differences

M. Frize is a Professor in the Dept. of Systems and Computer Engineering at Carleton University (with a joint appointment in the School of Information Technology and Engineering at University of Ottawa), Ontario, Canada, K1S 5B6 (phone: 613-520-2600 ext 8229; fax: 613-520-5727; (e-mail: mfrize@connect.carleton.ca).

R.C. Walker is a neonatologist at the Children's Hospital of Eastern Ontario and Professor in Pediatrics at the Faculty of Medicine at the University of Ottawa (e-mail: walker@cheo.on.ca).

D. Ibrahim graduated with a Master's in Electrical Engineering at the University of Ottawa in 2005. (e-mail: dibrahim2005@hotmail.com).

in admission protocols at various hospitals. In the CNNetwork database, only deaths to discharge are recorded because there was no follow up after discharge from the hospital. Moribund babies were excluded since they only receive comfort measures rather than aggressive therapy. The complications studied in this project were necrotizing entero-colitis (NEC) and broncho-pulmonary dysplasia (BPD), which were recorded in the database used for the tests. Table I lists the input variables used for each of the two studies: NEC and BPD.

TABLE I
VARIABLE NAMES AND DESCRIPTION

Variable name	Description
Po2fio2r	Lowest pO2/FiO2 ratio
Lurine	Lowest urine output
Lserum	Lowest serum pH
Apgar5	Apgar score at 5 minutes
Lplt	Lowest platelet count
Sga	Small for gestational age
Hsodium	Highest sodium
Hrespr	Highest respiratory rate
Hpco2	Highest pCO2
Bthwt	Birth weight
Lgluc	Lowest glucose
Ltempf	Lowest temperature
Hbloodp	Highest mean blood pressure
Ivhiii	Intra-ventricular hemorrhage or 'IVH' grade III
Ipe	Intra-parenchymal echodensities
Pvl	Peri-ventricular leukomalacia
Cryo	Cryotherapy
BPD	Broncho-pulmonary dysplasia outcome
NEC	Necrotizing entero-colitis outcome

The output variables NEC and BPD had two possible values. A negative output (-1) predicted normal cases, which meant: no NEC or no BPD, and a positive output (+1) predicted the presence of a medical complication (NEC or BPD). Mortality was excluded from the analysis of the outcomes under study since each of these outcomes could be a common cause of death in preterm babies.

For the NEC experiments, the database contained 18,306 complete cases with 17 input variables (see TABLE I) and NEC as output variable. The database was divided into three sets: 1/2 (9,152 cases) of the data for the training set and 1/4 (4,577 cases) for test set and 1/4 (4,577 cases) for validation sets. The original ratio of the abnormal cases for the NEC data was 1.74%, therefore artificial training sets were created containing 20% of these cases. But the testing used the original data sets. This approach improves the classification of positive cases [14].

The database used in predicting BPD had 17,649 complete matching cases with 17 input variables (see TABLE I) and BPD as output variable. The database was divided into 8825, 4412 and 4412 cases for training, test and validation sets respectively. The original ratio of the abnormal cases of BPD was 5.56%. Therefore artificial training sets were again created with 20% of positive cases.

B. Estimation of complications and identification of risk factors in NICU patients using (ANNs)

Several experiments were run to determine the optimal number of nodes in the hidden layer and by adjusting the nine ANN parameters. See Table II. The performance of the training, testing and validation experiments was assessed by recording the Sensitivity, Specificity, Correct Classification Rate (CCR), the Average Squared Error (ASE), and the area under the Receiver Operating Characteristic (ROC) curves.

The correct classification rate is the sum of the number of cases that were correctly classified. The model's sensitivity identifies the rate at which the positive outcome cases are correctly classified (i.e. true positives divided by the total number of positive outcome cases in the dataset). Specificity is the rate of correct classification of the negative outcome out of all negative outcome cases (i.e. true negatives divided by the total number of common outcome cases in the dataset). These parameters are sensitive to the prevalence of the situation under investigation.

TABLE II
THE RANGE OF THE INITIAL ANN PARAMETERS

Parameter	Range
Learning rate (lr)	0.00005-0.005
Learning rate increment (lr_inc)	1-1.3
Learning rate decrement (lr_dec)	1-0.7
Weight-elimination constant (λ)	0.00001-0.01
Weight-elimination constant increment (λ_inc)	1-1.3
Weight-elimination constant decrement (λ_dec)	1-0.7
Weight-elimination scale (w_0)	0.001-1
Momentum	0-0.99
Err_ratio	1.001-1.05
hidden nodes	0-15

TABLE II shows the ANN adjustable parameters and their available ranges. The manual tweaking of these parameters to get the best classification rate possible is a long and strenuous task. However, our research group has automated this process, thus enabling us to perform experiments without user supervision on several computers and 24 hours a day [15].

