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



# Risk Factors for Apgar Score using Artificial Neural Networks

Doaa Ibrahim, Monique Frize, and Robin C Walker

*Abstract*— **Artificial Neural Networks (ANNs) have been used in identifying the risk factors for many medical outcomes. In this paper, the risk factors for low Apgar score are introduced. This is the first time, to our knowledge, that the ANNs are used for Apgar score prediction. The medical domain of interest used is the perinatal database provided by the Perinatal Partnership Program of Eastern and Southeastern Ontario (PPPESO). The ability of the feed forward back propagation ANNs to generate strong predictive model with the most influential variables is tested. Finally, minimal sets of variables (risk factors) that are important in predicting Apgar score outcome without degrading the ANN performance are identified.**

*Keywords*—Neural networks, perinatal outcomes, Apgar score.

## I. Introduction

Although there is an increase in information retrieval techniques, there is a lot of information being lost in storage. One of the fields that could benefit the most from information analysis is the medical field. The PPPESO began collecting perinatal data 25 years ago. The mission of the PPPESO is to promote optimum perinatal care of childbearing families in Eastern and Southeastern Ontario in order to improve health and to achieve excellent perinatal health outcomes [1]. Perinatal is traditionally defined as the period of time during pregnancy, birth, and the first month after birth [1].

The MIRG (Medical Information technologies Research Group) has demonstrated in several articles that artificial neural networks (ANNs) are an ideal tool to estimate outcomes using large medical databases [2-5]. The current study discusses the use of ANNs to estimate an important perinatal outcome and identify its risk factors. Apgar score is measured at one minute and five minutes after delivery. The Apgar test is considered to be the first test for the newborn and it is used to quickly evaluate the newborn's condition after delivery. It tests five qualities: Appearance (color); Pulse (heartbeat); Grimace (reflex); Activity (muscle tone); and Respiration (breathing). A score is determined by awarding zero, one or two points in each category. Generally, a higher score means a better baby's condition. Scores of seven and over indicate the baby is in good condition. This valuable tool was developed in 1952 by the late pediatrician, Dr. Virginia Apgar [6,7]. The Apgar outcome in this study predicts the Apgar score at 5 minutes after birth (Apgar5).

## II. Methodology

### A. Automated Artificial Neural Network (ANN)

The MIRG automated network was developed to optimize each network parameter without user supervision. An effective technique was implemented that ran the network using the maximum and midpoint values of a parameter and compared the test set performance measure (either highest CCR, lowest MSE or highest log-sensitivity index value). The technique was recursively repeated until no further improvement was possible. During the optimization process the ANN monitored the CCR, ASE and log-sensitivity index for each run. If the classification performance did not improve or the error rate did not diminish within 500 epochs, the training was stopped. The test set performance over this number of epochs was quite stable and reliable [8,9]. In this study the log-sensitivity index was used as stopping criteria [10-12].

### B. ANN Architectures

The number of hidden layers and nodes in each layer affects the complexity and the performance of the networks. Too few nodes may cause under-fitting results. On the other hand, too many nodes may cause long training times and over-fitting. If a network with no hidden nodes performs as well as a network with a large number of hidden nodes, the problem appears to be linear. The majority of clinical studies use networks with one hidden layer [13]. Two and three-layer ANNs were tested in the MIRG automated software in order to get optimized performance.

### C. Variable Reduction

In 1991 Garson proposed a method for partitioning the network connection weights in order to determine the relative importance of each input variable in the network [14]. The Garson algorithm uses the absolute values of the connection weights when calculating variable contributions, and therefore does not provide the direction of the relationship between the input and output variables. In the application of Garson's algorithm by Goh in 1995 and all the researchers following Goh [15,16], the network connection weights between the hidden and the output layers were canceled. This means that the Goh equations did not account for the effect of the hidden output connections in the final calculation of the relative importance. This contradicts the main objective of Garson's algorithm to partition the hidden

D. Ibrahim. Author is graduated in 2005 with a Master's in Applied Science, School of Information Technology and Engineering, University of Ottawa, Ontario, Canada (e-mail: dibrahim@site.uottawa.ca).

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 (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).

