Automatic Breast Cancer Classification from Histopathological Images

SticAmsud - meeting #2

Fabio Alexandre Spanhol* — PhD student
Luiz Eduardo Soares de Oliveira*, PhD — professor advisor

*Department of Informatics
UFPR

Curitiba, 2014

Presentation Outline

Quoting

"*An urgent need in cancer control today is to develop effective and affordable approaches to the early detection, diagnosis, and treatment of breast cancer among women living in less developed countries.*"

Christopher Wild (2013), IARC's director

Motivation

Cancer incidence and mortality worldwide

Cancer is a significant public health problem in the world!

- increase of 20% in last decade resulting in 27 million of new cases until 2030 according to WCRF (World Cancer Research Fund);
- 8.2 million of deaths in 2012, according to IARC (International Agency for Research on Cancer)/WHO (World Health Organization);

Motivation
Breast Cancer

Breast cancer is second most common (excluding skin cancer) for women!

- mortality of breast cancer is very high with regard to the other cancer types;
- while cancer mortality increased by 8% in 2012, the mortality rate of breast cancer was 14% in the same period.

Motivation
Cancer and Breast Cancer in Brazil

INCA (National Institute of Cancer José Alencar Gomes da Silva)/MS (Health Ministry) estimates **576 thousand** new cases of cancer in 2014

- Compared to other countries, Brazil shows higher rates of diagnostics for breast and prostate cancer;
  - Breast cancer presents 59 cases for 100 thousand inhabitants (worldwide average about 43 for 100 thousand)

Motivation
Detection and Diagnosis of Breast Cancer

Diagnosis from a histopathology image remains the "gold standard" in diagnosing almost all types of cancer

- current pathology diagnosis is based on the subjective opinion of pathologists (experts)
- digitized tissue histopathology has become feasible to the application of computerized image analysis and machine learning techniques
  - large increases in computational power, cheap storage and improvement in image analysis algorithms
  - CAD (Computer Aided Diagnosis) systems have emerged for disease detection, diagnosis, and prognosis prediction to complement the opinion of the pathologist

Tumor

Tumor = Neoplasm

Neoplasia — Transformed cells in continuous replication produce a neoplasm (tumor)

- categories are based on a judgment of a tumor's potential clinical behavior
- **benign**
  - relatively "innocent" and slow-growing
  - remain localized
- **malignant**
  - **cancer is synonym!**
  - grow faster
  - lack of differentiation (*anaplasia*)
  - lesion can invade and destroy adjacent structures (locally invasive) and spread to distant sites (metastasize) to cause death

Tumor
Nomenclature

Stroma (neoplasic cells of the tumor) determines the tumor's name

- Benign — in general, designated by attaching the suffix *-oma* to the cell type from which the tumor arises, e.g., *fibroadenoma*.
- Malignant — when the origin are
  - "solid" mesenchymal tissues or its derivatives are called *sarcomas*
  - epithelial cells are called **carcinomas**
  - **breast cancer is a type of carcinoma.**

Pathology

pathos = suffering + *logos* = study

Pathology provides the scientific foundation for the practice of medicine.

- It involves the investigation of the causes of disease and the associated changes at the levels of cells, tissues, and organs, which in turn give rise to the presenting signs and symptoms of the patient.

Pathology
Main areas

Cytopathology

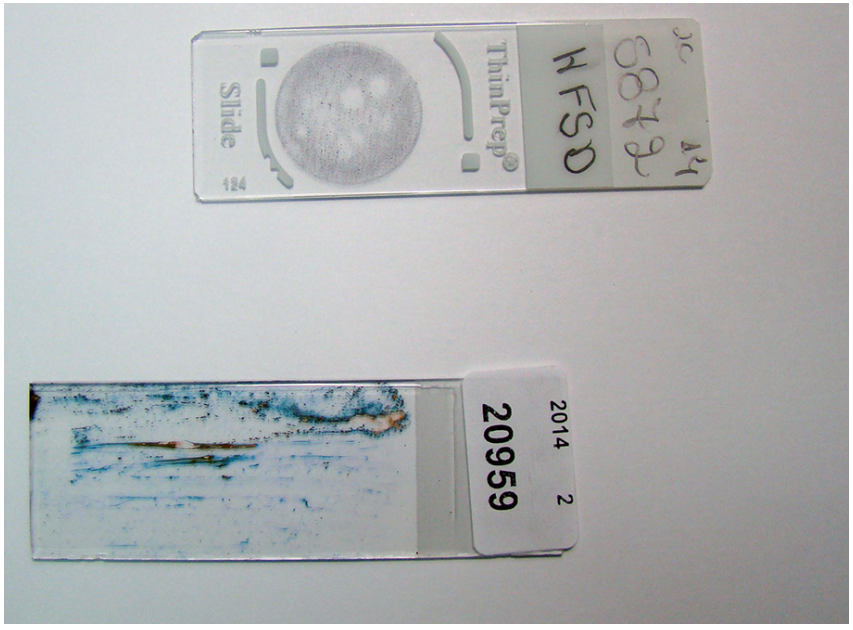
- study of cells, main cellular components (nucleus and cytoplasm)
- features including size, shape, color, texture and topography

