An Artificial Neural Network Derived Trauma Outcome Prediction Score as an Aid to Triage for Non-Clinicians

Adrian PEARL, Raphael BAR-OR and David BAR-OR
Swedish Medical Center and St Anthonys Central Hospital Trauma Research Lab.
Englewood, Colorado. USA

Abstract: In mass casualty events Emergency Medical Service Providers (EMS) choose treatment at Scene or a “scoop and run” approach. The latter requires clinically trained personnel at the reception site to triage patients. Current methodology based on Revised Trauma Score (tRTS) requires use of Glasgow Coma Scale, a method reliant on experience and clinical knowledge. This makes the system subjective and often inadequate for non-clinicians. This project attempts to develop a simplified outcome prediction score using an artificial neural network for use by a non-clinically trained EMS to aid triage. The project uses National Trauma Data Bank, Version 6.1. Tiberius Data Mining Software created Neural Network models. Variables considered were values that could easily be obtained during an event. Binary values were used for low SBP and low Respiratory Rate, coded using the RTS scoring table as a basis, and age indicators. A modified motor component of Glasgow Coma Score was created to negate the need for clinical knowledge. Best performing models, identified by Gini coefficient and ability to predict mortality, were with 8 and 10 neurons. On mortality prediction all even numbers of hidden neurons have similar performances. Training sets were compared to test sets, and found to be identical in Gini coefficient and performance. Models performed well in predicting mortality compared to standard outcome predictors. Possible additional variables such as gender or ethnicity might improve the Neural Network predictive ability. Pulse appears an essential variable not recorded by the NTDB.

Keywords: Artificial Neural Networks, mortality prediction, triage

Introduction

When there are multiple casualties in an incident, Emergency Service teams have to choose between conflicting protocols. They either provide treatment at the scene, performing triage and sending the serious casualties first, or operate “scoop and run” in which they pick-up casualties regardless of injury detail and triage as they arrive.

Triage principles whether performed in the field or at the receiving facility remain constant. Its role is to classify patients according to their medical need. With Field triage not only is it important to assess the actual severity of the injury, but location and distance to the local healthcare facilities must be considered. Additionally the resources at the disposal of the facility may affect triage decisions.
Following a traumatic injury there is a trimodal pattern to the deaths that occur\(^1\). Immediate death occurs within minutes of injury, often due to vascular or neurological damage. Treatment rarely improves outcome. The second peak occurs during the “golden hour”. Death usually occurs due to intra-cranial haematoma, or major thoracic or abdominal injury\(^2\).

Since time is critical, field triage requires a simple method to sort the injured. Standard triage is based on the Revised Trauma Score (RTS), using a simplified coded version - tRTS. It is determined by application of clinical experience to assess Glasgow Coma Score\(^3\) (GCS) and hard to apply for non-clinicians. Reliance on GCS, as has previously been argued\(^4\), creates a subjective system, dependent on interpretation, in which clinicians assess patient conditions. This dependency restricts the ability to obtain high correlations between injury and outcome predictions\(^5\). Since only clinically trained personnel could accurately determine true GCS, precious time can be lost waiting to assess injury and triage order. The Emergency Medical Services (EMS) are thus pressured into either adopting a ‘scoop and run’ policy or to ‘treat at scene’ until evacuation orders are arranged. If a new method of triage could give the EMS the ability to predict mortality at mass casualty events, it could ease pressure on Hospital Emergency Rooms while improving the survival chances of those injured.

There are a number of systematic protocols deployed to aid in first response triage, guiding injury selection based on observation of three primary functions; respiration, pulse and mental state\(^6\). These protocols are known by the acronym START. The variables form the basis of a simplified trauma outcome score we are attempting to design using application of artificial neural networks (ANNs) to establish predictive models to remove GCS scores from calculations\(^4\) in mortality prediction.

The use of ANNs in outcome prediction has become increasingly prevalent in physiological modeling. Mathematical models constructed on the basis of organic neural systems, these ANNs are flexible systems which are increasingly used in predictive modeling due to the ability of the ANN to learn and improve. This occurs by a methodology known as feed-forward/back propagation, in which the artificial neuron adds weights according to positive or negative deviation from a training set of data from which it ‘learns’.