The next step involved reducing the number of input variables one by one until the performance of the ANN started to degrade, as measured principally by the Sensitivity and Specificity of the training sets. This allows to identify the minimum number of input variables that can predict the outcome of interest (in this case, NEC and BPD). We refer to this as a minimum data set (MDS). These input variables become the model of risk factors that help predict complications that may arise while the infant is in the NICU.

Several methods exist to eliminate variables with small weights. However our approach was to extract weights at the nodes of the input layer after achieving an optimal performance of our ANNs, then begin to eliminate input variables one by one, starting with those with the smallest weights.

Reduction of input variables	Sensitivity	Specificity	CCR
0	42.5	90.8	90.0
1	38.8	91.4	90.5
2	38.8	90.9	90.0
3	35.0	92.2	91.2
4	37.5	90.3	89.4
5	40.0	91.2	90.3
6	40.0	90.4	89.5
7	47.5	89.0	88.3
8	45.0	90.6	89.8
9	43.8	90.5	89.7
10	43.8	89.3	88.5
11	40.0	90.0	89.1
12	**41.3**	**89.9**	**89.0**
13	22.5	94.9	93.6
14	22.5	94.5	93.2

III. Results

The best results for estimating both outcomes (NEC and BPD) were obtained with a three layer network (one hidden layer) and 5 nodes. TABLE III shows the performance results for the two complication outcomes, NEC and BPD.

TABLE III

PERFORMANCE OF ANN FOR TEST AND TRAINING SETS

Performance	NEC %	BPD %
Test Sensitivity	42.5	80.4
Train Sensitivity	64.6	83.0
Test Specificity	90.8	91.8
Train Specificity	90.4	91.9
Test CCR	90.0	91.2
Train CCR	85.2	90.1
Test ASE	28.1	25.9
Train ASE	43.1	31.0
Test ROC	77.5	91.7
Train ROC	85.9	92.0

After the training of the ANN, the testing results of both of NEC and BPD were observed in all the phases of testing. Table IV and V show that the ANNs performed very well in estimating the risk of NEC and BPD and can now be applied to determine the risk of single patients using a few variables collected at the time of admission.

Validation sets containing data not previously seen in either the training or testing sets are analyzed by the ANN and performance measures recorded. The results are shown in TableVI and VII for both NEC and BPD.

TABLE IV

REDUCTION OF VARIABLES FOR NEC

TABLE V

REDUCTION OF VARIABLES FOR BPD

The validation results are for experiments run on the entire input variables, and the results for experiments run after the reduction of variables. The minimum data set (MDS), that is, the input variables that remain after the reduction of smaller weights prior to a degradation of the performance of the ANN are: high values of partial pressure of carbon dioxide in the blood, the presence of BPD, low values of urine output, low temperature, and a low ratio of pO2/FIO2.

Reduction of input variables	Sensitivity	Specificity	CCR
0	80.4	91.8	91.2
1	78.8	92.1	91.4
2	77.6	92.6	91.8
3	75.1	92.4	91.5
4	78.4	92.0	91.2
5	75.9	93.1	92.1
6	77.1	92.5	91.6
7	79.6	91.5	90.9
8	76.7	92.4	91.6
9	79.2	91.1	90.5
10	77.1	92.1	91.2
11	78.0	91.8	91.0
12	82.5	90.1	89.7
13	78.4	92.2	91.4
14	81.2	90.4	89.9
15	78.0	91.4	90.7
16	76.3	91.7	90.8
17	**77.1**	**91.1**	**90.3**

IV. CONCLUSION

Although our analysis demonstrated that five input variables appear to have the most influence on predicting the potential complication called NEC, it is not clear at this time that input variables can be eliminated from the current list, as weights at all input nodes have not been of negligeable value; the factors with the highest values were: partial pressure of carbon dioxide in the blood, the presence of BPD, low values of urine output, low temperature, and a low ratio of pO2/FIO2. This model does correlate well with medical experience. Additionally, it is interesting to note that the presence of chronic lung disease seems to also be a predictor for NEC. Birth weight is the most significant risk factor to estimate BPD. This variable alone, of course, is not sufficient for predicting all cases of BPD. However a low birth weight does indicate the risk for an infant to develop this complication and several others. In previous studies, low birth weight was also a predictor of mortality. Our analyses fit well with the medical model and future work will integrate probability to the current ANN models, as physicians will definitely be interested in this additional information when making their treatment plan decisions. Other research will add other complications such as Neuro-Imaging Abnormality to the current list of outcomes studied.