Histopathology

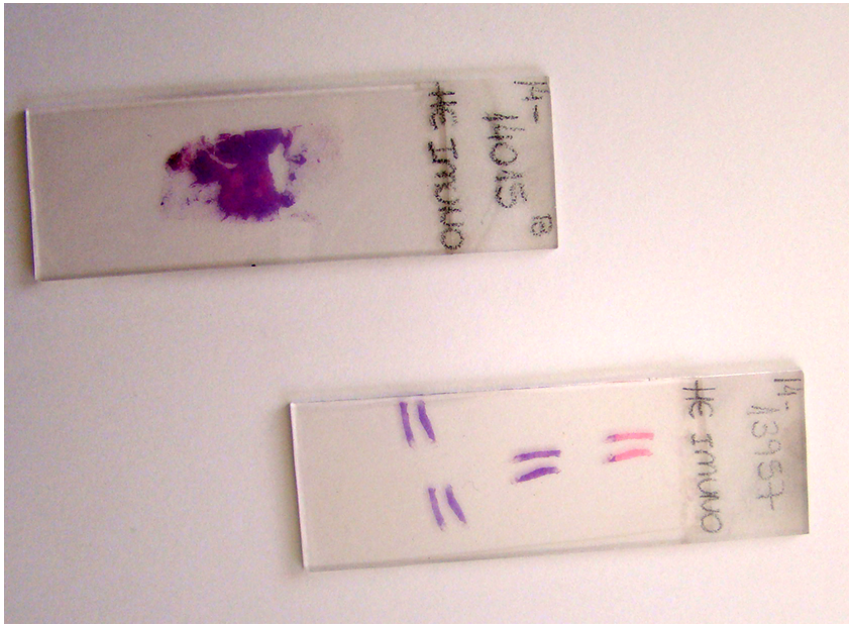
- anatomical study of the microscopic structure of tissues (grouped cells forming complex structures)
- analyze disease states at a cellular level by means of light and/or electron microscopy, histochemistry and immunochemistry

Pathology
Cytopathology x Histopathology

Samples may be fixed across a glass microscope slide for subsequent staining and microscopic examination.



Cytopathology: Liquid-Based Cytology (above) x Conventional (bellow)



Histopathology: Surgical Biopsy (above) x Core Needle Biopsy (bellow)

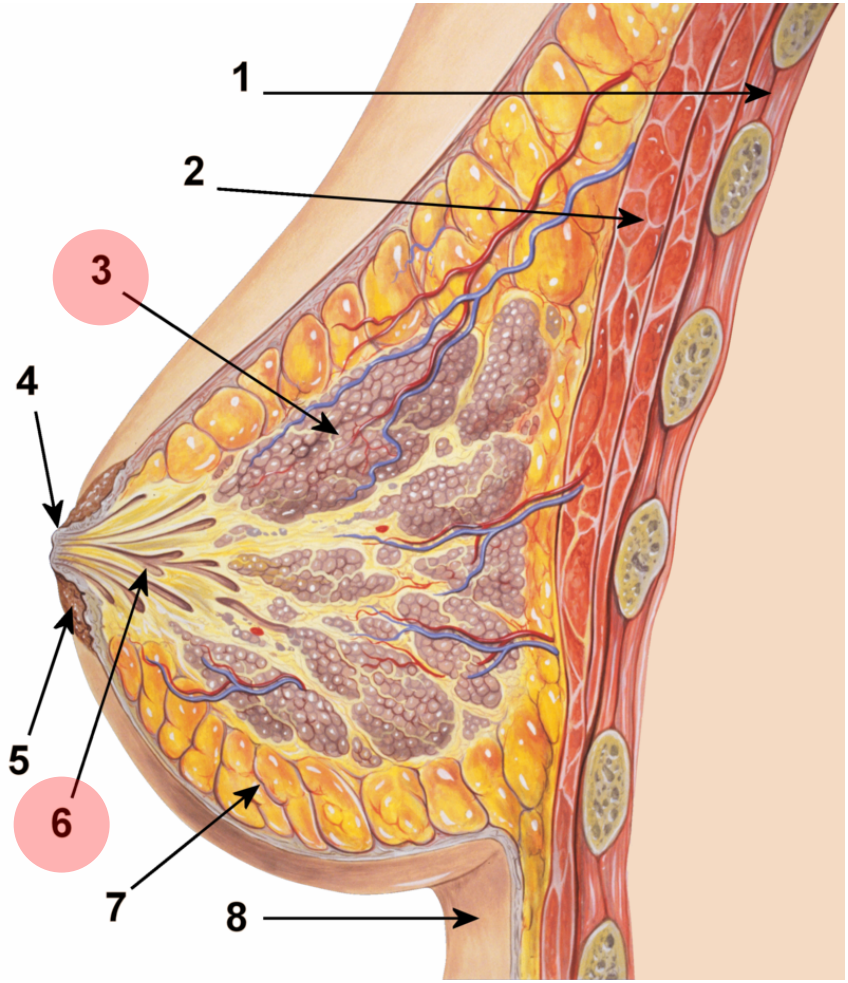
Breast Cancer

Breast Cancer is the cancer of breast tissue

- cells can invade nearby tissues and can spread through the bloodstream and lymphatic system (lymph nodes) to other parts of the body
- Early breast cancer usually does not cause pain and may exhibit no noticeable symptoms
- As the cancer progresses, signs and symptoms can include a lump or thickening in or near the breast; a change in the size or shape of the breast; nipple discharge, tenderness or retraction; and skin irritation, dimpling or scaliness.

