November 27, 2016 **Project Proposal**

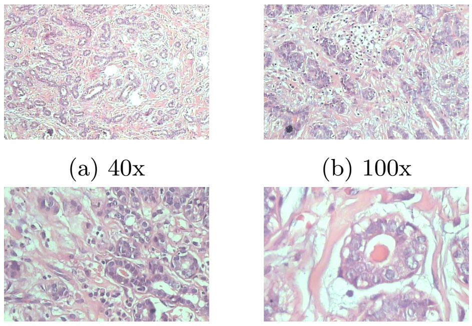
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Breast Cancer Classification

based on histopathological images

Introduction

Automatic disease classification based on medical data is becoming one of the hot topics in research. Indeed, nowadays clinicians spend a lot of their time analysing data such as medical images for disease diagnosis. Providing a tool for computer assisted disease classification would alleviate this extremely time-consuming task. Since the patient's life could depend on the correctness of the classification result, it is primordial to achieve very high classification accuracies. In this project, classification of breast tumors will be studied on a new large database of patho-histologic images containing different types of benign and malignant tumors. Both binary and multi-class classification will be considered.



(a) 40x

(b) 100x

(c) 200x

(d) 400x

Figure 1: Examples at different magnification

Data acquisition

The data that will be used in this project is contained in the BreakHis database [1]. This database, which has been published only very recently, has as objective to provide data scientists with a benchmarking dataset for their classification algorithms in the context of breast cancer histology. The full content of the database in terms of images is summarized in table 1. There is a total of 7909 (resolution 700x460) images of 82 patients for 4 magnification levels, 4 benign and 4 malignant tumor types. An example is shown in figure 1. Since this is a database specially made for benchmarking classification systems, it is already cleaned by the publisher and as such no or very little data cleaning will be needed.

Magnification	**Benign**	**Malignant**	**Total**
40x	652	1370	1995
100x	644	1437	2081
200x	623	1390	2013
400x	588	1232	1820
Total Images	2480	5429	**7,909**

Table 1: Summary of images in dataset

Data exploration

The data will be explored by visualising images from different classes, discussing their differences and similarities and inspecting basic properties and image features with histograms and scatterplots to check if basic features could provide a way of easy tumor type classification.

Data exploitation

The main goal of this project is to build a machine learning model which is able to discriminate between different tumor types. Spanhol et al. [1, 2] and Bayramoglu et al. [3] use either texture descriptors with basic classifiers like LDA, QDA, k-NN, SVM,... or convolutional neural networks, which learn the features themselves, to tackle this problem and report the accuracy they can achieve in discriminating between benign and malignant tumors. However, no multi-class discrimination results have been reported. Therefore, in this project, first a model will be built to do binary classification and once good accuracy will be achieved, multi-class classification will be considered as well. Besides the inherent challenges in classifier design, coping with the large amount of data on a personal laptop will also be challenging. Patch based classification or spatial subsampling could be considered.

Evaluation

The evaluation of the proposed methods will be done, for instance, by considering the classification accuracies compared to the ones reported in the literature and the malignant tumor detection sensitivity since this is the most important parameter in a clinical context. Training and testing times will be reported as well, but are not extremely important in the context of computer aided disease classification.

Bibliography

[1] Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, jul 2016.

[2] Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. 2016.

[3] Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä. Deep learning for magnification independent breast cancer histopathology image classification. 2016.