ITCR: A Unified Machine Learning Platform for Cancer Diagnosis

Linda Shapiro, Ph.D.
Principal Investigator

Joann Elmore, M.D.
co-Principal Investigator

Donald Weaver, M.D.
Consultant

Meredith Wu
Research Assistant

Ximing Lu
Research Assistant

Motivation

- GOAL: Develop a unified software package for sharing image analysis and machine learning tools to improve the accuracy and efficiency of cancer diagnosis
- REASONS (for breast biopsy diagnosis):
  - Millions of women depend on pathologists' interpretive accuracy.
  - Diagnostic errors are alarmingly frequent.
  - Particularly critical are the diagnostic thresholds of atypia and ductal carcinoma *in situ*, where up to 50% of cases are misclassified.

Prior work on two NIH/NCI- sponsored ROIs for scientific research studying breast and melanoma cancer biopsy diagnosis by practicing pathologists

- Our technical work has produced methodologies for a number of tasks that will lead to a unified software package for cancer diagnosis,

1. *localization of ROIs* in whole slide images
2. *analysis of pathologist search patterns*
3. *tissue type segmentation* of breast histopathology images
4. *automated diagnosis* of regions of interest

- This work has produced experimental computer programs that test the new methods, which could be extremely useful to other researchers or to practicing physicians, and has also motivated new research directions.

Aims

- Aim 1: Regions of Interest. Produce
  - 1a) a *ROI-finder classifier* and associated tools for use by researchers or pathologists for automatic identification of potential ROIs on whole slide images of breast biopsy slides and
  - 1b) a *ROI-analysis classifier* and associated tools that can point out image regions that tend to cause misdiagnosis and produce suitable warnings as to why such regions may either be distractors or indicate cancer.

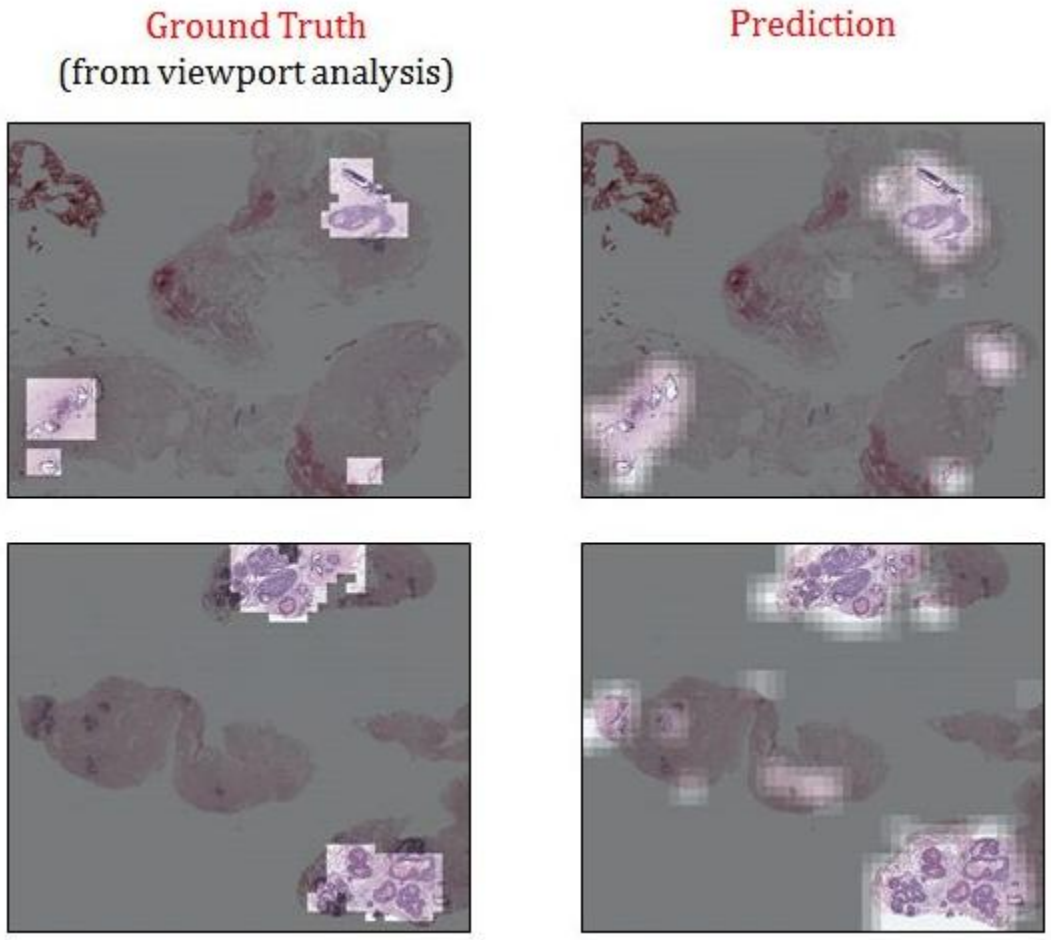
Aims (continued)