Breast Anatomy

Female, side view

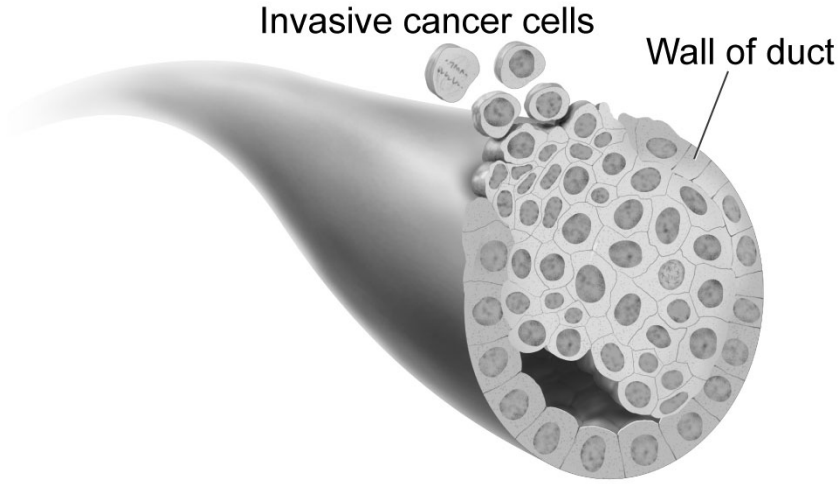


1. Chest wall
2. Pectoralis muscle
3. Lobules
4. Nipple
5. Areola
6. Milk Duct
7. Fatty tissue
8. Skin.

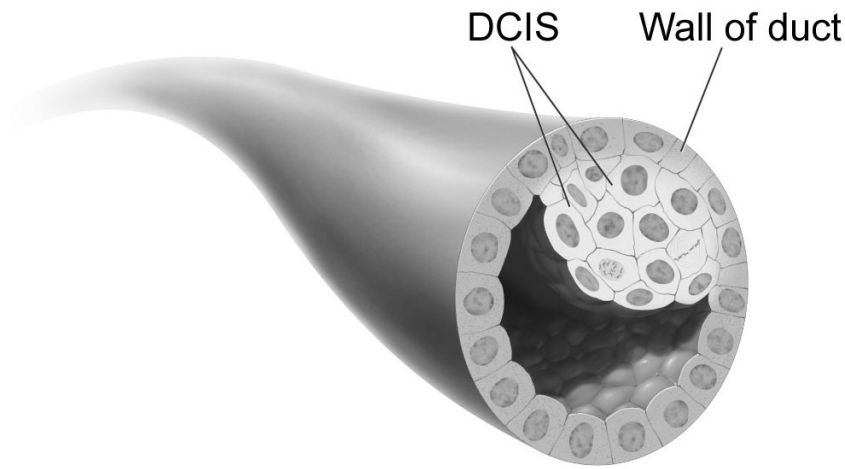
Breast Cancer
Invasive x *In Situ*

Breast cancers are classified according to whether they have or have not penetrated the limiting basement membrane.

Invasive (infiltrating) cancers do grow into normal, healthy tissues.



*Non-invasive (**in situ**) cancers stay within the milk ducts or lobules in the breast.*



Breast Cancer
Types/Subtypes

1. **Noninvasive**
   (a) Ductal carcinoma *in situ* (DCIS)
   (b) Lobular carcinoma *in situ* (LCIS)
2. **Invasive (infiltrating)**
   (a) Invasive ductal carcinoma ("not otherwise specified"), the most common subtype of invasive carcinoma. Term used for all carcinomas that cannot be subclassified into one of the specialized types;
   (b) Invasive lobular carcinoma;
   (c) Medullary carcinoma;
   (d) Colloid carcinoma (mucinous carcinoma);
   (e) Tubular carcinoma;
   (f) Other types.

