Author's response to reviews

Title: Characterization of Digital Medical Images Utilizing Support Vector Machines

Authors:

Dr Ilias Maglogiannis (imaglo@aegean.gr)
Elias P. Zafiropoulos (ezafir@power.ece.ntua.gr)

Version: 2 Date: 20 Oct 2003

PDF covering letter
Question
1. In a recent meta-analysis (Rosado et al., Archives of Dermatology, 2002) the following basic quality requirements for experimental studies on computer diagnosis of pigmented skin lesions were suggested
   a. Selection of lesions should be random or consecutive
   b. Inclusion and exclusion criteria should be clearly stated.
   c. All lesions clinically diagnosed as melanocytic lesions should be analyzed.
   d. The study setting should be clearly defined.
   e. To avoid verification bias, clearly benign lesions which were not excised should be included. For these lesions the diagnostic gold standard could be short term follow-up with digital dermoscopy.
   f. Instrument calibration should be reported.
   g. Repeatability analysis should be carried out (inter- and intra-instrument)
   h. Classification should be carried out on an independent test set
   i. Computer diagnosis should be compared with human diagnosis.

The authors do not provide sufficient data to answer the question whether these quality requirements have been fullfilled or not. For example, the way the lesions were selected is not reported. Selection bias may explain the extremely good results of this study.

Answer
a. Our sample was consecutive based on the patients that arrived at the Dept of Plastic Surgery in Athens General Hospital. We have included all the melanomas and all the suspicious dysplastic naevi. The collection of samples lasted 6 months. Total number of lesions: 17, Number of melanomas: 7, Number of dysplastic naevi: 10. For a justification of the sample size see next question.

b. For the training set we have used the clinical diagnosis made by the physicians either via visual or dermoscopical inspection, either via histological examination.

c. For the image acquisition we have developed a system that performed real time calibration described in (I. Maglogiannis, D. Kosmopoulo: “A Digital Image Acquisition System for Skin Lesions” In SPIE Proc. of Medical Imaging 2003, Image Perception, Observer Performance and Technology Assessment pp. 337-346 2003) and in (I. Maglogiannis, D. Kosmopoulo: “A System for the Acquisition of Reproducible Digital Skin Lesion Images” Technology and Healthcare, IOS Press to appear in 2003). The manuscript was revised to include this information (see section Methods pages 3-5).

d. The repeatability of results was ensured via the utilization of on-line calibration (see f).

h. We have used the same training set and the test set due to the small size of the available sample. This is applied in literature [Rocco C. M., Moreno J. Ali, “Fast Monte Carlo reliability evaluation using support vector machine”, Reliability Engineering and System Safety, 76 (2002) pp. 237-243]. The support vector machines calculated performed better than the rest of the methods presented in the paper while it used the same cases of melanoma.
and dysplastic nevus. However, this study is preliminary and it can be further improved including more medical cases so as to employ different test sets for a possible support vector improvement.

i. Computer diagnosis was actually compared with human diagnosis since the results of computer program classification were cross validated with the human opinion on the lesions.

Question
2. The authors stated that they used 17 lesions in the training set, which is a very low number. Why only 17 lesions? Did the investigators perform other experiments including more lesions? If yes, what were the results of these experiments. Did the authors only provide data of the experiment with the best results?

Answer
The number of melanomas used in our study is relatively small. This is due to the fact that malignant melanoma cases in primordial state are too rare. Although Greece is a country with an increasing number of melanomas, it is very common that many patients arrive at specialized hospitals (e.g. Athens General Hospital with partially removed lesions; lesion removal is a simple operation that may be performed in small health centers. The above sample was collected within a period of 6 months. The number of dysplastic naevi was intentionally kept low to avoid bias of the classification. The results are probably very good due to the low number of cases. We agree though that it is a preliminary study and it is now necessary to examine more patients in order to increase the number of cases, particularly during the classification phase. This will clarify the issue of selecting the most powerful variables for classification. The manuscript has been revised accordingly to acknowledge this (see Abstract page 1 and Conclusions page 15)

Question
3. The authors never tell us how many lesions were included in the test set.

