Reviewer's report

Title: Artificial neural network (ANN) velocity better identifies benign prostatic hyperplasia but not prostate cancer compared with PSA velocity

Version: 2 Date: 8 April 2008

Reviewer: Leif Peterson

Reviewer's report:

Major compulsory revision:

It was good to see a paper on ANN usage in PCa prediction. However, there were several issues that could be further described/developed.

1. The authors left out valuable descriptions on why they didn't use certain predictor variables when developing their artificial neural network model. What was the informativeness of histopathological scores such as biopsy Gleason, pathology Gleason, TNM, pathology stage, extracapsular extension, seminal vesicle invasion, lymph node involvement (yes/no), surgical margins, etc?

2. Please provide a table with statistical test results for the above variables using a k-group test such as F-test or Kruskal-Wallis for the 3 groups (i.e. k=3). What are the p-values for significant difference of the above across the 3 groups? How many missing data are there? Provide more discussion about variable selection prior to ANN implementation.

3. For the ANN models used, how many input nodes were there (don't use variable names). Rather, state e.g. "there were 6 input nodes based on the input variables x,y,z..."

4. How many hidden nodes were there in each ANN model?

5. How many output nodes in each ANN model (3?). This can be assumed, but it's better to say "there were 3 output nodes, one node for each diagnostic class."

6. What activation function was applied at the hidden nodes (logistic, tanh, etc.)?

7. What function was applied prior to output nodes (softmax, linear, etc.)?

8. How many epochs or sweeps were used per model, and what was the MSE stopping criteria? It is a good idea to provide a plot of MSE as a function of epochs (sweeps) for an example model.

9. How were the initial ANN weights set and what was their range?

10. Were the variables standardized prior to ANN usage? How was this done?

11. What were the values of the ANN learning rate and momentum?
12. Was the sample order randomly permuted before each training epoch so that the ANN training results are not biased toward the order in which the samples were arranged during training?

13. How were the samples partitioned during ANN training and testing. Leave-one-out cross validation? 10-fold cross validation? Hold out method, where e.g. 2/3 samples are used for training and 1/3 for testing?

Last, the data are purely phenotypic (clinical) and not mechanistic, since PSA is not a biological marker of cytokine signaling, apoptosis, cell survival, neovascularization, or invasion. Thus, the reported AUC values are understandable.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

No conflicts