Author’s response to reviews

Title: Identification of Disease-Causing Genes Using Microarray Data Mining and Gene Ontology

Authors:

Azadeh Mohammadi Miss (azade.mohammadi@gmail.com)
Mohammad Hossein Saraee Dr (saraee@cc.iut.ac.ir)
Mansoor Salehi Dr (m.salehi@med.mui.ac.ir)

Version: 2 Date: 31 October 2010

Author’s response to reviews: see over
Dear Editor:

Many thanks for considering our paper and constructive comments to improve the paper. We would like to inform you that we have addressed comments made by the reviewers and editorial board in the revised manuscript

Editorial Requests

Following changes are made based on editorial requests:

• Quality of written English:

We have revised the paper and have made changes/ improvement on structure and grammar. We have asked a native English speaking colleague (Dr C.H. Bryant) help us with this.

• Abstract Structure:

We have reorganized our abstract and structured it according to guidelines for Medical journals within the BMC series. Our abstract consists of four parts: Background, Methods, Results and Conclusions as the template
• Competing interests:
  we have added 'The authors declare that they have no competing interests' in
  Competing interests section.

• Acknowledgements:

We have acknowledged C.H. Bryant for proof reading of the paper. (we obtained
permission to acknowledge)

• Our revised manuscript conforms to the journal style and correctly formatted
  (based on http://www.biomedcentral.com/info/ifora/medicine_journals)

Reviewer1 comments

• 1) Quality of written English:

  • We have revised the paper and have made changes/ improvement on
    structure and grammar. We have asked a native English speaking
    colleague (Dr C.H. Bryant) help us with this.

  • "it don't" (page 2, line 12) has changed to “…method does not
    remove redundant genes”

  • "...genes are ignored also these methods..." is replaced by "...genes are
    ignored. Also these methods..." (page 9, last line)
• "...information is not completed for example..." is replaced by "... but it is incomplete, e.g., information for some genes is missing from GO ..." (first line page 11)

• "...genes as input, this list is obtained..." is replaced by “The greedy algorithm takes as input, from first stage, a list of 500 genes sorted in ascending order of discriminative power” (end of page 11)

• "fisher" is replaced by Fisher in all the paper.

• "TABLE 5 demonstrates" should be replaced by " Table 7 lists the genes predicted by our method to be colon cancer markers " (the number of tables are changed)(end of page 17)

• "SVM-RFE" should be replaced by "SVMRFE" in all the paper (except where it it the title of references)

• A through revision is made and several structural/grammatical changes are made.

2) Other errors:

• MRMR is a name of method which is used in reference[20] and in page 5 we have stated that is stand for Minimum Redundancy Maximum Relevance

• "accuracy", "sensitivity", "specificity" are defined in pages 14 and 15.

• We have added some tables to the paper and they are numbered from 1 to 8.
• “The label in second column of Tables 1 and 2 should be respectively "accuracy" and "sensitivity", not "specificity…” the tables are redesigned and now it is in a correct format.

• We have revised equations. The parameters of equation (1) are \( \mu_1(g) \) and \( \mu_2(g) \) and \( \sigma_1(g) \) and \( \sigma_2(g) \) (page 6)

• \( X_i \) and \( y_i \) are \( i^{th} \) sample and its label in the weight vector. We have defined them in page 7.

• 3)

• For the sake of clearness, in Table 7 we have highlighted with a * the genes which are reported in ref [43] as well as our method and genes reported in ref [11] by a +. We have done the same for table 8 (reported genes for DLBCL)

• 4)

• The accuracy, sensitivity and specificity results presented on Tables 1, 2 and 3 are the average value for repeated experiments (We have ignored standard deviation because it is not very considerable)

Reviewer2 comments

• 1) We have Revised the objective of the paper in the abstract and discussion part and we have stated the contribution of the paper to the state of the art. We have cleared that our method is a framework for combination of filter method and
embedded methods and an improvement to reduce redundancy using a greedy approach and incorporation Gene Ontology information “…In fact, we have combined the Fisher method and SVMRFE to utilize the advantages of a filtering method as well as an embedded method. Furthermore, we have added a redundancy reduction stage to address the weakness of the Fisher method and SVMRFE…” (page 2) or “In this paper, we propose a new framework for gene selection which combines the Fisher filter and the SVMRFE embedded method, with a greedy algorithm to remove the redundant genes.” (page 6)

• 2)
  
  • “Avoid using specialized phrases without giving a sense for their meaning. For example ‘Train SVM with linear kernel function…’ we have explained SVM with linear kernel function on page 7.
  
  • We have revised all equations and their associated explanation to avoid inconsistency.
  
  • We have tried to Discuss all relevant techniques with a unified vocabulary and notation in the context of microarray experiments and papers published in this context.

• 3,4)

  • We have described the datasets we have used in details on pages 13 and 14. Description of datasets including number of genes and samples, type of samples, type of microarray, … is provided.
• We have cited the original papers using each of these dataset for achieving more detailed information and also we have referenced the address for downloading datasets.

• Also the preprocessing steps for each dataset have been described.(for example normalization or missing value estimation …)

5,6)

• We have evaluated our methods on 3 different datasets and compared it with different methods based on accuracy, sensitivity and specification. The results are available at tables 1, 2 and 3 on pages 26 and 27.

• Also we have compared our results with results reported on other papers on the same dataset (tables on page 28)

• We have evaluated our method from biological view, one of the authors has PhD in Genetics and he is a member of Dept. of Genetics, Medical School. He has investigated the selected genes from our method. (we have done, and still are doing some experiments to check the validity of our results for a real situation. Our preliminary results looks very promising but we prefer to get more comprehensive results before we send the results of the experimental work for publication.)

• Also we have mentioned the limitations of the biological evidence in our paper for example: “The predictions made in this study can serve as a list of
candidates for subsequent wet-lab verification and might help in the search for a cure for colon, DLBCL or prostate cancer.” “It should be mentioned that the final approval of these genes as cancer factors, require more biological and laboratory investigations and selected genes, can only provide genetic researchers some clues for more research.” (page 20)

- 7)
  - We have updated our literature review with works published within the last three years. (different parts in pages 4, 5 and 6)

- 8)
  - We have cleared in abstract that our method is a framework for combination of filter method and embedded methods and an improvement to reduce redundancy using a greedy approach and incorporation Gene Ontology information “…In fact, we have combined the Fisher method and SVMRFE to utilize the advantages of a filtering method as well as an embedded method. Furthermore, we have added a redundancy reduction stage to address the weakness of the Fisher method and SVMRFE…” (page 2)

- 9) Quality of written English:
  - We have revised the paper and have made changes/ improvement on structure and grammar. We have asked a native English speaking colleague (Dr C.H. Bryant) help us with this.
• A thorough revision is made and several structural/grammatical changes are made.

Best Regards

Azadeh Mohammadi