ABSTRACT
Traumatic pelvic injury is one of the most dangerous injuries because it is often associated with severe hemorrhage as well as serious complications. It is therefore vital to provide immediate medical treatment to increase the survival rate of pelvic injury patients. However, it is often difficult to make treatment decisions, as cases are complex and display similar patterns. It has been suggested that the use of computer aided decision-making in a trauma support system is the most efficient way to reduce the cost of trauma care. In our previous work, we found that creating rules using all available variables results in lower accuracy than when using only significant variables. This is because less relevant attributes and/or less reliable attributes with regards to the means of measurement can result in random correlation that is clinically meaningless. Based on this knowledge, we designed an efficient computer assisted trauma decision making system for traumatic pelvic injuries using a machine learning algorithm. More specifically, a rule-based system was designed to create a reliable method of making predictions/recommendations on the status and exact outcome – i.e. home or rehabilitation - of pelvic trauma patients using a nonlinear regression and classification (CART) method. The resulting computer aided system can aid physicians in making rapid and accurate decisions. Three machine learning algorithms were compared to evaluate the proposed method.

Keywords
Computer-aid decision making, CART, Pelvic trauma injury

INTRODUCTION
Pelvic fracture is a dangerous injury which causes a variety of complications that often require long-term rehabilitation [8]. A report on childhood injury in motor vehicle crashes conducted by National Highway Traffic Safety Administration (NHTSA) states that the second most common injury diagnosis in frontal impacts is internal abdomen or pelvic injury, occurring in 58% of recorded cases [19]. The possible complications of undiagnosed hemorrhage result in a high mortality rate, and potential severe and permanent disability in survivors. Diagnosis requires a variety of medical information, in particular CT and MR images, and treatment requires the expertise of highly skilled trauma surgeons in assessing the patient's condition according to all available patient data [5]. The increasing cost of trauma care is also regarded as another reason for crowded trauma units across the United States. Furthermore, due to the need for rapid decisions by physicians on patient treatment, errors occur. In a recent study on error rates in medicine [1], a database was constructed using a 1992 dataset provided by the American Hospital Association. In total 1,116 hospitals participated, and the study found 17,338 medication errors that adversely affected patient outcome.

Decision-making in traumatic injury cases is therefore an extremely challenging task, due to the complex nature of patient medical information, the stressful work environment of the trauma unit, and the diverse injury types. Computer-aided systems can significantly improve patient care by using all available clinical information to offer a reasonable recommendation on a course of treatment. Consequently, they reduce the cost of care and improve its outcome for patients. However, even though several computer-based decision making systems have been designed for trauma medicine, none are commonly used in trauma centers for three main reasons:

1) Frequently, these systems apply non-transparent methods such as neural networks and support vector machines that fail to provide the reasoning behind the decisions/recommendations made by the automated system. 2) Important patient data - such as CT scans and physiological signals - is not considered or not optimally analyzed during the decision making process, while
some irrelevant patient data is fed to the classifier which compromises the applicability of the resulting rule-bases. 3) There is no comprehensive database integrating all relevant and available data for particular medical decision making processes [12].

Many studies have developed an easy and simple predictive model for survival of trauma brain patients. Signorini [21] describes a simple model containing variables such as age, Glasgow Coma Score (GCS), Injury Severity Score (ISS), pupil reactivity, and presence of hematoma on CT to predict the survival rate. Although this model is efficient to use, the small number of variables considered may limit the reliability of the generated rules. Rovlias [20] uses CART to predict brain injury using only variables which are easily and rapidly obtained, such as age, gender, blood pressure, GCS, computed tomography scan data, and intracranial pressure (ICP). He emphasizes the importance of GCS in outcome prediction, and states that CART is a useful tool to visually predict the severity of brain injuries. Recently, logistic regression and radial basis function neural network have been compared to predict the mortality rate for traumatic brain injury [11, 15]. Results suggest that neural networks perform better than logistic regression. However, neural networks are not sufficiently accurate or statistically reliable, and the knowledge stored in trained networks is not transparent. As a result, physicians cannot examine the reasoning behind the recommendations of the computer aided system [9, 22].