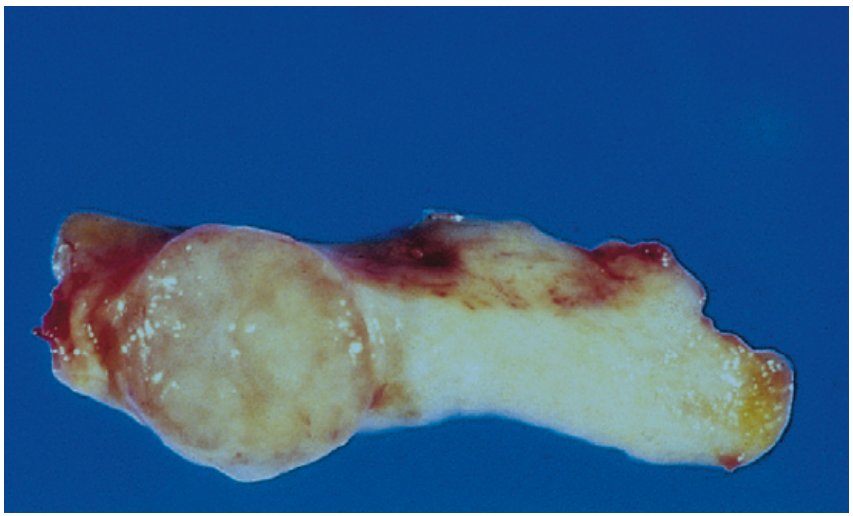
Breast Cancer
Types/Subtypes

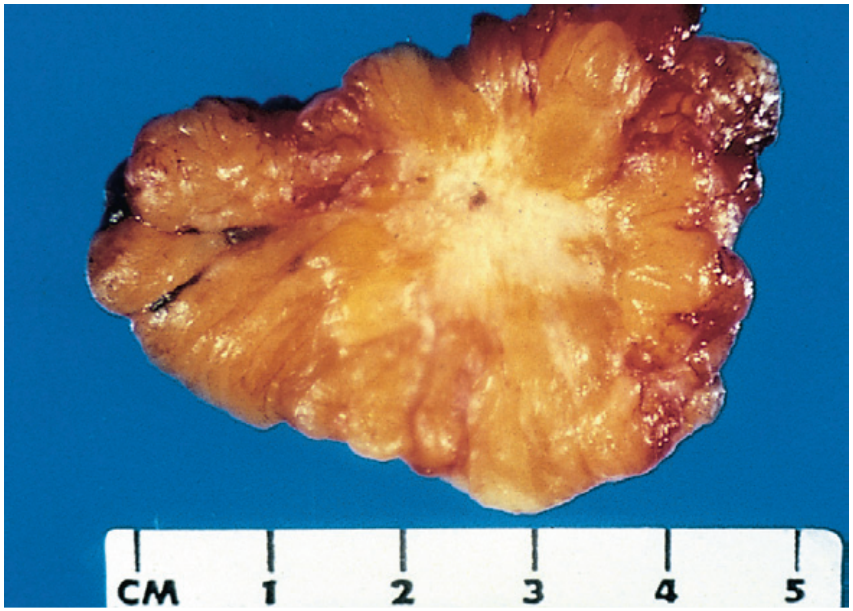
- Various types/subtypes can have **different prognoses and treatment implications**.
- Two main types
  - Invasive Ductal Carcinoma (IDC) — incidence $>$ **80 %**
  - Invasive Lobular Carcinoma (ILC) — $\approx$ **10-15 %**
- Remaining cases of invasive carcinoma are comprised of other special types of breast cancer
  - e.g. colloid medullary, micropapillary, papillary, tubular, etc.

Breast Cancer

Benign x Malignant

Macroscopic view: Fibroadenoma (benign, left) and Invasive Ductal Carcinoma (malignant, right).