**1-4244-0033-3/06/$20.00 ©2006 IEEE.**

output connection weights of each hidden neuron into components associated with each input neuron. We updated Goh's equations, which we now call partitioning weight algorithm. The following updated equations were applied in order to correctly calculate the importance $S_i$ of the input variable $i$ following the original idea of Garson.

$$S_i = \sum_{h=1}^{m} Q_{hi} \tag{1}$$

where $Q_{hi}$ is calculated by Equation 2.

$$Q_{hi} = P_{hi} * |w_{oh}| \tag{2}$$

where $w_{oh}$ is the connection weight from output $o$ to hidden node $h$ and $P_{hi}$ is calculated by Equation 3.

$$P_{hi} = \frac{|w_{hi}|}{\sum_{x=1}^{n} |w_{hx}|} \tag{3}$$

where $n$ is the total number of input variables in the input layer and $w_{hi}$, is the connection weight from hidden node $h$ to the input $i$.

The last step after calculating $S_i$ for each input variable is to obtain its relative importance $RS_i$ (Equation 4). Expressed as

$$RS_i = \frac{S_i * 100}{S_{Total}}$$

a percentage, this gives the relative importance or distribution of all output weights attributable to a given input variable [15].

(4)

In the case of ANNs with more than one hidden layer, the above equations can be generalized and the calculation can be done for each hidden layer. In general, calculating the importance of each node in a certain hidden layer will start backwards from the first hidden layer (before the output) until reaching the input layer. Generally, $n$ is the total number of nodes in the layer being calculated, $m$ is the

$$S_i = P_{hi} = |w_{oi}|$$

number of the nodes in the layer following it, and $S_i$ is the importance of the node $i$ in the current layer [15].

In the case of a two-layer ANNs (with no hidden layer), the importance of each input variable $i$ is equal to its absolute input-output weight:

(5)

The modified partitioning weight algorithm was applied in this study to the optimized network to obtain the minimal sets of variables that are important in predicting Apgar5 outcome.

### *D. Generalization*

The next step after training a network is generalization. Generalization ability of a network is a measure of its performance on data not presented in the training set [13]. The Niday-1999 database (not yet seen by the ANN) was used to create the generalization sets. A generalization set is an independent set that tests the performance of the network.

### *E. Description of the Database*

The Niday-2001 Enhanced Perinatal Database (Niday-2001 database) has 17,688 cases and 38 input variables. This large data set contains 16,183 complete cases and 1,505 cases with missing values. Cases with missing values were deleted for the analysis. Some variables were deleted for medical reason. Apgar score after 1 minute variable was deleted as it likely to be a cofounder variable for Apgar5 (the outcome). Cord blood gases variable, the five resuscitation variables, neonatal mortality variable and neonatal transfer during first week variable were deleted from the Apgar5 estimation list because they happen after the Apgar5 score is measured, so, they can not play any role in deciding the risk factors. The total number of input variables was reduced to 26 after deleting the identification number for each baby birth and baby's date of birth variables. These are important to keep track of the data but not for analysis. The final number of cases and variables used for estimating Apgar5 score was 16,183 cases with 26 input variables. The dataset was divided into 2/3 (10,685 cases) training set and 1/3 (5,498 cases) test set. The original dataset was skewed, containing 2.18% low Apgar score which means less than 7 Apgar score value (+1 type) and 97.82% high Apgar values which includes Apgar values of 7 to 10 (-1 type). Using the original training set with such a low +1 type ratio produced a very low sensitivity. Artificial training set was created to contain 20% of low Apgar5 cases (+1 type). It was found by other researches that this improves sensitivity. Data standardization was applied in order to get a uniform magnitude for each variable values. The mean was subtracted from each input value and the resulting differences were divided by 3 standard deviations over the entire database.

The methodology can be summarized in the following steps:

1. Use the automated software with different ANN architectures to find the optimized model (using the Niday-2001 database)
2. Calculate the relative importance for each input variables using partitioning weight algorithm
3. Apply variable reduction depending on the results of step 2 to obtain the minimal set of variables

4. Test the model of step 1 and the simplest model of step 3 on Perinatal Updated-1999 database

## III. Results

This section presents the ANN experimental results for estimating Apgar5 using Niday-2001 database. Finally, the predictive models are tested using Niday-1999 database.

