

Classification of mammographic masses using support vector machines and Bayesian networks

Maurice Samulski^a, Nico Karssemeijer^a, Peter Lucas^b, and Perry Groot^b

^aDepartment of Radiology, Radboud University Medical Centre, Geert Grooteplein Zuid 18, 6525 GA Nijmegen, The Netherlands;

^bRadboud University, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

ABSTRACT

In this paper, we compare two state-of-the-art classification techniques characterizing masses as either benign or malign, using a dataset consisting of 271 cases (131 benign and 140 malign), containing both a MLO and CC view. For a number of suspected regions in a digitized mammogram, 12 out of 81 calculated image features have been selected for investigating the classification accuracy of support vector machines (SVMs) and Bayesian networks (BNs). Additional techniques for improving their performance were included in their comparison: the Manly transformation for achieving a normal distribution of image features and principal component analysis (PCA) for reducing our high-dimensional data. The performance of the classifiers were evaluated with Receiver Operating Characteristics (ROC) analysis. The classifiers were trained and tested using a k-fold cross-validation test method (k=10). It was found that the area under the ROC curve (A_z) of the BN increased significantly (p=0.0002) using the Manly transformation, from $A_z = 0.767$ to $A_z = 0.795$. The Manly transformation did not result in a significant change for SVMs. Also the difference between SVMs and BNs using the transformed dataset was not statistically significant (p=0.78). Applying PCA resulted in an improvement in classification accuracy of both BNs and SVMs, from $A_z = 0.767$ to $A_z = 0.786$, and $A_z = 0.793$ to $A_z = 0.799$, respectively. The difference in classification performance between BNs and SVMs after applying PCA was small and not statistically significant (p=0.11).

Keywords: Methods: classification and classifier design, pre-processing, Modalities: mammography, Diagnostic task: diagnosis

1. INTRODUCTION

Machine learning techniques to diagnose breast cancer is a very active research area. Several computer-aided diagnosis (CAD) systems have been developed to aid radiologists in mammographic interpretation. These CAD systems analyze the mammographic abnormalities and classify lesions as either benign or malignant in order to assist the radiologist in the diagnostic decision making. There are already commercial systems available on the market. Some of them are based on Bayesian networks learned on mammographic descriptions provided by radiologists¹ or on features extracted by image processing.² Another classification technique that is widely used for the diagnosis of breast tumors are support vector machines which are known to be especially suited for dealing with classification data with a non-linear decision boundary.³⁻⁵ One of the shortcomings is the black-box nature of the model, whereas in Bayesian networks statistical dependences and independences between features are represented. In this study we compare both classification methods and use two techniques, namely dimension reduction by principal component analysis (PCA) and normality transformation, to further improve the accuracy rate of the classifiers. Recently, the combination of PCA and SVMs has been used in medical imaging, where principal component analysis is applied to extracted image features and the results are used to train a SVM classifier, but not specifically for mammograms.⁶

2. MATERIALS AND METHODS

The digitized mammograms that were used in this study have been obtained from the Dutch Breast Cancer Screening Program. In this program two mammographic views of each breast were obtained in the initial screening: the medio-lateral oblique (MLO) view and a cranio caudal (CC) view. In this study, 271 cases with both a MLO and CC view available were used, 131 benign and 140 malign. In the conducted experiments we used a subset of 12 features out of 81 features that were computed per region selected in previous research using feature selection.³ Combining the 12 features of the MLO with the 12 features of the corresponding CC gives

Further author information: (Send correspondence to Maurice R.M. Samulski)
M.R.M. Samulski, e-mail: m.samulski@rad.umcn.nl, telephone: +31 24 3619811, fax: +31 24 3540866

