Author's response to reviews

Title: A Comparative Analysis of Multi-Level Computer-Assisted Decision Making Systems for Traumatic Brain Injuries

Authors:

Soo-Yeon Ji (jisy@vcu.edu)
Rebecca Smith (smithr@vcu.edu)
Toan Huynh (toan.huyhn@carolinashealthcare.org)
Kayvan Najarian (knajarian@vcu.edu)

Version: 4 Date: 19 September 2008

Author's response to reviews: see over
Over the last 20 years, the development of computer decision support systems has been a growing field in the research medical domain. Trauma experts must make several difficult decisions considering a large number of patient attributes, often in a short period of time. These decisions are made at different stages of care giving, based on the available patient data at the time of each decision. Computer-aided systems can provide vital assistance to clinicians in deciding both diagnosis and treatment, and so provide a higher standard of patient care. This paper focuses on current medical issues, and proposes a predictive computer-assisted decision making system for traumatic head injury (TBI) using machine learning algorithms. Machine learning provides a good solution for dealing with missing values, a critical issue when working with medical data.

Thus, the main goal of this study is that develop a reliable rule-based decision making system that provides both recommendations and outcome predictions at critical stages of care giving and compare the existing machine learning methods available for medical informatics. We also suggest that using all available variables in rule generation may not be optimal for medical applications. The inclusion of less relevant and/or reliable attributes, with regards to some predefined means of measurement, can create random correlation in some rules that are ultimately clinically meaningless. Furthermore, retaining less informative and/or highly correlated attributes increases the length and complexity of the resulting rules, and so makes them less desirable for clinical application. To support this assertion, we compare five machine learning algorithms using both the all-available-variables and significant-variables-only methods, and then generate the most reliable rules using CART and C4.5, which are chosen due to their transparency. Sensitivity and specificity are also calculated to evaluate outcome identification.
Thank you for your suggestions to revise the manuscript. We greatly appreciate your feedback, and will address each comment in turn.

1. **Provide a clear rationale for the study**

As suggested, we have added our hypotheses to the paper introduction. This clarifies the rationale for our study, and why we believe the study is important.

1. We hypothesize that a rule-based system, attractive to physicians as the reasoning behind the rules is transparent and easy to understand, can be as accurate as "black-box" methods such as neural networks and SVM.
2. We hypothesize that when trained correctly, a computer-aided decision making system can provide clinically useful rules with a high degree of accuracy.
3. Based on our hypothesis that appropriate feature selection may have a critical impact on increasing prediction accuracy, we examined which variables are most significant in the recommendation/prediction making process. As a result, accuracy performance is improved to predict survival outcomes using significant variables such as age, pre-existing diseases, and complications in our study.

2. **Improved Description of MLE**

We have also improved the description of MLE. It now reads as follows:

*To test the significance of the individual model parameter, logistic regression is used log likelihood test. The likelihood ratio test itself does not tell us if any particular independent variables are more important than others. However, we can analyze the difference between for the full model and a nested reduced model which drops one of the independent variables. The test takes the ratio of the maximized value of the likelihood function for the full model (L1) over the maximized value of the likelihood function for the simpler model (L0). The resulting likelihood ratio is given by:*

\[
\text{Likelihood Ratio} = \frac{L1}{L0}
\]

A non-significant difference indicates no effect on performance of the model, hence we can justify dropping the given variable. For our study, only the significant variables (p value $\leq .05$) are selected. A statistical analysis tool, in this case SAS, is used to calculate the significance of individual attributes. We call this directed MLE.

Note that forward and stepwise model selections are also available to discover the significance of variables. In the literature of statistical regression, the
A stepwise method is commonly used to find the best subset of variables for outcome prediction, considering all possible combinations of variables. However, the stepwise approach may not guarantee that the most significant variables are selected due to the repetition of insertion and deletion. For example, age may not be selected as an important variable; however, physicians may believe that patient age is important in deciding treatment. Therefore, we prefer to use directed MLE for our medical application. Our other justification for using MLE is empirical; in our previous study, we found that the direct MLE method has slightly higher accuracy in finding significant variables than stepwise and forward model selection.