### 1) *Optimization*

Table I shows the results obtained with different ANN architectures for the prediction of Apgar5 using an artificial training set (tr20%). Three-layer ANNs with different number of hidden nodes in the hidden layer were tested. ANNs with 6, 7 and 13 hidden nodes achieved good prediction results. Three-layer ANN with 13 hidden nodes achieved the highest sensitivity of 45.69% with high specificity and CCR of 96.36% and 95.27%. This model was chosen as the best optimized model.

Table II shows the ANN performance results of the 10 Apgar5 test sets and their mean and standard deviation.

TABLE I
Apgar5 prediction using different ANN architectures

| Performance measure | Three-layer ANNs with different number of hidden nodes | | |
|---|---|---|---|
| Hid Nodes | 6 | 7 | 13 |
| Test Sens | 40.52% | 41.38% | 45.69% |
| Train Sens | 46.06% | 62.10% | 50.65% |
| Test Spec | 97.80% | 95.76% | 96.36% |
| Train Spec | 97.86% | 96.71% | 96.69% |
| Test CR | 96.57% | 94.59% | 95.27% |
| Train CR | 87.50% | 89.78% | 87.48% |
| Test ASE | 17.65% | 20.70% | 18.41% |
| Train ASE | 42.67% | 36.13% | 40.79% |
| Test ROC | 73.73% | 74.38% | 79.20% |
| Train ROC | 77.70% | 83.61% | 80.43% |
| Test Log-Index | 0.089 | 0.091 | 0.115 |
| Train Log-Index | 0.120 | 0.247 | 0.148 |

### 2) *Relative Importance (RI)*

TABLE II
Performance results for 10 test sets for Apgar5

| Test set | Sens. | Spec. | CCR | ASE | ROC |
|---|---|---|---|---|---|
| 1 | 45.69% | 96.36% | 95.27% | 18.41% | 79.20% |
| 2 | 40.98% | 94.84% | 93.62% | 23.77% | 75.38% |
| 3 | 40.31% | 95.20% | 93.88% | 23.33% | 74.98% |
| 4 | 43.10% | 94.96% | 93.85% | 23.08% | 78.07% |
| 5 | 42.75% | 95.75% | 94.46% | 20.86% | 72.23% |
| 6 | 34.17% | 95.30% | 93.94% | 22.96% | 78.49% |
| 7 | 40.00% | 93.19% | 91.86% | 55.87% | 68.30% |
| 8 | 39.23% | 95.10% | 93.75% | 22.70% | 74.46% |
| 9 | 35.90% | 94.56% | 93.29% | 24.54% | 78.25% |
| 10 | 40.31% | 95.20% | 93.88% | 23.33% | 74.98% |
| Mean | 40.24% | 95.05% | 93.78% | 25.88% | 75.43% |
| STD | 3.19% | 0.78% | 0.82% | 10.13% | 3.18% |

Table III shows the relative importance for each input variable in Apgar5 prediction model, ranked in order from highest to lowest.

TABLE III
Relative importance for the input variables in Apgar5 model

| Variable | Description | Rank | RI |
|---|---|---|---|
| gest | gestational age | 1 | 6.40% |
| spinal | pain relief: spinal | 2 | 5.89% |
| scalp | scalp blood gases | 3 | 5.54% |
| bf | mother intention to breastfeed | 4 | 5.44% |
| general | pain relief: general | 5 | 5.41% |
| weight | baby's weight | 6 | 5.13% |
| smoking | maternal smoking after 20 weeks gestation | 7 | 5.12% |
| pudendal | pain relief: pudendal | 8 | 4.70% |
| epidural | pain relief: epidural | 9 | 4.38% |
| nitoxide | pain relief: nitrous oxide | 10 | 4.03% |
| steroids | antenatal steroids | 11 | 3.72% |
| preterm | previous preterm babies | 12 | 3.55% |
| momtrans | mother transfer | 13 | 3.51% |
| hosptype | hospital type | 14 | 3.40% |
| parity | parity | 15 | 3.30% |
| gender | baby's sex | 16 | 3.26% |
| monitor | monitoring methods | 17 | 3.12% |
| labtype | labour type | 18 | 3.10% |
| assisted | forceps or vacuum | 19 | 3.05% |
| present | type of presentation | 20 | 2.83% |
| delby | physician or midwife | 21 | 2.74% |
| momage | mother age | 22 | 2.69% |
| term | previous term babies | 23 | 2.68% |
| deltype | delivery type | 24 | 2.50% |
| narcot | pain relief: narcotics | 25 | 2.48% |
| nbabies | number of babies | 26 | 2.03% |

### 3) *Variable Reduction*

Table IV shows different steps of applying reduction on the best optimized model (consisted of 26 input nodes) using the relative importance values of table III.