Answer
The training set and the test set were the same due to the small size of the available sample. For this reason all the results were cross validated. The cross validation procedure was basically; remove a sample, perform classification without this sample and then check if this sample is classified correctly. Then the sample returns to the training set, and a new sample is removed. The process is continued until every sample has been rotated out once. The manuscript has been revised to include this clarification (see Methods page 10).

Question
4. It seems that the authors used the test set for optimizing the parameter sigma (see section VGP-DSP comparison). In this case the test set is not independent.

The training set and the test set were the same due to the small size of the available sample. However this case has appeared in literature [Rocco C. M., Moreno J. Ali, “Fast Monte Carlo reliability evaluation using support vector machine”, Reliability Engineering
and System Safety, 76 (2002) pp. 237-243] and the authors tried subsets of the train set, just to detect which kernel function is less complex and results in a low number of support vectors. The Gaussian radial base kernel function with the specified values for the parameters was detected in this way and the support vectors were calculated using the full train set of the 17 cases and this kernel function.

Question
5. Regarding the references, the authors overlooked a large part of the literature which appeared in medical journals.

Answer
After a more recent literature review the following references have been added.


Advice on publication: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
Level of interest: A paper whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Declaration of competing interests: none
Reviewer's report
Title: Characterization of Digital Medical Images Utilizing Support Vector Machines
Version: 1 Date: 15 September 2003
Reviewer: Stefania Seidenari
Reviewer's report:
General
Although the design of the study appears interesting as a comparison of different classification methods, this study was improperly conducted. Aiming at a 100% sensitivity for the identification of all malignant lesions, the specificity is always lower, especially when comparing difficult to diagnose lesions such as melanomas and dysplastic nevi. In real conditions, a 100% sensitivity and specificity is not even reached by histology, the usual golden standard in this kind of studies. Certainly, these high sensitivity and specificity values are due to selection bias: evident melanoma versus evident nevus images are compared for the diagnosis. Even the identification of "easy to diagnose" lesions can be of value, but, if this was the case in this study, it has to be specified in the method section. Moreover, the study population and the basic statistical analysis have to be reported and described in details.

Discretionary Revisions (which the author can choose to ignore)
Comment
1 In Abstract: "Results" sub-section: abbreviations were not explained
Answer
Revised Accordingly

Minor Compulsory Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1 Some sentences referring to the methods are reported in Results
The sentence “A kernel function has a good performance if the support vectors that are calculated by using the corresponding transformation are few and the classification of the test data is successful” has been moved to the methods section (see page .8.).

2 The study population is not described: how many lesions? How many melanomas and what about their thickness? "On what basis a nevus is defined as "dysplastic" (histology or clinical atypia?)

Our sample was the patients that arrived at the Dept of Plastic Surgery in Athens General Hospital. We have included all the melanomas and all the suspicious dysplastic naevi. Total number of lesions: 17, Number of melanomas: 7, Number of dysplastic naevi: 10. A nevus was defined as "dysplastic" basically from the clinical evaluation. The mean thickness of melanoma lesions was measured during biopsy at approximately 1.5 mm penetration through the skin.
We agree though that it is a preliminary study and it is now necessary to examine more patients in order to increase the number of cases, particularly during the classification phase. This will clarify the issue of selecting the most powerful variables for classification. The manuscript has been revised accordingly to acknowledge this (see Abstract and Conclusions)

3 In the abstract it is reported that images were "acquired under reproducible conditions". It is not explained in the Materials and methods: what does "reproducible conditions" mean?
What kind of instruments or digital cameras were employed? Were images acquired with a fixed magnification and distance? Owing to an indefinite number of variables (light sources, angle of incidence of the light, skin scaling and skin hydration, etc ... ), I think that it is impossible to have "reproducible" clinical images, or it must be proven recording the same lesions several days after and comparing the obtained parameter with the appropriate statistical tests.