3. **Clarify similarities and differences between methods and provide the requested ROC curves.**

As requested, we have also included ROC results – tables comparing the AUCs (Table 1 and Table 2) for each of the methods we tested. This enables us to see the similarities and differences between methods in terms of performance, and how significant the differences are. The similarities and differences between the machine learning methods themselves are also covered in the paper. We also include an ROC curve plot for the logistic regression method; in the interest of brevity, we do not do this for the other methods.

4. **Provide the tests for the linearity assumption for each variable entered in the logistic regression model.**

As there are many variables involved, we have included the linearity assumption test results for two variables in the paper. We will now provide results for all variables across all datasets. This section also explains how the tests are conducted, and includes scatter plots and residual plots for logits and their predictors. We have added a shorter description of the method to the paper itself, to clarify our approach.

In order to check linearity assumption between logit and predictors, scatter plot and residual plot are used. Figure 1 presents the block diagram for the individual predictor test processing. This analysis was performed using statistical analysis software (SAS).
Once logistic regression is applied, the scatter plot between logit and predictors is presented to check linearity. In order to provide more accurate results, residual analysis is performed based on logit and each predictor using regression analysis. If the linear assumption is satisfied, we would expect the residuals to vary randomly – i.e. there would be no pattern. If the residual plot appears to form a curve, there may be a nonlinear relationship in the variable.

This document is organized as follows. First, we present the scatter plot and residual analysis results when using all available variables for survival prediction. This is followed by the same analysis using only significant variables for survival prediction. In addition, we present the results using the helicopter transport dataset. Figure 2 describes the organization of this document.

4.1. Survival Prediction: All Available Variables’ Linearity

Scatter plots of logit and each independent variable are presented in Figure 3. Since all the multi-valued pre-existing conditions and complications variables are replaced with dummy variables (Yes/No), we only choose one of them (Myocardial Infarction).
(a) AIS Head Score

(b) AIS Thorax Score
(c) AIS Abdomen Score

(d) EDEYE
(e) EDMotor

(f) EDVERBAL
(g) Safety

(h) Age
(m) ISS

(n) Gender
Figure 3. Scatter Plot between logit and predictor using all available variables (X axis indicates logit, Y axis indicates predictor)

Most of the scatter plots in Figure 3 show a linear relationship, except those for BP, EDRTS, and GCS. Thus, we perform regression analysis using logit and each predictor in order to verify the relationship with residual analysis. If the residual plot appears to form a curve, there may be a nonlinear relationship in the variable as mentioned above. Figure 4 presents the residual plots of logit versus predictor.
(b) AIS Thorax Score

(c) AIS Abdomen Score
Figure 4. Residual plots (X axis indicates predicted value of predictor. Y axis indicates residual)

Some of the plots in Figure 4 do not show random scatter. For example, there is a definite pattern with two vertical lines on the Gender plot (plot (n)). However, there are only two levels of gender and so there are at most two distinct predicted values, one for males and one for females. All females will have the same predicted value, and all males will have the sample predicted value.
These correspond to the two vertical positions on the plot. The scatter within each vertical line represents the variability of individuals in their survival (Alive/Dead) within their group.

BP (R square=0.0035), and FURR (R square=0.0145) variables do not show strong curvature pattern, and therefore may have weak linear relationships. These plots do not suggest violations of the linearity assumptions. Although GCS, ED EYE, and EDMOTOR residual plot also show a weak curvature pattern, they are not selected as significant variables in our final model.

4.2. Survival Prediction: Significant Variables’ Linearity

Head AIS score, Thorax AIS score, Insulin Dependent, Myocardial Infarction, Acute Respiratory Distress Syndrome (ARDS), Coagulopathy, Age, EDRTS, and ISS are selected as a significant variable.