Machine learning algorithms such as CART are commonly used in medical informatics because they are known to deal effectively with missing values as well as categorical variables. They are also able to extract rules from data directly and, unlike neural networks, provide physicians with the reasons for the decisions generated by the computer algorithm. However, CART has shown limited success in separating and identifying significant variables in rule creation. Logistic regression, by contrast, does not assume linear relationship or normal distribution of the attributes, and is useful in finding statistically significant regression coefficients for modeling tasks with binary outcomes. We therefore perform statistical analysis using logistic regression for this phase of rule extraction [7] in our previous work [12]. As our previous work demonstrates [12, 13], combining machine learning and logistic regression can increase understanding of data patterns and consequently generate the most reliable rules. Creating rules using all variables may not be optimal for medical applications, because less relevant attributes and/or less reliable attributes with regards to the measurement methods used can result in random correlation that is clinically meaningless. We found that using direct maximum likelihood estimation (direct MLE) to select significant variables has better accuracy than forward or stepwise model selection methods. We discuss direct MLE in more detail in the methodology section.

In this paper, a rule-based computer aided system is proposed to predict the exact outcome, home or rehabilitation, of pelvic trauma patients using nonlinear regression methods (CART). Three machine learning algorithms such as AdaBoost, CART, and RBF neural network are compared. We also perform quantitative measures over the statistical reliability and accuracy of the predictions/recommendations for traumatic pelvic injuries.

The next section describes the methodology, beginning with a description of the dataset used in this study. Subsequent sections explain the algorithm used, and provide results and conclusions.

**METHODOLOGY**

For the purposes of this paper, it is assumed that there is potentially a significant difference between the applicability and usability of the resulting medical rule-bases when using all variables or only significant variables. When all variables are used, many of the rules are usually neither meaningful nor applicable to real-world medical diagnostic applications. Therefore, as proposed by our previous study [12, 13], significant variables are extracted via direct maximum likelihood estimate. Logistic regression [10] performs the significance test for each variable by comparing the log-likelihood ratios for the model with and without the variable using chi-squared. Two of the most popular logistic regression methods to test variable significance are stepwise and forward model selection. The stepwise model selection method performs repeated addition/deletion until certain criteria such as Akaike information criteria (AIC) or Bayesian information criteria (BIC) are satisfied, while the forward method keeps adding variables [17]. However, both methods are time consuming, so we directly use the results of log-likelihood ratio using chi-square to select significant variables. We call this method direct maximum likelihood (direct MLE). As previously mentioned, our prior work [12] found that direct MLE has slightly better accuracy than forward and stepwise selection.

It is commonly stated that Glasgow Coma Score (GCS) and Injury Severity Score (ISS) are significant factors in predicting the outcomes of trauma patients [4, 5, 15, 20], and the treatment advised by experts. Thus, we include those two factors as significant variables. Ten-fold cross-validation methods are applied ten times to test the validation. The dataset is randomly divided into ten mutually exclusive subsets, and in each training-testing attempt nine partitions are used as the training dataset, with the remaining partition used as the testing dataset. Repeating the entire process ten times generates more accurate rules. After each repetition, a number of rules are associated with the decision making system based on the resulting tree. This paper focuses on predicting the exact outcome of traumatic pelvic injuries, i.e. whether the patient is sent home or
to rehabilitation, using a combined approach: logistic regression (for filtering of non-significant attributes) and classification methods (for rule extraction).

Dataset and its use

The traumatic pelvic injury dataset used in this study was created in collaboration with Carolinas Healthcare System and Virginia Commonwealth University. The database contains not only demographics and physiological data but also measurements such as the Glasgow Coma Score (GCS) and Injury Severity Score (ISS), standard metrics that are frequently used in assessment of injury severity. These scores can be taken either at the site of the accident or at the hospital. Any treatments - for example, blood transfusions - and all medical tests and procedures performed at any stage are also included in the database. Table 1 shows all the available variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable</th>
<th>Variable</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>PRE-HOSP EYE</td>
<td>SCENE RESP</td>
<td>CHEST CT</td>
</tr>
<tr>
<td>SEX</td>
<td>PRE-HOSP FLUIDS</td>
<td>SCENE REV TS</td>
<td>HEAD CT</td>
</tr>
<tr>
<td>CHIEF COMPLAINT</td>
<td>SCENE GCS</td>
<td>ED GCS</td>
<td>ISS</td>
</tr>
<tr>
<td>POSITION</td>
<td>SCENE MOTOR</td>
<td>ABD. U/S</td>
<td>Diagnosis type</td>
</tr>
<tr>
<td>SCENE BP</td>
<td>SCENE</td>
<td>NEEDLE</td>
<td>ABD CT</td>
</tr>
<tr>
<td></td>
<td>THORAX</td>
<td></td>
<td>SCENE PULSE</td>
</tr>
</tbody>
</table>

