Pre-CADs in Breast Cancer

Joana Lopes da Fonseca  
INESC TEC and  
Faculdade de Engenharia  
Universidade do Porto

Jaime S. Cardoso  
INESC TEC and  
Faculdade de Engenharia  
Universidade do Porto

Inês Domingues  
INESC TEC  
Faculdade de Engenharia  
Universidade do Porto

Abstract—In this study we present a pre-CAD system that aims to help the radiologists in the analysis of the high number of mammograms that they have to evaluate each day, helping prevent the increased number of misclassification that could happen, due to the repetitive task to which they are submitted. The method consists in extracting features from mammograms, previously classified by experts according to the breast density and then classify them into normal or abnormal mammograms for each of the tissue density types.

I. INTRODUCTION

Breast cancer (BC) is the leading cause of death, by cancer, among Portuguese women. According to Instituto Nacional de Estatística the number of deaths by BC is assuming an increasing perspective. From 2006 to 2010, the standardized mortality rate for BC has increased from 3.7 to 30.3 deaths per 100000 women [1]. Each year, 4500 new cases of BC are diagnosed and 1500 women will die of BC. If early detected and then submitted to a correct treatment, 90% of all BC diagnosed are curable and the mortality rate could decrease up to 30%.

A pre-CAD system, functioning as a “first look” in the analysis of the mammography could help reduce the work of the radiologists, providing with the opportunity to enhance their performance among the most difficult cases. Unlike to a pre-CAD system, a CAD system may be used as a second opinion for reviewing a mammogram after the radiologist has already made an initial interpretation.

II. STATE-OF-THE-ART

Wulf was the first author mentioning the idea of breast density classification in 1976. The author used solely visual classification of mammograms to classify them according to the density into four categories. He concluded that a relation between the four categories used to classify the breast density and the risk of developing BC exists. As the density in the breast increases, also the risk of developing BC increases [2][3].

The idea that there is a relation between the breast density and the risk of developing BC was followed in 1995 by Boyd et al., where the authors used a quantitative computerized thresholding method in order to measure the density of the breast [4].

In 2002 Bovis and Singh investigated a new approach in the classification of mammograms according to the tissue type using a combined classifier paradigm. They have studied this task by following two different paths: the four-class classification problem and the two-class problem. The results obtained by Bovis and Singh indicated that using the two-class classification method improves the performance and robustness of the classification, as expected.

In 2010 Elshinawy et al proposed an algorithm to separate the mammograms into dense and fatty categories and to extract features from each category individually. A better accuracy is obtained when using majority voting with rotation invariant Local Binary Pattern(LBP) features [5]. In the following work of the authors, rather than using only LBP features, they used also Grey Level Co-occurrence matrix(GLCM) features with only first-order statistics and GLCM features with both first-order and second-order statistics. The authors concluded that GLCM features were more accurate than simple LBP and rotation invariant LBP features. Using GLCM features with both first and second-order statistics improves the system so much, that they managed to reach a rate of False Negative (FN) of 0% and a rate of True Positive (TP) of 100% for both dense and fatty mammograms. For dense mammograms they reach a rate of True Negative (TN) of 96.31% and a rate of False Positive (FP) of 3.69%. For fatty mammograms the rates decrease a little in terms of TN to 95.85% and increases in terms of FP to 4.14% [6]. The author was able to conclude that separating the dense and fatty mammograms reduces the FN rate in each tissue type individually when GLCM and LBP features were extracted.

In 2012 Domingues et al used a two block method to classify normal mammograms. In the first phase the authors classified the mammograms into dense and fatty and extracted GLCM and LBP features from only fatty mammograms since dense mammograms were immediately sent to expert evaluation. In a second phase they used Support Vector Machines (SVM) classifiers created with an Radial Basis Function (RBF) kernel function to classify the mammograms one more time, but in this case in terms of malignancy [7].

III. MATERIALS AND METHODOLOGY

A. Mammogram Database and Density Classification

In this work we used the Cranio-Caudal (CC) view of mammograms belonging to INbreast Database, a recently proposed database. The images contained in this database were acquired at the Breast Center in Centro Hospitalar São João with MammoNovation Siemens Full-field Digital Mammography.
sets that we used. As preliminary results of the present study we used only the GLCM set of features and the K-Nearest Neighbour Classifier (KNN). Given the original purpose, - not miss any malignant cases - we introduced two variations in the KNN classifier: 1) the concept of asymmetric distances, using a parameter $\alpha$ to distinguish the distance of the malignant classes, approaching them to the test instance and 2) if one of the neighbours is malignant, the test is classified as malignant. With this two variations we obtained better results. Some of the results can be seen in Table I and II.

### Table I

<table>
<thead>
<tr>
<th>TissueType</th>
<th>K</th>
<th>$\alpha$</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense</td>
<td>9</td>
<td>0.24</td>
<td>100%</td>
<td>0%</td>
<td>24%</td>
<td>76%</td>
</tr>
<tr>
<td>Fatty</td>
<td>4</td>
<td>0.105</td>
<td>100%</td>
<td>0%</td>
<td>8%</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.11</td>
<td>98%</td>
<td>2%</td>
<td>14%</td>
<td>86%</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>TissueType</th>
<th>K</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense</td>
<td>6</td>
<td>100%</td>
<td>0%</td>
<td>6%</td>
<td>21%</td>
</tr>
<tr>
<td>Fatty</td>
<td>10</td>
<td>95%</td>
<td>5%</td>
<td>8%</td>
<td>94%</td>
</tr>
</tbody>
</table>

### V. Conclusions and Future Work

The presented results, although not quite satisfying, are a first step in the implementation of a pre-CAD system. Future work consists in the extraction of LBP features and the classification through the other classifiers mentioned in the methodology: Support Vector Machines and Random Forests. The performance of LBP and GLCM features along with the three different classifiers should be compared, in order analyse for each tissue type which one is more suitable.

### REFERENCES