REFERENCES

[1] Gray, JE, Richardson, DK, McCormick, MC, et al., "Score for Neonatal Acute Physiology (SNAP) and risk of intra-ventricular hemorrhage (IVH)," *Ped. Res*., vol. 31, pp. 249A, 1992.

[2] De Courcy-Wheeler, R.H.B., Wolfe, C.D.A., Fitgerald, A, Spencer, M,Goodman,J.D.S., Gamsu, H.R., "Use of the CRIB score in prediction of neonatal mortality and morbidity," *Arch Dis Child Fetal Neonatal*, vol. 73, pp. F32-36, 1995.

[3] Bard, H., "Assessing neonatal risk, CRIB versus SNAP (editorial)," *Lancet*, vol. 342, pp. 449-450, 1993.

[4] Richardson, D.K., Gray, J.E, McCormick, M.C, Workman-Daniels, K., Goldmann, D., "Score for Neonatal Acute Physiology (SNAP): Validation of a new physiology-based severity of illness index," *Pediatrics*, vol. 91, pp. 617-623, 1993.

[5] Escobar, G.J., Fischer, A., Li, D.K., Kremers, R., Armstrong, M.A.
Richardson D, Corcoran JD, Escobar GJ, Lee SK, for the Canadian NICU Network, "Kaiser Permanente Neonatal Minimum Data Set Wide Area Network, SNAP-II Study Group. SNAP-II and SNAPPE-II: simplified newborn illness severity and mortality risk scores," *J Pediatr*, vol. 138, pp. 92-100, 2001.

[6] Lee S, Normand C, McMillan D, Ohlsson A, Vincer M, Lyons C, Canadian Neonatal Network, "Evidence for changing guidelines for routine screening for retinopathy of prematurity," *Arch Pediatr Adolesc Med*, vol. 155, pp. 387-395, 2001.

[7] Richardson, D.K., Tarnow-Mordi W.O., "Measuring illness severity in newborn intensive care," *J. Int. Care Med*, vol. 9, pp. 20-33, 1994.

[8] Castella, X., Artigas A, Bion J, Kari A."A comparison of severity of illness scoring systems for intensive care unit patients: results of a multi-center, "Multinational Study, *Crit Care Med*, vol. 23(8), pp. 1327-1335, 1995.

[9] Fery-Lemonnier, E., Landais P, Loirat P, Kleinknecht D, Brivet F. "Evaluation of Severity Scoring Systems in ICUs - Translation, Conversion and Definition Ambiguities as a Source of Inter-observer Variability in APACHE II, SAPS and OSF," *Int Care Med*, vol. 21, pp. 356-360, 1995.

[10] Hyzg, R.C., "ICU scoring and clinical decision making [Editorial]," *Chest*, vol., 107 no 2, pp. 1482-1483, 1995.

[11] LeGall, J., Lemeshow S, Leleu G , Klar J, Huillard J, Rui M, et.al., "Customized probability models for early severe sepsis in adult intensive care patients," *JAMA*, vol. 273 no 8, pp. 644-650, 1995.

[12] RC Walker and M. Frize, "Are artificial neural networks 'ready to use' for decision- making in the NICU?," *Journal of Pediatric Research*, (Invited paper), 2004.

[13] CR Walker, CM Ennett and M. Frize, "Use of an Artificial Neural Network to Estimate Probability of Mortality and Duration of Ventilation in Neonatal Intensive Care Patients," *Proceedings of Medinfo'01*, 2001.

[14] C.M. Ennett and M. Frize, "Weight-Elimination Neural Networks Applied to Coronary Surgery Mortality Prediction," *IEEE Transactions of Information Technologies in Biomedicine*, vol. 7(2), pp. 86-92, 2003.

[15] M. Frize, CM Ennett, E. Charette, "Automated Optimisation of Neural Networks Performance," *Medical Engineering and Physics*, vol. 26, pp. 321-328, 2002.