Table 1. Summary of all available variables

As Table 1 shows, there are twenty variables in the pelvic injury dataset. The dataset includes all possible scene and ED values. The dataset we use includes all possible informative records except signal and image information. The variables in Table 1 are fed to the model as the input to extract significant variables, and the exact outcome is treated as the output of the model. The output is grouped into two classes: the patient is sent home, or to rehabilitation. A total of 681 cases, including 381 home cases and 300 rehab cases, are used for this study. In order to predict the outcome, only alive patients’ data are used.

METHODS

Significant variables extraction: Direct maximum likelihood estimation in logistic regression is used to discover the significant variables in outcome prediction. Statistical analysis software (SAS) is used to perform the significance test. The intention behind maximum likelihood parameter estimation is to determine the parameters that maximize the probability (likelihood) of the sample data. It therefore has optimal properties such as consistent and efficient parameter estimation [18]. In addition, it provides efficient methods for quantifying uncertainty through dataset confidence. These advantages make direct MLE in logistic regression suitable for finding significant variables.

Data validation: Ten cross validation is performed to measure the generalization quality and scalability of the rules. The cross-validation method investigates whether similar outcome distribution is demonstrated in each of the ten subsets of the data.

Rule Extraction: Rules are extracted using CART, because it is able to creating rules visually; in other words, it provides transparent rule generation. Once trees are constructed as a result of CART, the symbolic rules containing variables can be directly interpreted and compared with existing biological knowledge, providing useful information for the clinicians. As a result of ten cross validation, ten trees are generated and a variety of rules are extracted. Each individual rule extracted from each tree is tested.

Comparison machine learning algorithms: Three machine learning algorithms are compared: AdaBoost, classification and regression trees (CART), and RBF neural network. AdaBoost, introduced by Yoav Freund and Schapire [6], is an iterative process algorithm for constructing a robust classifier as a linear combination of weak classifiers with a voting scheme. The main objective of a neural network is to transform the input into meaningful outputs, and Radial Basis functions (RBF) are particularly well suited for solving pattern classification problem due to the resulting network’s simple topological structure and rapid learning ability. An RBF neural network has an input layer, a hidden layer and an output layer [14, 23]. CART, designed by L. Breiman [2, 3], applies information-theoretic concepts to create a decision tree, allowing it to capture rather complex patterns in data and express them in form of transparent grammatical rules [16]. Even though CART is an older
algorithm, its nonlinear extensions are still extended and widely used in data mining and machine learning because of its ability to deal with missing data and categorical variables. Another advantage is that CART can process multiple data types such as numerical and categorical variables [7].

**Performance of Rule Quality:** The performance of each rule is measured as the probability of correctly case predication using the accuracy measure, \( \text{acc}_R = \frac{\text{Po}(R)}{\text{Po}(R) + \text{Ne}(R)} \). In other words, the accuracy of each rule is measured with its number of positive matches (\( \text{Po}(R) \)) divided by the total number of positive matches and negative matches (\( \text{Po}(R) + \text{Ne}(R) \)). Both \( \text{Po}(R) \) and \( \text{Ne}(R) \) are evaluated as follows: assuming that \( D \) is a given dataset including the instance \((x_i, y_i)\), \( D_r \) is the training set, \( D_t \) is the testing set where \( D_t \in (D \setminus D_r) \), and \( y_j \) is the outcome. \( I(D_t, w_i) \) indicates the matching survival classification of testing class \( D_t \) on the dataset using an induction algorithm (nonlinear regression/classification tree and C4.5), i.e:

\[
\text{eval}(R) = \sum_{(y_i, w_j) \in D_t} \delta( I(D_t, w_i), y_j )
\]

where \( \delta(i, j) = 1 \) if \( i = j \) (\( \delta(i, j) = 0 \) otherwise). Once a reliable rule based system is created, sensitivity and specificity are measured to check how successfully the positive cases and negative cases are identified.

