Computational Methods for Classifying Breast Cancer Molecular Subtypes in Mammograms

Vivek kumar Singh*

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
vivekkumar.singh@urv.cat

1 Purpose

There is an unstoppable tendency towards personalized medicine in order to achieve both, diagnosis and treatment, and monitoring more effective for each patient. In this line, we propose the P-BreasTreat work, aimed at the personalized treatment of breast cancer by developing new computational techniques for image and data analysis. The ultimate purpose is to improve the effectiveness of current methods for determining the level of malignancy associated with that cancer tumors and also to propose models to prevent relapse and improve the quality of life of the patients.

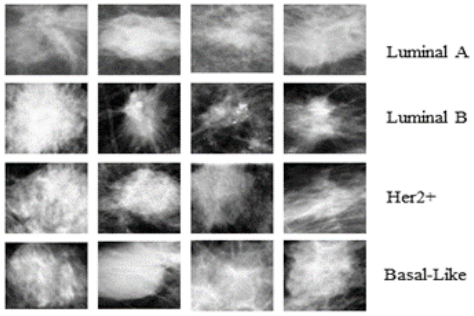

Fig. 1: 16 examples of input image samples, 4 for each of the 4 molecular subtypes of breast cancer

In proposed work, we will develop computer technologies for distinction and initial screening of the 4 molecular subtypes of breast cancer shown in figure 1 (Luminal A, Luminal B, Her2+ and Triple Negative) as advanced support to the traditional pathologic analysis. The impact will be focused

* PhD advisor: Dr. Santiago Romani Also and Dr. Domenec Puig Valls

on reducing both the number of biopsies and adverse psychological effects on patients. To do this, we will design specific methods of medical image analysis by using Computer Vision and Artificial Intelligence techniques, aimed at designing new adaptive biomarkers.

Once detected the molecular subtype, we will design customized models for the diagnosis and monitoring of patients treated with neoadjuvant therapy, conservative surgery and radiotherapy, in order to provide new tools for predicting relapse (either local or remote) of breast cancer, anticipating corrective measures to improve the rate of recovery. These models will also highlight critical points of the treatment or disagreements with the clinical standards (analysis of adherence). In order to do this, we will apply automatic process mining techniques to the evolutionary data of the patients.

2 Related Works

Numerous approaches have been proposed to classify the BC tumor subtypes based on histological information. The method designed by Perou et al. [2] performed a BC classification into certain "intrinsic" subtypes based on gene expression patterns. Herbeck et al. [3] presented the guidelines for the BC molecular subtype categorization based on several immunohistochemistry (IHC) biomarkers such as estrogen receptors (ER), progesterone receptors (PR), human epidermal growth factor receptor 2 (HER-2) and antigen KI-67 (Ki67).

Torrents-Barrena et al. [4] presented the first work to determine the feasibility of using a CAD system to differentiate among all BC molecular subtypes in mammograms. Authors designed two classification experiments: "Luminal A vs. Luminal B", and "Luminal A vs. Luminal B vs. Her-2+ vs. Basal Like". Support Vector Machines (SVM) and Local Binary Patterns (LBP) yielded the best accuracy: 75% and 52.17%, respectively. Moreover, they designed in [5] a new methodology based on fractal texture analysis and unsupervised / supervised classifiers. SVM also achieved the best performance (76.48% and 55.67%, respectively). The main drawback of both works was the limited number of Her-2+ and Basal Like samples.

3 Proposed Method

In this abstract, we propose a semi-automatic CAD system to classify the four molecular subtypes of BC from full-field digital mammograms (FFDM). A modified VGG16 [6] convolutional neural network architecture is presented to learn the underlying micro-texture patterns of the mammogram image pixels for each subtype.

Our hypothesis is that a CNN conveniently designed can learn the prototypical underlying micro-textures of each cancer subtype and that those

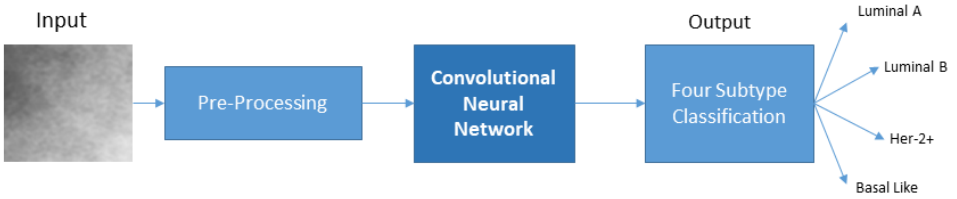

Fig. 2: Work flow process of four molecular subtypes classification of breast cancer

prototypes are characteristic of each subtype shown in figure 3, i.e., they are similar to all samples of the same subtype but different from the micro-texture prototypes of the other cancer subtypes. Hence, the trained CNN should be able to predict the subtype of any new breast tumor, given a ROI sample extracted from its corresponding segmented mammography.