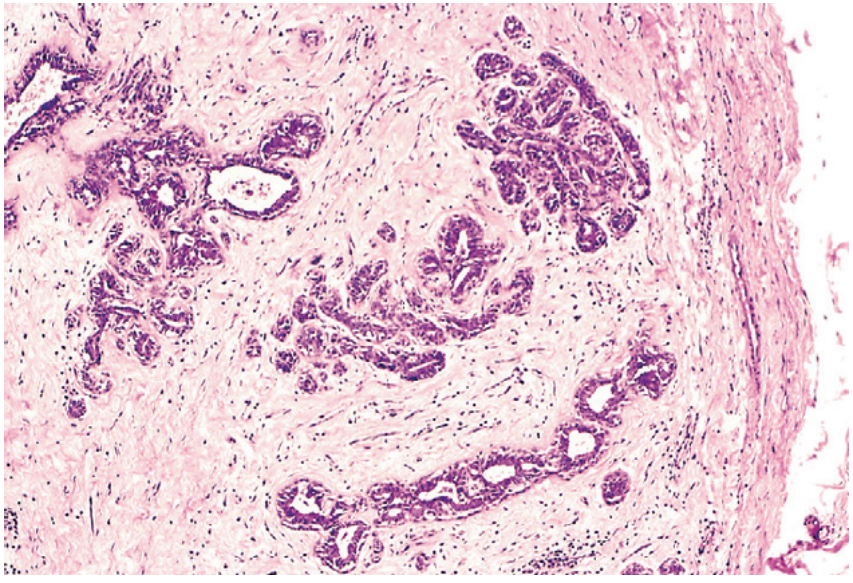




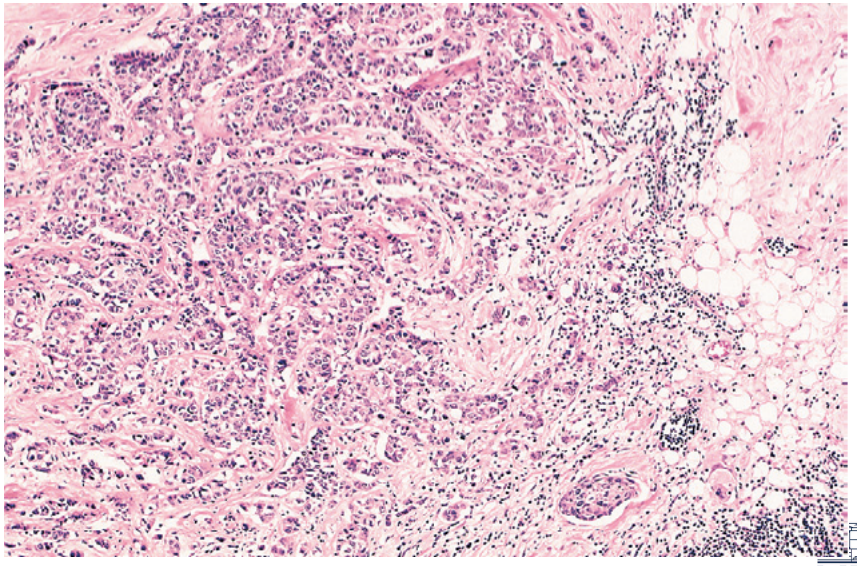
Breast Cancer
Benign x Malignant

Microscopic view: Fibroadenoma and IDC.

fibrous capsule sharply delimits the tumor from the surrounding tissue



shows invasion of breast stroma and fat by nests and cords of tumor cells



Detection and Diagnosis of Breast Cancer
Methods

Non invasive methods (**imaging procedures**)

- Diagnostic mammograms (x-ray)
- Magnetic resonance imaging (MRI) of the breast
- Breast ultrasound (sonography)
- Thermography

Biopsy — is the only way to tell if cancer is really present.

- Fine Needle Aspiration (FNA)
- Core Needle Biopsy (CNB)
- Vacuum-assisted
- Surgical

Goals

The main goal of this work is propose a **automatic classifier to assist pathologists in breast cancer diagnosis**.

Requirements

- work with microscopic scanned images (CCD acquisition)
- differentiating benign from malignant lesions
- subtyping and grading of tumors
- quick classification
- high rates of accuracy
- smallest false-negative rate
- support inexperienced specialists

Goals
Challenges?

Pattern recognition approaches related to breast cancer classification face some difficulties:

- Available training datasets are tipically inbalanced with many more benigns cases than malignant ones;
- Scarcity of public histopathological images datasets specific to breast cancer;

Goals
Specific Goals

Specific goals:

- Evaluate existing segmentation techniques and propose new alternatives;
- Evaluate existing feature extraction techiniques and propose new alternatives;
- Assess the proposed approach in real environments related to breast cancer diagnosis (clinical pathology laboratories);

Goals
Specific Goals

Specific goals:

- create a public histopathology image database related to breast cancer:
  - scanned microscope images of tissue histology labeled by expert;
  - well balanced database: instances are proportionally divided among benign and malignant cases;
  - representative population: should include carcinoma subtypes.

Current Stage

Elaboration and Documentation of New Image Dataset in Progress

Images are being collected in collaboration with a Pathological Laboratory from western region of Parana (**P&D** located at Cascavel city). New database contains digitized microscopic images from breast tissue:

- H&E stained sections ($\approx 3\mu m$ thickness) of breast tumor (benign and malignant);
- tumoral area identified by pathologist;
- confirmed by ImmunoHistoChemistry (IHC);
- for each sample, randomic shots are taken from peripheral tissue, transitional border and within tumor area
  - 10-15 color images (700x480, RGB, 24-bit) at 40x, 100x, 200x, and, 400x optical magnification