Figure 5 show the scatter plot between logit and its predictor. As mentioned at the beginning of this document, pre-existing disease (Insulin Dependent, Myocardial Infarction) and complications (Acute Respiratory Distress Syndrome (ARDS), Coagulopathy) are replaced by dummy variables (Yes/No). We have demonstrated their linearity in Figure 3 and Figure 4; therefore, we do not show their results in this section in order to avoid redundancy.
(b) Head AIS score

(c) Thorax AIS Score
Most of the scatter plots in Figure 5 show linear relationship except Thorax AIS score, and EDRTS. Regression analysis using logit and each predictor is performed in order to verify the relationship using residual analysis. Figure 6 presents the residual plots of logit versus predictor.
(a) Age

(b) Head AIS score
Figure 6. Residual plot between logit and predictor using significant variables (X axis indicates predicted value of predictor. Y axis indicates residual).

The Figure 6 plots show the linearity of age, head AIS score, and ISS. However, Thorax and EDRT show a weak linear relationship.

If the plots of the residuals versus the predictors show curvature, a quadratic term should be tested for statistical significance, as this may suggest a better model. If the coefficient for this quadratic term is significant, the quadratic term should be included.

Even though our model does not show any strong curvature, we add a quadratic term to test significance. For the significance test, head AIS score (discrete variable) ISS (continuous variable), and EDRTS are tested. The model is \( \logit = \alpha + \beta x + \gamma x^2 \) where \( \alpha \) is an intercept term, \( \beta \) is a coefficient of predictor, and \( \gamma \) is a coefficient of squared predictor. For head AIS score, the estimate of \( \beta \) is -0.1820 (p value = 0.0015), and the estimate of \( \gamma \) is -0.0124 (p value = 0.2058). For ISS, the estimate of \( \beta \) is -0.0306 (p value = 0.0002) and the estimate of \( \gamma \) is 0.00008 (p value = 0.4607). These p values (> 0.05) indicate that the model does not need a quadratic term; therefore, the predictors show a linear relationship.

For EDRT predictor, the parameter estimation of \( \beta \) and \( \gamma \) is not significant when we are added quadratic term. However, the parameter estimation of \( \beta \) is significant (p value < 0.0001) when the quadratic term is removed from the model. This indicates that it may have weak linear relation with its logit.
4.3. Helicopter transport patients

According to our approach, age, BP, Head & Neck ISS, and ISS are selected as significant predictors. As mentioned in our paper, age, respiratory rate, pulse, and airway are selected by physicians because ISS and Head & Neck ISS are not measured at the scene. In the interest of brevity, instead of showing all results, we showed the result of significant variables (age, BP, GCS, respiratory rate, pulse rate, and airway) and some results of all available variables (Head & Neck ISS, and ISS). Figure 7 presents the scatter plots of the logits and their predictors.

(a) Age

(b) Airway
(e) Head and Neck ISS

(f) ISS
Most of the plots in Figure 7 demonstrate a linear relationship, although Age shows a slightly weak linear relationship. Thus, we do regression analysis using logit and each variable in order to verify the relationship using residual analysis. Figure 8 presents the residual plots of logit versus predictor. Airway result is not provided here since it is clear that it has a similar linear relationship to the Gender variable previously mentioned.
(a) Age

(b) Airway
Figure 8. Residual plot between logit and predictors

The residual plots demonstrate linear relationships, except for Head and Neck ISS. For the Head and Neck ISS residual plot, we do not see strong curvature. Thus, we add a quadratic term and perform a significance test as described above. The coefficient value of $\beta$ is 2.2610 (p value=0.0011) and the $\gamma$ value is -0.3727 (p value=0.0006). This indicates that the predictor needs a quadratic term. However, this predictor is not selected for our final model.
Based on overall analysis, most of the predictors show a linear relationship with the logit, and some show a slightly weak linear relationship. However, we do not think that these slightly weaker relationships violate the assumption of linearity.