BCCT.core Model Enhancement With Interpretability and Lateral Information

Hélder P. Oliveira and Jaime S. Cardoso
Universidade do Porto, Faculdade de Engenharia, INESC Porto
helder.oliveira@fe.up.pt, jaime.cardoso@inescporto.pt

Abstract

Breast Cancer Conservative Treatment (BCCT) is considered the gold standard of breast cancer treatment. The heterogeneity of the aesthetic result and the limited reproducibility of the subjective evaluation motivated the research towards objective methods, such as, the recent computer system named BCCT.core, based on machine learning techniques, namely support vector machines (SVMs).

In the current work we investigate the accuracy of different interpretable methods against the model currently deployed in the BCCT.core software and the improvement of the model by introducing lateral information extracted from patients images. Experimental results show only a marginal improvement in the performance of the new models, suggesting that is essential to use more robust models, such as 3D approaches.

1. Introduction

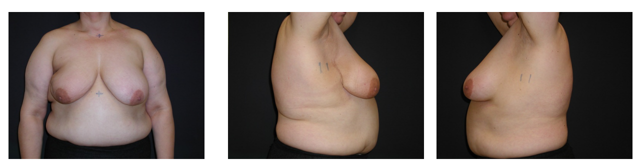
Breast cancer is the most common cancer to affect women in Europe and many women are expected to live a longer time with the aesthetic results of their local breast cancer treatment, increasing the importance of a good aesthetic outcome. It is known that auditing this problem and developing techniques for improving aesthetic results is a difficult task due to the absence of a measuring standard method. Formerly breast cancer surgical treatment consisted essentially of only one of two procedures: mastectomy or BCCT (excision of only the tumour) of the affected breast. Although considerable research has been put on breast-conservation techniques, different forms of performing them, as well as incorrect working practices, contribute to different aesthetic results. Therefore, the aesthetic result evaluation in any institution performing breast cancer treatment is fundamental to improve current strategies [3].

In order to overcome the poor objectivity and reproducibility of the subjective assessment, objective methods were introduced, consisting on measurements between identifiable points on patient photographs [3]. These kinds of methods are essentially based on asymmetries between treated and non-treated breasts, such as the BCCT.core [1] software.

The objective of the current work is to investigate improvements on BCCT.core, by comparing the performance of interpretable methods against the currently deployed in BCCT.core and by introducing lateral information as a complement of the frontal information currently used.

2. BCCT.core design

The development of BCCT.core [1] involved the automatic extraction of several features from the frontal patient's photographs (Figure 1(a)), directly related with the cosmetic result, such as: breast asymmetry, skin colour changes due to the radiotherapy treatment and surgical scar appearance. After that, a SVM classifier was trained in order to predict the overall cosmetic result from the recorded features.



(a) (b) (c)

Figure 1. Typical photographs.

2.1. BCCT.core enhancement

BCCT.core is currently being used by many international groups in prospective studies; however, improvements on interpretability of the model were often suggested. SVMs have proven to be very useful in machine learning; however, there is a significant drop of understandability of the learned hypothesis, especially when using nonlinear kernels, as is the case of the radial basis function (RBF) kernel.

In this work we researched the accuracy of interpretable methods (LINEAR SVM and classification trees) against the model currently deployed (RBF SVM). We used a set of 143 photographs recorded at different breast centers and a ground truth from the patients evaluation made by an international panel of experts in order to train the classifier. The distribution of the patients over the four different classes was: Excellent - 20; Good - 74; Fair - 34; Poor - 15. (The data set was posteriorly divided in a training set and testing set, with 113 and 30 respectively). A simplified feature selection was also conducted on this study and three different sets of features were selected [4]. The addressed problem involves classifying examples into classes which have a natural ordering, consequently, was adopted specific classification methods for this kind of data [2]. A "grid-search" on the parameters of the models was performed using cross-validation in the training set. From experimental tests, the best achieved result for the RBF SVM, was 0.30 for the misclassification error (MER) and 0.52 for the weighted MER while for LINEAR SVM was 0.37 and 0.52 respectively. Weighted error represents the fact of caregivers consider an error in a true excellent or true poor patient more penalizing than an error in the middle classes (fair or good). Moreover, failure to a contiguous class is not as serious as failure to a non contiguous class. From these considerations, and in collaboration with clinical experts, we empirically defined a cost matrix reflecting the penalty of classifying samples from one class as another:

$$C = \begin{bmatrix} 0 & 2 & 4 & 6 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 6 & 4 & 2 & 0 \end{bmatrix}$$

where $C(i, j)$ is the cost of classifying a point into class j if its true class is i. This cost matrix was used during training and testing phase.

Analyzing the obtained results we can say that the LINEAR SVM follows closely the accuracy of the nonlinear model, in particular if we focus on the weighted error.

Almost all methodologies reported until now, namely the used on BCCT.core, only use patient frontal images, that could discard some important information related with volume perception. Lateral information extracted from patients side-views (Figure 1(b) and 1(c)) can provide to model more information about breast shape, and was investigated in this work as a complement to the currently used frontal information.

On this part of the study a portion of the initial dataset was used, namely 63 photographs, due to the following reasons: definition of the normalization factor between images; fiducial points annotation; poor quality of some of the lateral photographs. The new distribution of the patients over the four different classes was: Excellent - 10; Good - 35; Fair - 11; Poor - 7. Differently from frontal analysis, there are not many studies reporting the definition of features in lateral images. Features were defined from the comparison between the treated with the non-treated breast. This represents an additional problem because we have different images for each breast, and a normalization factor must be found to have all measurements in the same scale. Features were defined from measurements obtained using fiducial points manually marked on images [5].

From experimental tests, the best achieved result using both frontal and lateral information was 0.32 for the MER and 0.53 for the weighted MER while for the model only with frontal information was 0.37 and 0.60 respectively. With the integration of these new lateral features the classification error decreases; however, it did not represent a significant improvement.

3. Conclusions

We have developed different accurate and interpretable models for the assessment of aesthetic result of BCCT. We have shown that the linear model achieves a performance very similar to the RBF SVM, with the obvious advantages of simplicity and interpretability, without sacrificing the performance. Additionally, we investigated the introduction of lateral features on the developed model that showed a better performance, however it did not represent a significant improvement.

4. Acknowledgments

This work is financed by the ERDF European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT Fundao para a Ciłncia e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-007235.

References

[1] J. S. Cardoso and M. J. Cardoso. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. *Artificial Intelligence in Medicine*, 40:115–126, 2007.

[2] J. S. Cardoso and J. F. P. da Costa. Learning to classify ordinal data: the data replication method. *Journal of Machine Learning Research*, 8:1393–1429, 2007.

[3] D. R. H. Christie, M.-Y. O'Brien, J. A. Christie, T. Kron, S. A. Ferguson, C. S. Hamilton, and J. W. Denham. A comparison of methods of cosmetic assessment in breast conservation treatment. *Breast*, 5:358–367, 1996.

[4] H. P. Oliveira, A. Magalhaes, M. J. Cardoso, and J. S. Cardoso. An accurate and interpretable model for bcct.core. In *Proceedings of the 32^{nd} Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6158 – 6161, 2010.

[5] H. P. Oliveira, A. Magalhaes, M. J. Cardoso, and J. S. Cardoso. Improving the bcct.core model with lateral information. In *Proceedings of the 10^{th} IEEE International Conference on Information Technology and Applications in Biomedicine*, 2010. (ACCEPTED).