Dataset

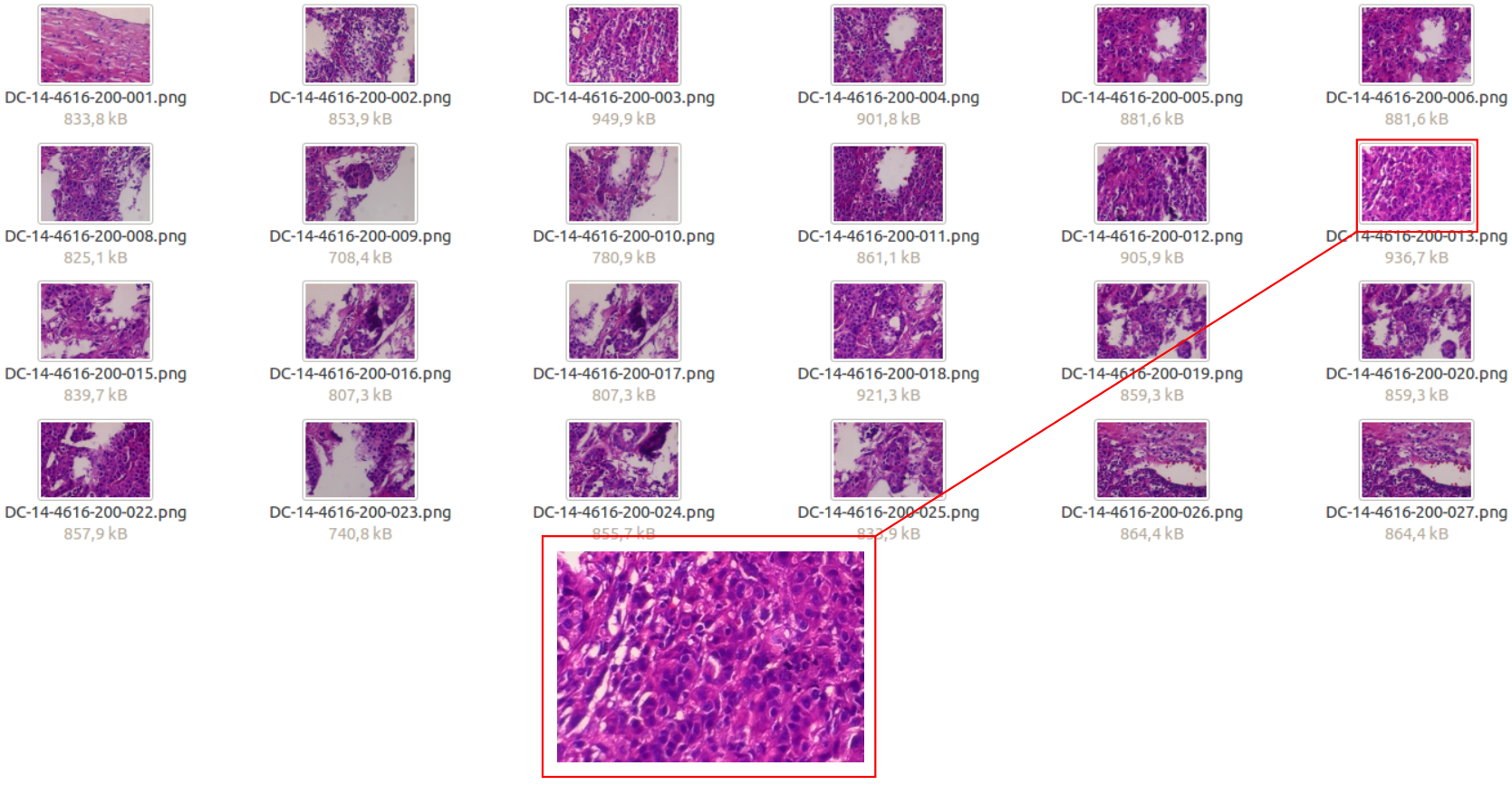
Elaboration and Documentation of New Image Dataset

Project database currently contains about **7800 images**

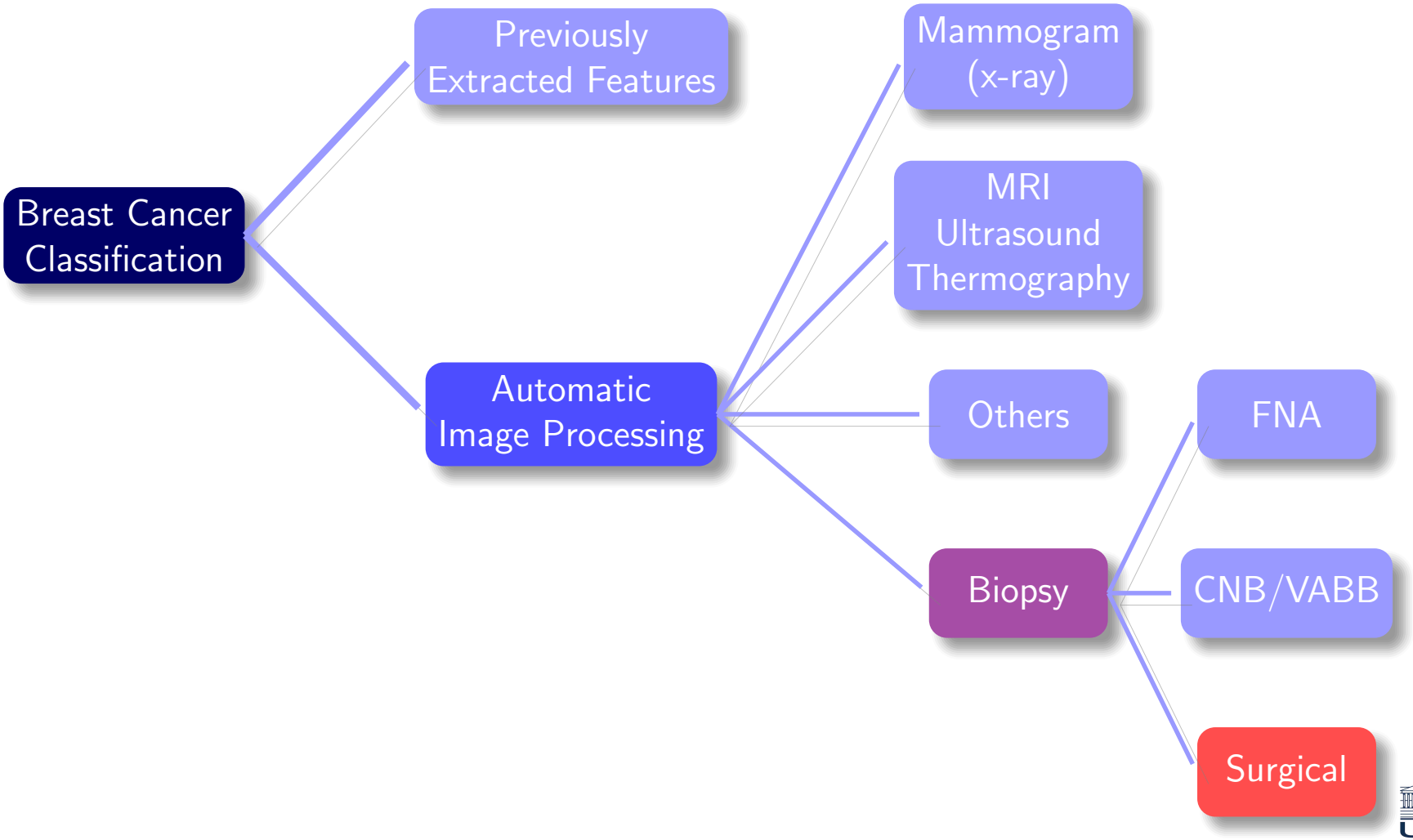
Magnification	Normal	Benign	Ductal	Lobular	Mucinous	Papillary	Total
40x	51	365	1137	181	176	88	1998
100x	33	372	1222	215	199	89	2130
200x	34	363	1141	199	157	87	1981
400x	25	317	986	174	135	95	1732
Total	143	1417	4486	769	667	359	**7841**

