Reviewer’s report

Title: Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model

Version: 1 Date: 7 March 2012

Reviewer: Patrick Tighe

Reviewer’s report:

Thank you very much for the opportunity to review this manuscript.

The authors should be commended on constructing a well-written manuscript that appropriately details an interesting project involving the use of an alternating decision tree to predict axillary lymph node metastasis in patients with primary breast cancer. The authors diligently describe a multi-stage approach, including their steps in preprocessing, model development using hold-out datasets, testing for accuracy and accounting for potential bias, and comparison with prior models.

Importantly, the authors should provide greater detail on the method of ADTree induction and Ensemble creation, and soften the strength of their conclusions.

Specific concerns are listed as follows. I believe each of these items can be easily remedied via rephrasing of content or inclusion of additional methodologic detail.

Major Compulsory Revisions:

Background.

1. In the third paragraph, the authors describe their use of the alternating decision tree. Why did the authors choose the ADTree instead of, perhaps, a Bayesian network or support vector machine? If the purpose of this project was to specifically explore the use of the ADTree, please say so. As currently written, the authors correctly point out the limitations of logistic regression-based classifiers, but fail to pivot on this argument and defend their use of the ADTree.

Methods

2. The authors very nicely describe the use of bootstrapping to account for bias, as well as their inclusion of a sensitivity and missing value analyses. More detail is needed for each of these accountings. For instance, were there constrains on the range of imputed values? Were averages, max-min, or other methods employed? How many bootstrap iterations were performed? Additionally, this reviewer is slightly confused on the total number of datasets generated; were 200 generated altogether, or 200 each for the bootstrap, sensitivity and missing value analyses for a total of 600?
3. The authors need to expand their description of several important features of the ADtree. For instance, how was the Ensemble generated? How many boosting iterations were performed with each run of the ADTree? How many nodes were expanded with each trial? (I may be confusing these items for the author’s inclusion of the number of nodes and trees, this is a big unclear.) How did the authors choose n=10-20 for number of nodes, random seed 1-10, and number of trees 2-20? Specific details on the implementation within Weka would be of interest. Similarly, the authors should include more description of the types of statistical comparisons used within the Methods section, including the rationale for a Kruskal-Wallis test.

Results

4. I would encourage the authors to use a bit softer language in describing their results. “The model successfully discriminated….at statistically significant levels” refers back to Figure 3. While Figure 3 does suggest statistically-significant differences in the probability of LN metastases in the LN- versus LN+ groups, there is considerable overlap in the probability distribution. This sentence should thus be reworded. Critically, the type of statistical test needs to be included, preferably with the mean and 95% confidence interval (or better yet, the mean difference with confidence intervals) for each comparison in this Figure.

Discussion

5. “...using the ADTree ensemble technique successfully predicted AxLN metastasis...” may be a bit too strong. I would consider softening this to read “....technique improved upon older models such as the MSKCC nomogram”

Minor Essential Revisions:

6. On page 12, “mammary grand” should read “mammary gland” I believe.

7. In Figure 3, each plot should be titled with its location as indicated in the Figure Legend.

8. Likewise, in Appendix D, each plot should be titled with its respective location.

Discretionary Revisions:

Background

9. In the first paragraph, 4th sentence ending with “...and could avoid this procedure” should be rephrased. Perhaps specify that it is the patients who could avoid this procedure?

10. In the second paragraph, the authors state that “tolerance against missing values” is “an important feature of the prediction model”. Which prediction model? The MSKCC nomogram? All classification systems? This paragraph would benefit from improved transitions and greater specificity of terms.
11. In paragraph 3, the authors may wish to use the term "machine learning" or "data mining", but perhaps not both.

Methods
12. The authors included sample populations from three distinct locations, but also from two distinct time frames. Differences in location were used for training, validation and testing. The authors may wish to mention whether there were any differences in screening, incidence, or testing over the two time intervals. Such differences may partially explain the fall-off in model accuracy with the test-set.

13. For readers who may not have a oncologic background, the authors may wish to reiterate that the “pathological findings from the surgical specimens” was the outcome variable which labeled each instance for classification purposes.

14. Within the Data Collection and SLN Biopsy section, the authors may want to comment on the possibility of different sensitivities for SLN detection across the three methods described. If the sensitivities differ significantly, the authors may wish to include an additional analysis whereby the location of surgery is included within the feature set, and a 10-fold cross-validation is conducted using all 465 subjects.

Results
15. I would rephrase the sentence “This indicates that these variables were more important than the other variables….". The sensitivity analysis simply suggests that the model was more sensitive to changes in these attributes, not necessarily that the variables themselves were more important. I state this because the sensitivity analysis altered each attribute independently, and it is unknown how each attribute correlated with other attributes in each, or the overall, sample.

Discussion
16. On page 13-14, the authors make a great point; the MSKCC nomogram and Russells Hall Hospital scoring systems, like many decision support tools generated from multiple logistic regression techniques, are considered “good” because they use a restricted number of variables. This also makes them “easy” for clinicians to apply by themselves. With the advent of electronic medical record systems and improved classification algorithms, however, the use of high-dimensional data may offer substantial improvements in accurate classifications. The authors may wish to expand on this difference.

Again, congratulations to the authors on a nicely written manuscript detailing a very interesting project.

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:

I declare that I have no competing interest.