We will base our design on the VGG_{16} architecture, since it uses small area filters (3x3) that we expect they are well suited for micro-texture prototype learning, in contrast to other CNN architectures (e.g. AlexNet) that use larger filters (11x11) to look for edges, macro-textures or other salient features of the objects.Since our CNN must learn just pixel-wide micro-texture prototypes (not the full tumor shapes) of only four classes of cancer subtype, we have checked several simplifications of the VGG_{16} original architecture.Concretely, we have defined smaller sets of filters and reduced the number of neuron layers.

4 Experimental Results

The experimental data are composed of 179 DICOM mammograms (CC and MLO views) distributed in 64, 63, 25 and 27 samples for classes Luminal A, Luminal B, Her-2+ and Basal-like, respectively. These medical image samples are provided by a Oncology Group in Spain.

Firstly, we have checked the performance of our model by training and validating the network with regards to the first two classes, Luminal A and Luminal B, which correspond to the less aggressive cancer subtypes. Our network has performed really well on Luminal A samples, achieving a 95% of accuracy. On the other hand, just 61% of Luminal B samples had been correctly classified, while the remaining 39% had been misclassified as belonging to Luminal A.Nevertheless, our network renders an overall accuracy around 78%, which is quite a good result taking into account the evident lack of visual patterns in the image samples. The second experiment corresponds to the full 4-class classification, i.e., including all breast cancer subtypes. From the individual accuracies, we can obtain an overall accuracy as the weighted

average with respect to the number of test samples of each class, obtaining a fair 67% of good predictions.

5 Conclusion and Future work

In this abstract, we have presented a supervised BC molecular subtype classification method based on a CNN that analyse manually selected areas of breast tumors found in DICOM images of mammograms. To the best of our knowledge, this is the first effort to predict the molecular subtypes of malignant tumors just from image excerpts of digital mammograms using CNNs. Before, we tried other approaches to the same problem using classical texture descriptors (Uniform Local Binary Patterns, Histogram of Gradients, Gabor filters, Fractal dimension), but with less degree of accuracy ([6]: 75% — 52%; [8]: 76% — 56%; current approach: 78% — 67%). Other authors have only focused on automatic detection of tumors and determining if the tumor is benign or malignant. Future work will aim at validating our approach on larger datasets of MRI images, with the ultimate objective of gradually bringing computerized assistance to BC molecular subtypes classification into clinical practice.

Acknowledgement. This research has been partly supported by the University Rovira i Virgili through Marti-Franques Research Grant and Spanish Government through project DPI2016-77415-R.

References

[1] G. Shieh,,C. Bai, and C. Lee Identify breast cancer subtypes by gene expression profiles. *Journal of Data Science*, 12:165–75, 2004.

[2] C. M. Perou, T. Srlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees and Fluge. Molecular portraits of human breast tumours .*Nature*, 6797: 747–752, 2000.

[3] N.Harbeck, C. Thomssen, and Gnant Brief preliminary summary of the consensus discussion. *Breast care*, 8(2): 102–109, 2013.

[4] J. Torrents-Barrena , D. Puig, L. Diez-Presa, M. Arenas, and P. Radeva Assessment of a multidiscriminant supervised classifier driven by textural features to distinguish molecular subtypes of breast cancer. *In International Conference of Computer Assisted Radiology and Surgery (CARS)* 10(1): S31-S32, 2015.

[5] J. Torrents-Barrena, A. Valls, P. Radeva, M. Arenas and D. Puig Automatic Recognition of Molecular Subtypes of Breast Cancer in X-Ray images using Segmentation-based Fractal Texture Analysis. *In 18th International Conference of the Catalan Association of Artificial Intelligence (CCIA)* 277: 247-256, 2015.

[6] K. Simonyan, and Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv,* 277: 1409.1556, 2014.