The first reduction removed nbabies variable, which had

TABLE IV
Variable reduction for Apgar5 model

| Reduction | Number of input nodes | Sens. | Spec. | CCR |
|---|---|---|---|---|
| 0 | 26 | 45.69% | 96.36% | 95.27% |
| 1 | 25 | 41.38% | 91.57% | 90.49% |
| 2 | 24 | 40.52% | 97.08% | 95.87% |
| 3 | 23 | 44.83% | 88.81% | 87.86% |
| 4 | 22 | 37.50% | 97.77% | 97.78% |
| 5 | 21 | 37.07% | 97.77% | 96.46% |
| 6 | 20 | 40.52% | 97.20% | 95.98% |
| 7 | 19 | 38.79% | 95.97% | 94.74% |
| 8 | 18 | 36.00% | 96.67% | 95.51% |
| 9 | 17 | 39.66% | 97.41% | 96.16% |
| 10 | 16 | 39.66% | 97.41% | 97.85% |
| 11 | 15 | 38.00% | 97.80% | 96.57% |
| 12 | 14 | 36.55% | 97.42% | 96.22% |
| 13 | 13 | 35.54% | 97.42% | 96.22% |
| **14** | **12** | **36.80%** | **97.16%** | **95.94%** |
| 15 | 11 | 31.80% | 92.59% | 91.55% |
| 16 | 10 | 35.00% | 95.17% | 93.99% |
| 17 | 9 | 36.50% | 98.18% | 96.96% |
| 18 | 8 | 34.50% | 96.87% | 97.85% |
| 19 | 7 | 34.50% | 96.87% | 95.64% |
| 20 | 6 | 35.20% | 97.54% | 96.31% |
| 21 | 5 | 36.50% | 87.91% | 87.04% |
| 22 | 4 | 33.50% | 91.29% | 90.34% |
| 23 | 3 | 32.00% | 96.67% | 95.40% |
| 24 | 2 | 31.90% | 97.63% | 96.27% |
| 25 | 1 | 31.89% | 97.27% | 95.94% |

the lowest relative importance of 2.03% (see table III). Reduction continued until reduction fifteen. It removed preterm variable which had 3.55% relative importance ratio. The removal of preterm variable resulted in a clear degradation of the ANN performance. The sensitivity decreased to 31.8% for the first time. Keeping preterm variable resulted in the following reduced model of 12 input nodes: gest, spinal, scalp, bf, general, weight1, smoking, pudendal, epidural, nitoxide, steroids and preterm.

#### 4) Generalization

The prediction models (optimized and reduced) built using Niday-2001 database tested with a data not presented in the training set. Table V shows the results of predicting Apgar5 using Niday-1999 database.

## IV. Conclusion

The risk factors needed to predict low Apgar score without degrading the best measured performance were determined using the Niday-2001 database. Using the log-sensitivity index as a measure of the ANN performance helped in creating the optimized model with the best sensitivities and specificities. The three-layer ANN with 13 hidden nodes in the hidden layer was found to be a strong predictive model with high generalization ability.