Skin images were capture with the help of a prototype image acquisition system that includes a standardized illumination and capturing geometry with polarizing filters and a series of software corrections: Calibration to Black, White and Color for color constancy, Internal camera Parameters adjustment and Pose extraction for stereo vision, Shading correction and Noise Filtering for color quality. The validity of the calibration procedure and the images’ reproducibility were tested by capturing sample images in three different lighting conditions: dark, medium and intense lighting. The validity of the calibration procedure and the images’ reproducibility were tested by capturing sample images in three different lighting conditions: dark, medium and intense lighting. For each case the average values of the three color planes RGB and their standard deviations were calculated; the measured error differences ranged between 0.7 and 12.9 (in the 0-255 scale). Preliminary experiments for stereo measurements provided repeatability of about 0.3mm.

The specific system is described in described in (I. Maglogiannis, D. Kosmopoulos : "A Digital Image Acquisition System for Skin Lesions" In SPIE Proc. of Medical Imaging 2003, Image Perception, Observer Performance and Technology Assessment pp. 337-346 2003) and in (I. Maglogiannis, D. Kosmopoulos : "A System for the Acquisition of Reproducible Digital Skin Lesion Images” Technology and Healthcare, IOS Press to appear in 2003). Manuscript has been revised to include this information (see section methods pages 3-4)

4 The definitions "Melanoma's Vertical Growth Phase" and "Melanoma's Radial Growth Phase" referred to the whole lesion and to dark areas inside the melanoma, respectively, are incorrect and misleading terms, referring to histological behaviour of melanomas without a necessary correlation with clinical aspects. These definitions have to be changed.

These terms were introduced by the medical personnel participated in our research. We have used this discrimination on the basis that RGP phase was indicated separately on the digital images from the expert dermatologists and had different chromatic properties from the rest of melanoma.

5 The "Melanoma's Radial Growth Phase" (RGP) is a manually selected portion of a melanoma images, corresponding to a dark area inside the lesion. It is not specified in how many melanomas it was present. Moreover, the comparison between a selected dark area inside a lesion presumptive known to be a melanoma with a whole lesion is methodologically incorrect. It is obvious that a "dark area" differs for color and shape from a whole pigmented skin lesion, independently of its nature. The selected dark areas in melanomas can be at least compared to similarly selected dark areas in dysplastic nevi, in order to identify differences in color components.

RGP phase was present on all melanomas. For the dysplastic naevi the whole lesion was taken into account. For this reason we selected to compare DSP (dysplastic naevi with both RGP (dark area inside the melanoma) and VGP (the whole lesion) classes.
6 Parameters employed for discrimination between RGP and dysplastic nevi by means of discriminant analysis and neural network approaches have not been reported.

Two more sections regarding discriminant analysis and neural network approaches have been introduced.

7 The "greatest diameter", one of the two parameter on which the discrimination between melanomas and dysplastic nevi is predominantly based, represents the dimension of the lesion. For discriminating algorithms based on image analysis, all geometric parameters depending on the lesion dimension should be avoided, since melanomas are obviously larger than common nevi and the selection of the study population depends on the human choice (selection bias). Moreover, in order to obtain additional information from a computer in respect to the clinical evaluation, color and texture characteristics should be investigated rather than obvious criteria depending on the lesion selection, such as the dimension.

The greatest diameter feature along with thinness ratio was indicated by a stepwise discriminant analysis (carried out using the SPSS package) using forward selection of significant predictors (with the criterion of Wilks' lambda statistic, indicating the separation between the groups in the multivariate space) as the most significant features for discrimination. Perhaps in another paper we may be constrained in color based features.

8 Finally, basic statistics (mean, standard deviation and statistical tests for comparison between the different groups) for the image analysis parameters were not included.

We have incorporated this information in the following Table.

