Author's response to reviews

Title: Computer-assisted Lip Diagnosis on Traditional Chinese Medicine Using Multi-class Support Vector Machines

Authors:

FuFeng Li (li_fufeng@yahoo.com.cn)
Changbo Zhao (changbo0612@126.com)
Zheng Xia (zxia@tmhs.org)
Yiqin Wang (wangyiqin2380@sina.com)
Xiaobo Zhou (xzhou@tmhs.org)
Guo-Zheng Li (gzli@tongji.edu.cn)

Version: 3 Date: 26 May 2012

Author's response to reviews: see over
Response letter

Dear editor,

This is the upload again cover letter.

First of all, we would like to thank you and the reviewers for interesting, thorough, and helpful comments and suggestions. We have included most of them into the new version of our paper. For your convenience, we have highlighted the major changes with yellow in the new version. Besides, there are many minor changes throughout the text, which we do not point out them. Below, we first summarize the major changes, and then present the detailed responses to all of the reviewers.

Major Changes:

1. We have added many new paragraphs and tables about some algorithms to elaborate details of our complementary experiment according to referee 1 and 3. Meanwhile, we beautified the graphs and tables for readers’ convenience and comprehensibility. Besides, we removed the FLD algorithm and used other two algorithms of MAPLSC and naïve Bayes for instead, due that other two algorithms significantly perform better than FLD.

2. In order to obtain more stable results of experiment, we also checked and optimized our experiment code carefully. Therefore, based on that, we modified a little bit data results and drew
Response to the reviewers:

Reviewer #1

Comment 1:
The paper needs serious editing in terms of English and scientific writing, and the specialized words in the paper should be used more carefully and appropriately. For example, more widely used descriptions of the four diagnostic methods in traditional Chinese method should be “inspection, olfaction & auscultation, interrogation and palpation”. Avoid using “simply speaking” and writing in the first person. Any abbreviations introduced in the paper for the first time are needed to give their full version, for example, “CHD” in the 4th paragraph of the introduction section.

Our response:
Thanks to the reviewer. We had done our best to make our English language more understandable and accurate. Furthermore, we modified the descriptions of the four diagnostic methods and gave the full spelling of “CHD”. In addition, we checked all the paper to guarantee phrase accurately.

Comment 2:
There are many types of classification methods such as (A novel and
quick SVM-based multi-class classifier, Pattern Recognition, Vol. 39, No. 11, Nov.2006, 2258-2264; k-NS: a Classifier by the Distance to the Nearest Subspace, IEEE Transactions on Neural Networks, Vol.22, No.8, 2011, 1256-1268; Classification Using Distances from Samples to Linear Manifolds, Pattern Analysis and Applications ). Why the authors choose Support Vector Machine, and why not use other more common classifiers as the comparative methods, such as Naïve Bayesian, Bayesian networks, neural networks, etc. The authors should introduce the methods and compare them.

Our response:

Thanks very much for your kindly suggestion. We have read the papers carefully you suggested. One is a novel multi-class classifier based on SVM, and the other presents a novel and simple classifier called k-NS based on k-NN. However, our paper already contained SVM and k-NN, the two classifiers you suggested may be similar to our paper. In order to compare more different type classifiers, we used the classifiers of Naïve Bayes and MAPLSC (Pattern Recognition 2010) instead.

Comment 3:

There are lots of feature selection methods, why the authors use recursive feature elimination? 9 selected features used in the experiment are chosen based on SVM, according to the reference [16] in the paper,
the RFE depending on the classifier, therefore, I have a question that are the 9 selected features used by WSVM, kNN, and FLD same to the selected features of SVM?

Our response:
Thanks a lot for the reviewer’s reminder. We have added two filter type feature selection methods named mRMR (minimum Redundancy Maximum Relevance) and IG (information gain). More details have elaborated in our paper. For the second question, because the feature selection wrapper method is very sensitive to training data and classifier, the selected features are difficult to keep completely inconsistent on different classifiers. However, some partial important features can be selected on certain datasets.

Comment 4:
The experimental dataset may be too small. As an extension work could the author give an extensive experimental results based on a larger dataset in this paper? The accuracies for Pale lip prediction are low and the Pale lip only has 12 instances, do they have any relation?

Our response:
This question is really critical, while the clinical data collection is relatively difficult especially concerning image samples of patients. The effective sample 257 cases meet the experimental requirements in this
paper, in the future we will continue to collect data based on this work. On the second question, the pale lip sample is small compared with other classes, the imbalanced distribution results in the low accuracy of this class.

Comment 5:
The medical meanings of the features used should be introduced in the paper simply, and the affections of each type of features for lip image classification should be shown in the experiments. It is more helpful and valuable to the other researchers who have the related research interests.