Most models use large numbers of variables, and do not align themselves for triage application. Others are determined from low volumes and restrict the ability of the model to accurately train and test the model. Our own previous work used low volumes with three measured physiological variables plus binary variants of them. Its resultant models were not as effective at low levels of sensitivity and specificity, and conclusions postulated that by using increased volume might improve the model\(^4\).

### 1. Materials & Methods

#### 1.1.1. Patient Population

Study data was taken from the USA National Trauma Data Bank (NTDB), version 6.1. Variables extracted from the Registry included patient demographics, Scene and ED physiological variables, and the Revised Trauma Score (ED_RTS) & Revised Probability of Survival (RPS) scores for comparison against the ANN derived model. There were no exclusion criteria.
1.1.2. Creating New Variables for analysis

Neural Networks are good for classification problems, working best with data in a binary format. New variables thought to affect survival rates were created, and coded based upon the categories of the RTS, age predictors used in TRISS – Trauma Injury Severity Score, and ability to obey simple commands. These variables will be used to test for improved performance. The following variables were created as defined below:

- LowSBP – Systolic Blood Pressure less than 40
- LowRR - Respiratory rate shallow (less than 10)
- OCmd – Ability to obey simple commands: 0 = mGCS 1-5; 1 = mGCS=6
- PedAge – Age of patient class for under 16 years: 0 = >16; 1= ≤ 16
- ThirdAge – Age of Patient greater than 55 years: 0 = ≤ 54; 1= >54

1.2. Development of the Artificial Neural Network

This study used Tiberius software (www.philbrierley.com), operating multilayer perceptron (MLPs) methodology to create ANN model algorithms. The process consists of two steps; the forward pass, where predicted outputs corresponding to given inputs are evaluated, and the backward pass, where partial derivatives are propagated back through the network. The chain rule of differentiation gives very similar computational rules for the backward pass as the ones in the forward pass. The network weights can then be adapted using any gradient-based optimization algorithm. The whole process is iterated until the weights have converged.

The neural network was designed using seven input variables (three measured, two binary calculated and two age related) and one output variable. It consists of three layers, one being a hidden layer of neurons. The numbers of neurons within the hidden layer affect the number of degrees of freedom in the optimization process, and therefore the performance of the model. Adding extra neurons increases the non-linearity of the model. Therefore a higher value can be used to extract a more complicated feature. For this project 85% of the data (1,217,125) were randomly selected as a training set, and 15% (215,899) were used as the test set. There were 5011 records excluded due to missing outcome.

1.2.1. ANN design

Initial model design was based on that previously constructed for analysis of SMC data for eight input variables, using the respiratory rate, systolic blood pressure and mGCS components as recorded at Scene. Age variables were encoded as outlined above. Pulse was missing from the NTDB dataset, so only seven input variables were used for this model. Initial design started with 5 neurons in the hidden layer, and extra neurons were added to identify the best model as identified by the Gini co-efficient, and the predictive performance of the model. The Gini coefficient is a measure of equality, and can be employed as a means of comparison between ANN models.

Each model design was allowed 100,000 epochs minimum in which to identify the best training algorithm before adding an extra neuron to the hidden layer. The training sets were compared in each design case to their test set, and found to be identical in Gini co-efficient and performance.
1.2.2. Statistical Analysis of Performance by Discrimination (ROC curve analysis)

Discrimination is the ability of the model to separate the population into two groups. In this instance we discriminate between those who live or die. Receiver Operator Curves (ROC) are independent of outcome prevalence, and are a useful tool in the performance evaluation for separating two populations. An ROC plot is the graph of all observed (1-specificity, sensitivity) pairs. Each point on this empirical plot can be represented by a 2x2 contingency table. Two different tests on the same patient can then be compared.

1.2.3. ANN Model Analysis

Analysis of model design was by comparison of the performance in mortality prediction and survival prediction. The coefficients calculated by the ANNs were then used to calculate the probability of survival according to the following equation:

\[ P_{\text{survival}} = \frac{1}{1 + e^b} \]

where \( b \) is calculated from:

\[ b = b_1(\text{lowRR}) + b_2(\text{RR}) + b_3(\text{SBP}) + b_4(\text{lowsbp}) + b_5(\text{OCmd}) + b_6(\text{3rdAge}) + b_7(\text{PedAge}) \]

Having determined the probability of survival (mortality prediction) the values, identified as PDT, were then plotted in an ROC graph for comparison by model and with standard trauma outcome measurement scores.