<table>
<thead>
<tr>
<th>Features</th>
<th>DSP</th>
<th>RGP</th>
<th>VGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irregularity A</td>
<td>0.058 (0.028)</td>
<td>0.030 (0.016)</td>
<td>0.041 (0.016)</td>
</tr>
<tr>
<td>Irregularity B</td>
<td>3.38 (0.20)</td>
<td>3.06 (0.56)</td>
<td>4.05 (0.47)</td>
</tr>
<tr>
<td>Thinnness Ratio</td>
<td>0.66 (0.04)</td>
<td>0.98 (0.35)</td>
<td>0.48 (0.10)</td>
</tr>
<tr>
<td>Red (Average)</td>
<td>69.5 (10.6)</td>
<td>36.3 (19.6)</td>
<td>104.5 (48.8)</td>
</tr>
<tr>
<td>Green (Average)</td>
<td>66.1 (19.4)</td>
<td>36.5 (8.4)</td>
<td>78.5 (31.3)</td>
</tr>
<tr>
<td>Blue (Average)</td>
<td>49.5 (18.3)</td>
<td>28.1 (9.7)</td>
<td>67.2 (33.0)</td>
</tr>
<tr>
<td>Red (St.Dev.)</td>
<td>22.1 (10.8)</td>
<td>18.3 (7.9)</td>
<td>37.4 (14.0)</td>
</tr>
<tr>
<td>Green (St.Dev.)</td>
<td>23.3 (8.9)</td>
<td>16.5 (6.0)</td>
<td>30.2 (12.7)</td>
</tr>
<tr>
<td>Blue (St.Dev.)</td>
<td>21.3 (8.1)</td>
<td>16.2 (5.5)</td>
<td>29.3 (12.3)</td>
</tr>
<tr>
<td>I1 (R+G+B / 3)</td>
<td>62.8 (9.9)</td>
<td>33.6 (15.6)</td>
<td>92.0 (43.0)</td>
</tr>
<tr>
<td>I2 (R-B)</td>
<td>20.0 (19.9)</td>
<td>8.3 (14.0)</td>
<td>37.3 (21.4)</td>
</tr>
<tr>
<td>I3 (2G-R-B /2)</td>
<td>6.61 (11.34)</td>
<td>4.25 (6.85)</td>
<td>-7.36 (11.22)</td>
</tr>
<tr>
<td>Feature</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Average Intensity</td>
<td>62.8 (13.4)</td>
<td>33.6 (11.8)</td>
<td>83.4 (37.1)</td>
</tr>
<tr>
<td>Average Hue</td>
<td>1.23 (0.83)</td>
<td>1.66 (0.84)</td>
<td>1.08 (0.75)</td>
</tr>
<tr>
<td>Average Saturation</td>
<td>0.27 (0.13)</td>
<td>0.30 (0.11)</td>
<td>0.24 (0.13)</td>
</tr>
<tr>
<td>Average L</td>
<td>109.8 (21.7)</td>
<td>60.0 (20.9)</td>
<td>148.3 (65.6)</td>
</tr>
<tr>
<td>Average Angle A</td>
<td>1.13 (0.11)</td>
<td>1.09 (0.10)</td>
<td>1.12 (0.09)</td>
</tr>
<tr>
<td>Average Angle B</td>
<td>0.74 (0.17)</td>
<td>0.87 (0.22)</td>
<td>0.67 (0.08)</td>
</tr>
<tr>
<td>Greatest Diameter</td>
<td>84.1 (28.9)</td>
<td>156.3 (50.0)</td>
<td>279.8 (90.4)</td>
</tr>
<tr>
<td>Asymmetry</td>
<td>13.5 (11.1)</td>
<td>10.9 (4.1)</td>
<td>26.2 (10.3)</td>
</tr>
</tbody>
</table>

Table 1. Mean values (standard deviations in parentheses) of features, by group.

9 The highest number of scientific articles on image analysis on pigmented skin lesions are reported in dermatological journal: the most relevant ones should be included in references.

After a more recent literature review the following references have been added.

Advice on publication: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
Level of interest: A paper whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Declaration of competing interests: none