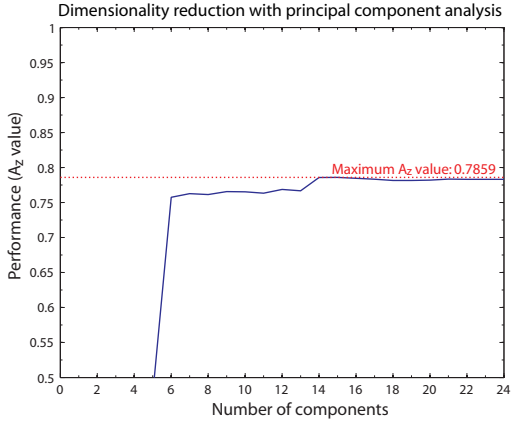


Figure 1: Case based performance naive Bayes classifier after dimensionality reduction with PCA, averaged over 5 runs.

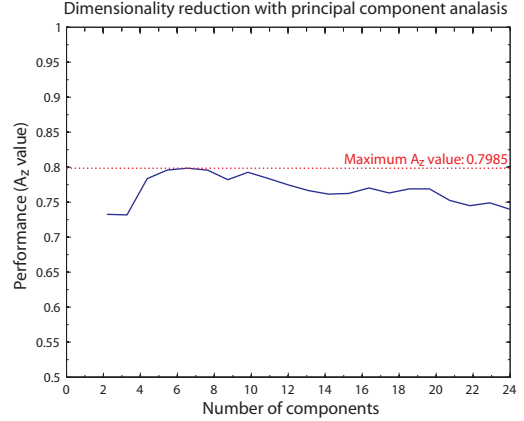


Figure 2: Case based performance SVM classifier with radial kernel function after dimensionality reduction with PCA, averaged over 5 runs.

a total of 24 features per case. The continuous output of the classifier is analyzed using ROC methodology, using the LABROC program⁷ of Metz et al. The statistical significance of the difference between ROC curves was tested using the CLABROC program⁸ of Metz et al. The classifiers were trained and tested using a k-fold cross-validation test method (k=10), in which each of 10 different combinations of training and test data sets included 244 and 27 cases, respectively. For each test partition, the classification accuracy was evaluated as the area A_z under the ROC curve. Many Bayesian learning algorithms that deal with continuous nodes are based on the assumption that the features are normally distributed. Unfortunately, most of the image features we use do not follow a normal distribution. We used Manly's power transformation to make the non-normal data resemble normal data by reducing skewness. One might think that the use of more features will automatically improve the classification power of the classifier. However the number of samples needed per feature increases exponentially with the number of features to maintain a certain level of accuracy. Dimensionality reduction is therefore a good choice to overcome this problem. One of the most well-known dimension reduction techniques is principal component analysis, that has been used in our study as a preprocessing step to reduce the dimensionality of our dataset.

3. RESULTS

We evaluated the classification performance of the naive Bayes classifier after applying the Manly transformation on the dataset to remove skewness. The calculated area under the ROC curve (A_z value) of the Bayesian classifier without transforming the dataset was 0.767. After applying the Manly transformation it increased to 0.795, which is a significant statistical improvement ($p=0.0002$). The performance of BNs with SVMs using the transformed dataset showed that the difference was not statistically significant ($p=0.78$).

Additionally, we evaluated the classification performance of the naive Bayesian and SVM classifier after applying dimensionality reduction on our dataset. Figure 1 shows the classification performance of the naive Bayesian classifier where horizontally the number of principal component vectors is plotted and vertically the area under the ROC curve. The principal component vectors are calculated using the training set only. These principal component vectors are then used to transform both the training and test set. The best result was obtained with 14 principal components and remained almost constant when adding more dimensions. With SVMs the best result was obtained with only 6 principal components and decreased gradually if more components were added which is shown in Figure 2. The difference in classification performance between BNs and SVMs was statistically insignificant ($p=0.11$) when we used the optimal number of principal components for the classifier. In an additional experiment we trained a SVM on all the available features (81 per view) that led to the classification results shown in Figure 3. The maximum performance was reached in 10 components ($A_z = 0.799$) and was slightly higher than the experiment with the subset of the 12 most important features ($A_z = 0.793$).

