Reviewer's report

Title: Applying data mining techniques to improve diagnosis in neonatal jaundice

Version: 2 Date: 17 May 2012

Reviewer: Andrew P Bradley

Reviewer's report:

This is quite an interesting paper which compares predictive clinical "rules of thumb" to those extracted via data mining. I am happy with the conclusions they draw, but request some additional information to strengthen their paper.

Major Compulsory Revisions

The experimental methodology is not clearly described:
1. please state exactly how many features were measured and what they were (say listed in an appendix)
2. please state how variables were eliminated and/or defined to be "similar"
3. more clearly define the 3 feature sub-sets, say in an enumerated list, stating how many features are in each sub-set
4. please give standard error for all your AUC measurements. You can also estimate this for previous literature using the method proposed by Hanley and McNeil in 1983 (in Radiology)
5. In the discussion I can not see how you can state that results are "substantially improved" or "statistically significant" without stating how you tested this hypothesis. please add this information.
6. The discussion should include either an interpretation of the models generated from the data mining (say the decision rules from Cart or J48) and/or a discussion of the trade off between using clinical rules of thumb and (slightly) more predictive "black box" rules.
7. Table 2 needs improving to highlight all of the highest performing models than are statistically equivalent, e.g., Bayes, Cart & Logistic on CRF features (as per 4 & 5 above)
8. Showing Se and Sp from different operating points is meaningless, e.g., on the full feature set Logistic has better AUC than Bayes, BUT Bayes has both better Se and Sp than logistic (the ROC curves clearly cross! and so AUC is not a good single number summary of performance). Please report the differing Sp values for a single (clinically meaningful) Se value. i.e., compare each classifiers Sp when their Se in (say) 90%.

Minor Essential Revisions
for an early treatment -> for early treatment
system, as -> system, has
put "one of the areas of machine learning" in brackets
contributing with new -> contributing new
as methodology -> as its methodology
move such as data mining to after explore new methodologies
do not -> does not represent any additional costs
better precision -> greater accuracy
brothers -> siblings
all hospital stay -> the complete hospital stay
average performance -> average performance on the test set
sensibility -> specificity :-)
please state ethics reference number
medium -> median
than the presented -> than those presented

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

i declare I have no competing interests