Dataset

Elaboration and Documentation of New Image Dataset



Related Works



Initial Experimental Results

Some Experiments Conducted on Image Dataset

Considering only **two classes: Benign and Malignant**

- images from benign tumor subtypes and breast cancer subtypes were properly grouped

1. Texture descriptors
2. *Cuda*-ConvNet
   - High-performance C++/CUDA implementation of convolutional (feed-forward) neural networks by Alex Krizhevsky, Google's neural network team
   - Python scripting facility

Initial Experimental Results
Texture Descriptors

- Experiment setup
  - randomly selected images at 100x
  - training with 616 images: 200 benign and 416 malignant
  - testing with 200 images
- Features extracted include
  - GLCM, LBP, LPQ, SURF
- Classification
  - classifiers k-NN and SVM
  - $\approxeq 80\%$ **of Accuracy**
- Apply (automatic) region of interest segmentation?
  - intelligent approach is required!

Initial Experimental Results

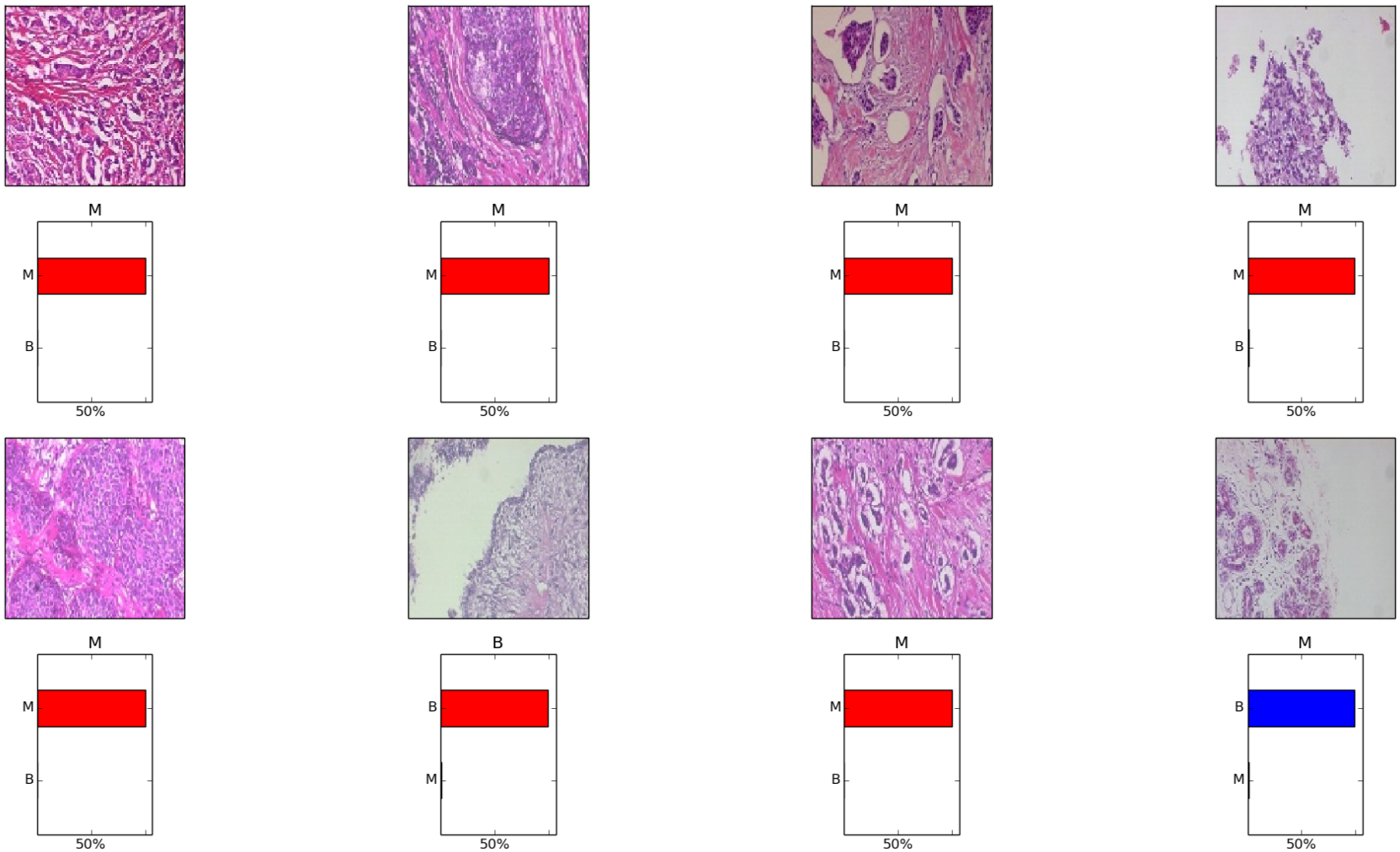
Cuda-ConvNet

- Experiment setup
  - 816 images at 100x: 299 benign and 517 malignant
  - k-fold 60-20-20
- Preprocessing
  - images were resized to 128x128
- Classification
  - 30 epochs
  - **95.7% of Accuracy** was achieved

Initial Experimental Results

Cuda-ConvNet

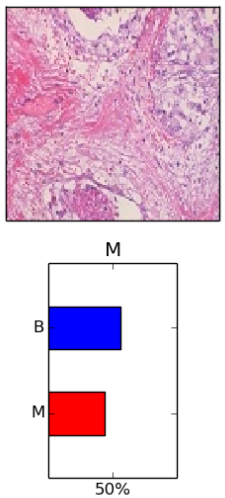
Correct test case predictions

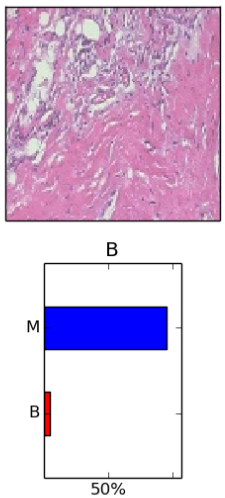


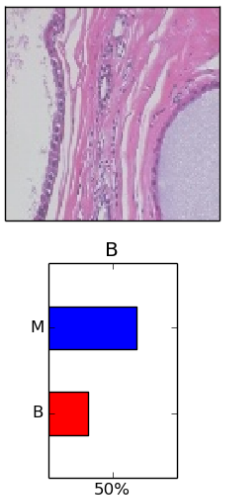
Initial Experimental Results

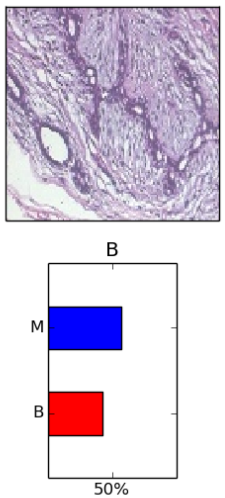
Cuda-ConvNet

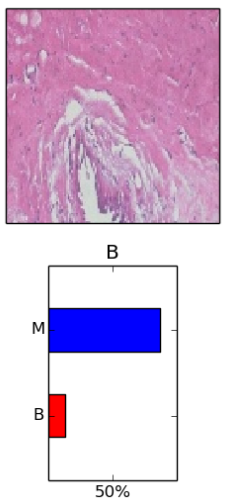
Mistaken test case predictions

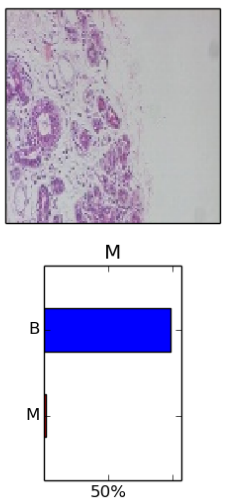


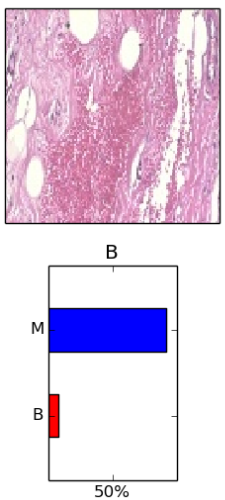













Thank You!