2. Results

Initial design started with 5 neurons in the hidden layer. The best performing models were identified by their Gini co-efficient and ability to correctly predict mortality. The training sets were compared in each design case to their test set, and found to be identical in Gini co-efficient and performance. Results are shown in Tables 1 and 2.

Table 1: Comparison of Gini Coefficient & RMS error in ANN models

<table>
<thead>
<tr>
<th>Nos Neurons</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini</td>
<td>0.6204</td>
<td>0.6228</td>
<td>0.6284</td>
<td>0.6153</td>
<td>0.6187</td>
<td>0.6180</td>
<td>0.6228</td>
<td>0.6192</td>
</tr>
<tr>
<td>RMS Error</td>
<td>0.1997</td>
<td>0.1997</td>
<td>0.1999</td>
<td>0.1999</td>
<td>0.1997</td>
<td>0.1998</td>
<td>0.1997</td>
<td>0.1998</td>
</tr>
</tbody>
</table>

The best performing models, as identified by their Gini co-efficient and ability to correctly predict mortality were constructed with 8 and 10 neurons, though on mortality prediction all the models with an even number of hidden neurons have similar performances. None of the models performed as well in predicting mortality when compared to using the model that included values for patient pulse.

Table 2: Comparison of Predictive performance in ANN models

<table>
<thead>
<tr>
<th>Nos Neurons</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Alive</td>
<td>72.4</td>
<td>85.3</td>
<td>71.5</td>
<td>85.0</td>
<td>72.7</td>
<td>83.0</td>
<td>73.4</td>
<td>85.7</td>
</tr>
<tr>
<td>% Died</td>
<td>78.2</td>
<td>65.7</td>
<td>79.8</td>
<td>66.1</td>
<td>77.7</td>
<td>69.4</td>
<td>77.5</td>
<td>65.2</td>
</tr>
</tbody>
</table>
2.1.1. Statistical Analysis of Performance by Discrimination (ROC curve analysis)

For each of the best performing models (pdt8 & pdt10, 8 and 10 neurons respectively) probability statistics were calculated and then the ROC graph plotted, with the ED Revised Trauma Score and the TRISS Probability of Survival included as comparisons. In addition a five neuron model (pdt5) is included since previous research indicated that a five neuron model worked well. In comparisons it is clear that the models work as well as current outcome predictors in mortality prediction.

![Receiver Operator Curve](image)

Figure 1. Receiver Operator Curve for artificial neural network models and comparison mortality predictors

3. Discussion

Using the greater volume of data improves the ability of the ANN to train, and therefore apply its predictive equation to a test set. However, when compared with the performance of the original model determined from the SMC registry data, with a Gini co-efficient of 0.493 and RMS error of 0.186, we can see that predictive accuracy is lost when excluding the patient pulse from the equation.

The original intent of the project was to create a model that could easily be applied by non-clinicians to more accurately predict mortality in a triage situation without requiring lots of variables to be taken into consideration. This analysis shows that while increased numbers improves the models ability to train, decreased variables in the model reduces the accuracy.

Using the NTDB dataset could be detrimental, as large volume of data may cause the ANN to loose important trends that could aid in mortality prediction. It may be that
the size of the population cohort used to create an artificial neural network has an optimum level.

4. Conclusion

Previous ANN-based studies conducted used population-based registries, with large numbers of inputs\(^1, \, 12\). This study concentrated on a smaller number of factors, previously considered, to give good correlation in survival prediction, to create an efficient simple method for ‘field triage’. The method was designed to require no application of clinical knowledge, and relate directly to physiological characteristics that could be measured by emergency service personnel.

Compared to the current outcome predictors, these models performed well in mortality predicting. However when compared to the model which included the patient pulse, as previously reported\(^4\), none of the models performed as well. The pulse appears to be an essential variable in this model and needs to be recorded in the NTDB.

Further work using multi-center data to provide sufficient sample sizes that include pulse should be performed to try and improve the methodology. Possible additional variables such as gender or ethnicity might further improve the Neural Network predictive ability. Additional studies may need to be conducted to identify if population cohort size affects the ability of Neural Networks to predict mortality, and that systems designed on too small as well as too large a population dataset mask important trends.

References