4. CONCLUSION

We performed a study to compare two state-of-the-art classification techniques characterizing masses as either benign or malign. We evaluated the effectiveness of dimension reduction and normality transformation in improving the classification accuracy. The Manly transformation method significantly improved classification accuracy of

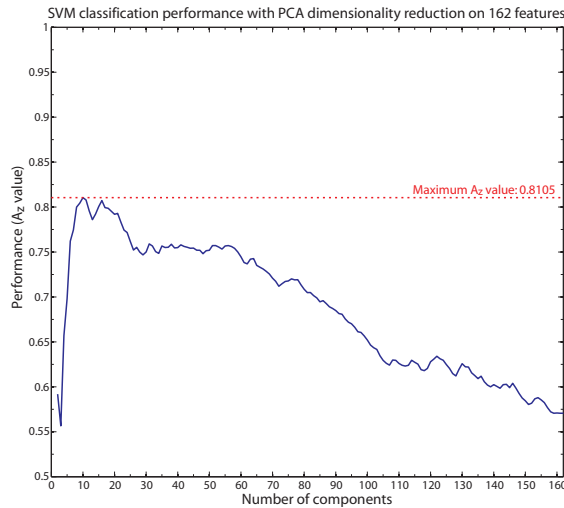


Figure 3: Case based performance SVM classifier with radial kernel function after dimensionality reduction of all features (81 per view) with PCA, averaged over 5 runs.

the Bayesian networks by transforming the distribution of the non-normal data closer to the normal distribution. We also found that this transformation does not work for all data, i.e., transforming features that were already approximately normal and also discrete variables had a negative impact on the classification accuracy. After transformation, the difference in performance of SVM and BN was not statistically significant. In contrast to SVMs, Bayesian networks allow incorporating background knowledge, which may be exploited to improve their performance. Despite the major drawback of principal component analysis, i.e., it can eliminate a dimension that is good for discriminating positive cases from negative cases, this unsupervised dimension reduction algorithm improved the classification accuracy of both classifiers. The performance of the two classifiers after applying PCA was very similar, with no statistical differences in the area under the ROC curve. Future research could concentrate on building multiple classifier systems combining models of different types to produce more reliable results.

REFERENCES

1. E. Burnside, D. Rubin, and R. Schachter, "A Bayesian network for mammography," in *Proceedings of AMIA Annual Symposium*, pp. 106–110, 2000.
2. X. Wang, B. Zheng, W. Good, J. King, and Y. Chang, "Computer assisted diagnosis of breast cancer using a data-driven Bayesian belief network," *International Journal of Medical Informatics* **54**, pp. 115–126, May 1999.
3. S. Timp, *Analysis of Temporal Mammogram Pairs to Detect and Characterise Mass Lesions*. PhD in medical sciences, Radboud University Nijmegen, 2006. ISBN 9090205500.
4. T. Nattkemper, B. Arnrich, O. Lichte, W. Timm, A. Degenhard, L. Pointon, C. Hayes, and M. Leach, "Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods," *Artificial Intelligence Medical* **34**, pp. 129–139, June 2004.
5. M. Mavroforakis, H. Georgiou, N. Dimitropoulos, D. Cavouras, and S. Theodoridis, "Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines," *European Journal of Radiology* **54**, pp. 80–89, April 2005.
6. S. Li, T. Fevens, and A. Krzyzak, "Automatic clinical image segmentation using pathological modeling, PCA and SVM," *Engineering Applications of Artificial Intelligence* **19**, pp. 403–410, June 2006.
7. C. Metz, B. Herman, and J. Shen, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Statistics in Medicine* **17**, pp. 1033–1053, 1998.
8. C. Metz, P. Wang, and H. Kronman, "A new approach for testing the significance for differences between ROC curves measured from correlated data," in *Information Processing in Medical Imaging*, F. Deconinck, ed., pp. 432–445, The Hague: Nijhoff, 1984.