**EXPERIMENT RESULTS**

The results presented here are the rules created using logistic regression combined with CART. Only the rules with high accuracy (> 85%) were incorporated into the rule base. Since a small number of examples were used in training (less than 800), the rules were tested using the entire dataset. The transparent nature of the rule-based system used in this study means that the reliable rules can be used as an assistive system not only to help pelvic injury experts predict the exact outcome for new patient, but also to provide the reasons behind the predictions/recommendations.

**Significant Variables**

Based on the direct MLE estimation in logistic regression, we found that age (p value<0.0001), pre fluids (p value=0.0317), and chest CT (p value=0.0311) are most significant. Chest CT has three values: positive, negative, and not performed. Pre fluid has four values: <500, 500-2000, >2000, and not performed. GCS value of ED and ISS are also added as significant variables.

**Comparisons**

In this section, the results of three machine learning methods are compared. The comparison was done using only significant variables, since this generates highly accurate and meaningful rules. Table 2 shows the resulting accuracy of each method on the testing set. It is clear that there is little variation in the accuracy of the methods, indicating that the variable selection is efficient and appropriate. It was found that CART has the highest accuracy in predicting exact outcomes using the pelvic dataset.

<table>
<thead>
<tr>
<th>ML Method</th>
<th>AdaBoost</th>
<th>CART</th>
<th>RBF Neural network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing accuracy</td>
<td>68%</td>
<td>69.7%</td>
<td>67.8%</td>
</tr>
</tbody>
</table>

Table 2. The results of machine learning comparisons

The required accuracy threshold was determined via discussion with physicians, and consequently the rules with >85% accuracy are determined as reliable. Table 3 presents the reliable rules generated by using CART with significant variables. Twenty-one reliable rules were generated, with a total accuracy of 91.2%. Rule sensitivity and specificity are 90.2% and 89.6% respectively. These high values indicate that the rules constructed from the pelvic injury dataset are well classified. The following rules are selected after discussion with medical experts.

<table>
<thead>
<tr>
<th>Pelvic injury Exact Outcome Rules</th>
<th>Testing Accuracy</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE:=46.50 &amp; EDGCS:=&lt;=13 &amp; ISS:=&lt;=14 &amp; (PRHFLU='TVF Unk Amount' or PRHFLU='&gt;=2000') Then rehab</td>
<td>11/12(91.7%)</td>
<td>CART</td>
</tr>
<tr>
<td>EDGCS:=13 &amp; (CHE_CT='Not Performed') &amp; ISS:=&lt;=9 &amp; AGE:=&lt;28.65 Then Home</td>
<td>31/31(100%)</td>
<td></td>
</tr>
</tbody>
</table>
CONCLUSION & FUTURE WORKS

The aim of this paper is to develop an efficient computer-aided rule-based system that helps physicians make accurate decisions. To achieve this goal, we introduced a method which generated high quality rules using combined CART and logistic regression. The most significant variables are statistically determined using direct MLE, and CART is used to generate variety of reliable rules that are easily understood by physicians. The resulting computer-aided system supports physicians in making fast and accurate decisions. In addition physician can examine the reasoning behind the recommendations. The algorithm can also help with decision making in rural and remote areas where physicians with extensive training may not be available. The results in this paper provide a framework to improve the physicians’ diagnostic accuracy with the aid of machine learning algorithm. Optimal resource utilization and rule-based decision-making offer the best way of assisting physicians in providing optimal care for traumatic injury patients. Diagnostic decisions can be made for future patients by comparing all possible rules associated with their symptoms. For the future work, other ML algorithms such as C4.5 will be applied and compared with CART in order to generate more reliable computer-aided rule based systems. Comparing two methods can provide better choices of classifiers or better opportunities to combine the resulting rules.

ACKNOWLEDGMENTS

This research was partially funded by grants from Health System Organization, Carolinas Healthcare System, and Virginia Commonwealth University.

REFERENCES