TABLE V
TESTING APGAR5 PREDICTION USING NIDAY-1999 DATASETS

| Performance Measure | Best Model | Reduced Model |
|---|---|---|
| Test Sensitivity | 35.75% | 43.02% |
| Test Specificity | 94.65% | 76.85% |
| Test CCR | 92.54% | 75.64% |
| Test ASE | 27.11% | 70.76% |
| Test ROC | 70.77% | 63.33% |
| Test CP | 96.4% | 96.4% |

As this was the first Apgar score prediction model to our knowledge, there is no previous model to compare with. Having gestational age, gest, the most important factor was highly expected as it is a very important indicator of the baby's health condition. The preterm baby, with gestational age of less than 37 weeks, could have more medical problems than a term baby. A question about maternal pain relievers arises here, as five out of six pain relievers are in the risk factors list. So, could the use of pain relievers seriously affect the newborn's condition? The only pain reliever that is not included in the above list was narcotics, which had the second least relative importance (table III). This may signify a safer use of narcotics as a pain reliever, which may not affect the newborn condition as much as other drugs. It was not a surprising result to find smoking, even if they mother smoked very little after 20 weeks gestational age, the seventh important factor in the list. Scalp and steroids were expected to be in the list since having these done to a newborn indicates medical problems. The most surprising non-medical factor was pf, the intention of the mother to breast feed, and how it became the fourth important variable in the list. So, can the mother's decisions about her future baby affect the baby's health? There are many questions that require more study. It is hoped that the list of Apgar's most important factors can help eliminate some of the risk factors to positively affect the newborns' condition and outcome.

## Acknowledgment

The authors would like to thank Dr. Paula Stewart for providing access to the perinatal database. Thanks to the Physicians Services Incorporated Foundation for a grant.

## References

[1] PPPESO strategic plan 2005. www.pppeso.on.ca/english/strategicplan_2005.pdf
[2] Frize M, Ibrahim D, Seker H,Walker RC, Odetayo MO, Petrovic D, Naguib RNG. Predicting Clinical Outcomes for Newborns Using Two Artificial Intelligence Approaches. Proceedings of IEEE Engineering in Medicine and Biology Society Conference; 2004 Sept. 1-5, San Francisco CA; 2004(2). p. 3202-3205.
[3] Ennett CM, Frize M, Walker CR. 2001. Influence of missing values on artificial neural network performance. Medinfo 10(Pt 1): 449-453.
[4] Walker RC, Ennett CM, Frize M. 2003. Use of artificial neural networks to identify the contribution of different physiologic variables to neonatal mortality in newborn infants admitted to neonatal intensive care units. Pediatr Res 53: 444A.
[5] Frize M, Ennett CM, Hebert P. 2001. Improving the efficiency and effectiveness of artificial neural networks as decision-support systems." Medinfo 10(Pt 1): 574.
[6] http://www.alexian.org/progserv/babies/thirdtri/apgarscore.html
[7] Apgar V, "A proposal for a new method of evaluation of the newborn infant," Curr Res Anesth Analg, vol. 32, pp.260-267, 1953.
[8] Frize M, Enntt CM and Charette E, "Automated Optimisation of the Performance of Artificial Neural Networks to Estimate Medical outcomes," Proc. 3rd ITAB Conference (Information Technology Applications in Biomedicine) and ITIS (International Telemedical Information Society), pp. 168-173, Washington, Nov 2001.
[9] Ennett CM, Frize M., Charette E, "Improvement and automation of ANNs to estimate medical outcomes," Medical Engineering &Physics, vol. 26, pp. 321-328, 2004.
[10] Scales N, "Mortality prediction in an ICU and coronary surgery database using neural networks," Term project, Department of Systems and Computer Engineering, Carleton University, Dec 2001.
[11] Ennett CM, Frize M, Scales N, "Logarithmic-sensitivity index as a stopping criterion for neural networks," Proc IEEE EMBS-BMES, Texas, 2002.
[12] Ennett CM, Frize M, Scales N, "Evaluation of the logarithmic-sensitivity index as a neural network stopping criterion for rare outcomes," Proc IEEE ITAB, 2003.
[13] Penny W, Frost D, "Neutral networks in clinical medicines," Medical Decision Making, vol. 16, pp. 386-398, 1996.
[14] Garson, G.D., "Interpreting neural-network connection weights," Artificial Intelligent Expert, vol. 6, pp. 47–51, 1991.
[15] Goh, A.T.C. "Back-propagation neural networks for modeling complex systems," Artif. Intell. Eng., vol. 9, pp. 143–151, 1995.
[16] Olden J., Jackson D., "Illuminating the 'black box': a randomization approach for understanding variable contributions in artificial neural networks," Department of Zoology, University of Toronto, Ecological Modelling, 154:135–150, 2002.