Author's response to reviews

Title: Automated Detection of Regions of Interest for Microarray Experiments: An Image Texture Analysis

Authors:

   Bilge Karacali (bilge@drexel.edu)
   Aydin Tozeren (aydin.tozeren@drexel.edu)

Version: 2 Date: 7 December 2006

Author's response to reviews: see over
Dear Editor,

Please find attached a revised version of the article entitled “Automated Detection of Regions of Interest for Tissue Microarray Experiments: An Image Texture Analysis” by B. Karacali and A. Tozeren (5134365331133852).

Also included with this cover letter is our detailed response to reviewer comments. We are grateful to the reviewers and the editors of *BMC Medical Imaging* for providing us with insightful evaluation of our manuscript. We hope that you will find the revised manuscript appropriate for publication in the journal and look forward to hearing from you soon.

Sincerely yours,

Bilge Karacali
Detailed Response to Reviewer Comments

Reviewer: Anil Parwani

Major Revisions Requested: None was requested.

Minor Essential Revisions

The quality of the images is somewhat lower than expected. Please modify the images by improving their resolution. Some images appear to be out of focus.

We have regenerated the images using higher resolution settings to improve clarity. There may also be a slight discrepancy in image quality between the images supplied to BMC Medical Imaging and the images transmitted to the referees by the journal.

Discretionary Revisions

The machine learning algorithm was tested for three broad categories....More work needs to be done in terms of applying this algorithm to different tumor types, non-neoplastic conditions that mimic cancer and pre-neoplastic conditions that often coexist with the neoplastic cells.

That is correct. We added the following paragraph to the Conclusions section of our revised manuscript:

“The statistical learning algorithm developed in this study was tested with success for three broad categories of texture images observed in normal or diseased breast tissue. Validity of our automated method of identification of cancer- and normal-specific tissue image textures is yet to be illustrated on a large set of images gathered in a clinical trial study. The method presented is a first step towards automated identification of clinically relevant image textures for cancer. It is expected that the method will require further refinement and improvement as it is challenged with tissue images gathered from a much larger pool of breast tumors that may contain images of a variety of non-neoplastic and pre-neoplastic conditions. Here, we have clearly demonstrated that given a set of learning texture images from histopathology, it is possible to recognize with very good accuracy similar textures in other histopathology images of breast tissue. Further improvements of the algorithm must include its adaptation to recognize texture images in a wide variety of tumor types. In the analysis of comprehensive image subsets involving different types of malignancy and/or tumors of different organs, the parameter set used in this article (B, P, H) can readily be revised and enriched with additional texture parameters causing minimal change in the rest of the log-likelihood estimation algorithm.”

Reviewer: Vincenzo Della Mea

Major Revision Requested
A pitfall of the paper is given by the low number of cases in the dataset. It is true that the total number of images and the covered area is somewhat high, but the images come from a limited number of slides (14), which furthermore come from just 6 specimen...This suggests to me that the paper could be considered as preliminary (with some appropriate wording in the title), with some comment on this in the discussion section.

Yes, the manuscript required a cautionary note along the lines suggested by Dr. Della Mea. Please see the new paragraph included into the Conclusions section of the revised manuscript and quoted in the previous page.

**Minor Revisions**

*In this area, Klaus Kayser wrote some papers, which could be read by Authors to find some related work.*

Yes, Dr. Kayser’s valuable work such as the one on automated analysis of lung cancer histology images needed to be cited in the manuscript. This we have done in our Introduction section.

*Authors often use the expression "microarray" to refer to tissue microarrays. This is not correct, as microarray is more often the short name of DNA microarrays (gene chips). If Authors want to shorten the name, they can use the commonly adopted acronym TMA.*

Thank you. We corrected this throughout the manuscript.

*Chapter 2.3, 5th line: blocs should be blocks*

Yes, corrected.

**Reviewer: Donald O'Shea**

**Major Requested Revision:** None.

**Minor Considerations and Revisions:**

*The paper is seems like a nice approach to tackling the problem. The simplicity of using greyscale image segmentation with an estimation of a likelihood ratio test to determine if a segment of a whole section is cancerous seems straightforward...It would be good to allow the segmentation algorithm to change (and parameters such as B, P, H).*

Yes, we agree with the referee completely. We have added the following statement to the Results section of the revised manuscript.

“In the analysis of comprehensive image subsets involving different types of malignancy and/or tumors of different organs, the parameter set used in this article (B, P, H) can
readily be revised and enriched with additional texture parameters causing minimal change in the rest of the log-likelihood estimation algorithm."

*The classification should of course be able to accept new observed data sets.*

Yes, that is also correct. We have added the following statement into the Conclusion section.

“The automated texture image recognition algorithm developed for this article can readily be adapted to the recognition of additional histopathology textures. Incorporation of new data in the learning procedure is both possible and feasible: It only requires classifying them with respect to the reference sets used in randomized nearest neighbor classifications, and supplying the initial log-likelihood ratio estimates at the new data points to the support vector regression. As the amount of data incorporated into the system after the initial training grows large, a re-estimation of the log-likelihood ratios with new nearest neighbor reference sets may be performed to maintain maximum fidelity to all available data.”

*Machine learning usually has some form of feedback loop where the algorithm gets more 'intelligent' as it sees more data. This algorithm compares with a set of known results to make a decision.*

In the revised manuscript, we replaced the phrase machine learning in describing our automated analysis with statistical learning.

*I think a workflow diagram might help this paper.*

We agree. Our new Fig. 1 is the flow chart requested by Dr. O’Shea.

*I'm not sure why a Weibull distribution was used to model the luminance indices.*

In our image dataset, the luminance indices behaved according to an extreme value distribution of maxima; as they are limited from above by the luminance of the background, and exhibiting lower luminance indices depending on the amount of stained tissue each pixel carried as seen in the histogram in Figure 3. A two-parameter Weibull distribution is a special case of the generalized extreme value distribution which we have found to be well suited to capture such behavior.

*The determination of the upper bound for unstained regions could have large variation. It may be different for different types of tissue. However as unstained regions (no cancer) are not of interest they can be ignored (less processing).*

We agree. We chose the conservative route and adopt a high threshold for unstained tissue content in order to prevent exclusion of potentially important tissue sections.