Our response:
Thanks very much for the reviewer’s thorough comments. To be honest, these features have no direct pathological significance. The medical significance need to be combined with many medical biochemical indexes. We tested different features (which are color space features, Haralick co-occurrence features and Zernike moment features) for lip images classification, and found that the result in color space is slightly better than those of the Haralick co-occurrence features and Zernike moment features. However, the result turned out better when we combined all the three types of features. Due to the limitation of scope, we didn’t list all the results of each feature. Here in this paper gives the
classification results of the above three types of features.

Comment 6:

The importance of feature selection to classification is a common view. In the experiment, the authors should analyze how the selected features effect on the accuracy.

Our response:

Thanks a lot for the reviewer’s reminder. We have added more descriptions in our paper.

Comment 7:

Why the performance of WSVM is lower than SVM’s? According to the description in section 2.5 “the data is highly unbalanced and the sample size is rather small”, in the experiments, the performance of WSVM should be better than SVM’s. The authors should analyze the reasons.

Our response:

Thanks very much for the reviewer’s thorough comments. At first, we have same doubts about this question when we receive your review. So we checked our code carefully and diligently, and found that the data was fluctuated a little big. After we optimized the code, the result showed a reasonable accuracy in our paper. Sincerely thanks again for these thorough comments.
Comment 8:

The authors need conclude their work more carefully. What an explicit representation of the knowledge is? Does Lip diagnosis in TCM only indicate the four color lip classification? Could the experimental results sufficiently support the conclusion that the proposed approach is an effective technique for solving problems with multi-class in clinical practice of TCM?

Our response:

Thanks very much for the reviewer’s comments. The traditional diagnostic approach is mainly based on observation by doctor’s nude eyes, which is non-quantitative and subjective. The non-quantitative approach on the doctor’s experience affects accuracy of the diagnosis and treatment in TCM. This research aims to achieve the objective and quantitative information of TCM lip consultation, TCM clinical information collection and diagnosis. This work has clinical value considering compensation for the lack of the human eye.

Of course not, four-color lip is used commonly in clinical practice, the remaining occurs rarely. Most of classification tasks of TCM clinical data are multi-value, this work is a try. We need to point out that classification results need be in line with the interpretation of Chinese medicine clinician’s conclusion.
Reviewer #2

We are grateful for such a comprehensive evaluation by the reviewer. Your support is a great encouragement to our future work.

Reviewer #3

Comment 1:

Data set description.

The focus of the paper is the variables collected in 257 human subjects of whom no mention is made. Some simple descriptive statistics of this population according to age, gender, diagnosis, race, could help the reader in putting the results in the appropriate clinical context.

Our response:

Thanks very much for the reviewer’s thorough comments and suggestions. 257 cases refer to patients in different disease systems. This paper mainly focuses on classification of lip color of patients through feature extraction and classification, without considering the difference of lip features among diseases.

Comment 2:

Gold standard definition

Apparently the gold standard used to train the different classificatory is
the judgment of color made by two experts. If this is true this need to be clearly stated.

Our response:
Thanks very much for the reviewer’s suggestions. The physicians performed lip diagnosis using the clinical classification scale for facial diagnosis of TCM (see Appendix A).

Comment 3:
Statistical comparisons: in Table 1 and 2 P value of statistical comparisons among the classifiers must be added in order to ascertain the superiority of a classifiers vs another.

Our response:
Thanks very much for the reviewer’s thorough comments and suggestions. We have added a new paragraph and a new table to clarify this problem. Furthermore, we also obtained a total rank for the classifiers through the P value of statistical comparisons.

Comment 4:
Digital camera and LED light

In order to allow other researchers to reproduce the results, some technical specification of the apparatus employed is required

Our response:
We have added some technical specification of the apparatus. I.e, The compatible face image acquisition system includes LED (the light source, color temperature value about 5600K, Ra=90) and camera (the photographing medium). Its size is 36cm × 40cm × 28cm. The other photographing conditions include distance 33cm (between the camera and patients’ faces), Tv (1/15s), Av (5.6), ISO (80), white balance, custom mode and horizontal photography. The size of photographing windows is 220mm x 170mm.

Comment 5:

*Figure Legends: the legends should be included in the main text file immediately following the references, rather than being a part of the figure file.*

Our response:

Thanks very much for the reviewer’s suggestions. We have modified that.

Comment 6:

*Table 1 and 2: the author should specify if the figures after ± refer to Standard deviation or Standard error of mean.*

Our response:

It means that mean ± standard error of mean, but in this version of paper, we changed it to be mean ± variance, in order to meet readers’ intuitive
comprehension.

Best Regards

Sincerely,

Fufeng Li

Email: Li_fufeng@yahoo.com.cn