- Aim 2: Diagnosis: Produce a *diagnostic classifier* and associated tools that can not only suggest the potential diagnosis of a whole slide image, but can also produce the reasons for the diagnosis in terms of regions on the image, their color, their texture, and their structure.
- Aim 3: Dissemination: Develop a *unified software package* containing this suite of tools, so they can be easily shared and provided (standalone and through the existing Pathology Image Informatics Platform (PIIP)) to both cancer researchers and clinical pathologists.
- Provide the *methodology to train related classifiers* for other biopsy-diagnosed cancers, such as melanoma, prostate, lung, and colon cancer.

Aim 1: Bag-of-Words Approach to Finding ROIs

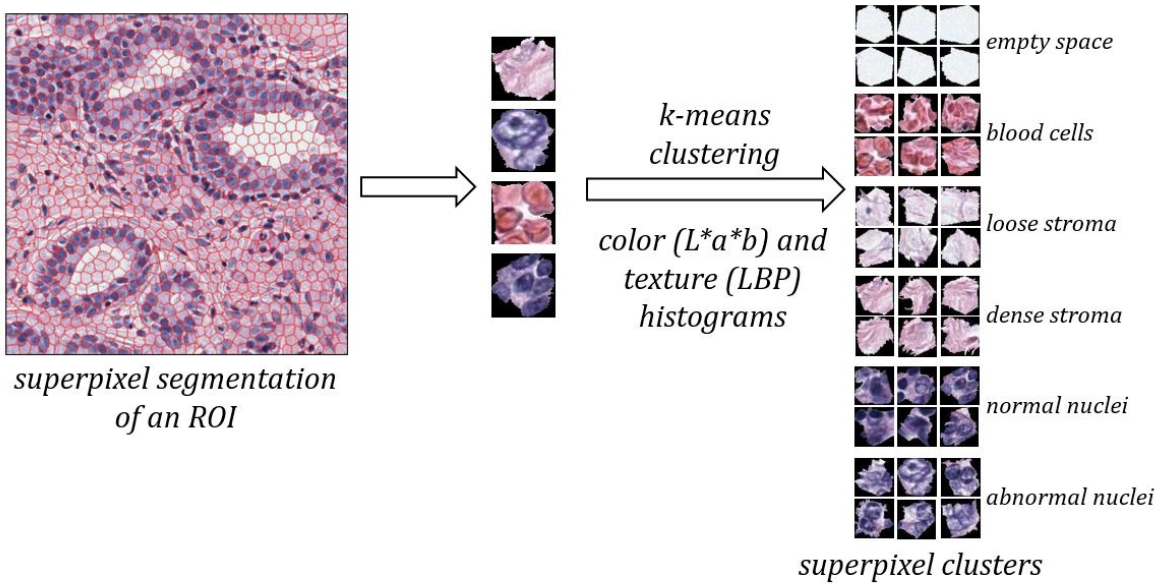
- K-means clustering of small patches produces “visual words” that form a visual dictionary.

- Histograms of larger regions characterize them and are used to train a classifier to distinguish between ROIs and non-ROIs.



Aim 2: Automated Diagnosis

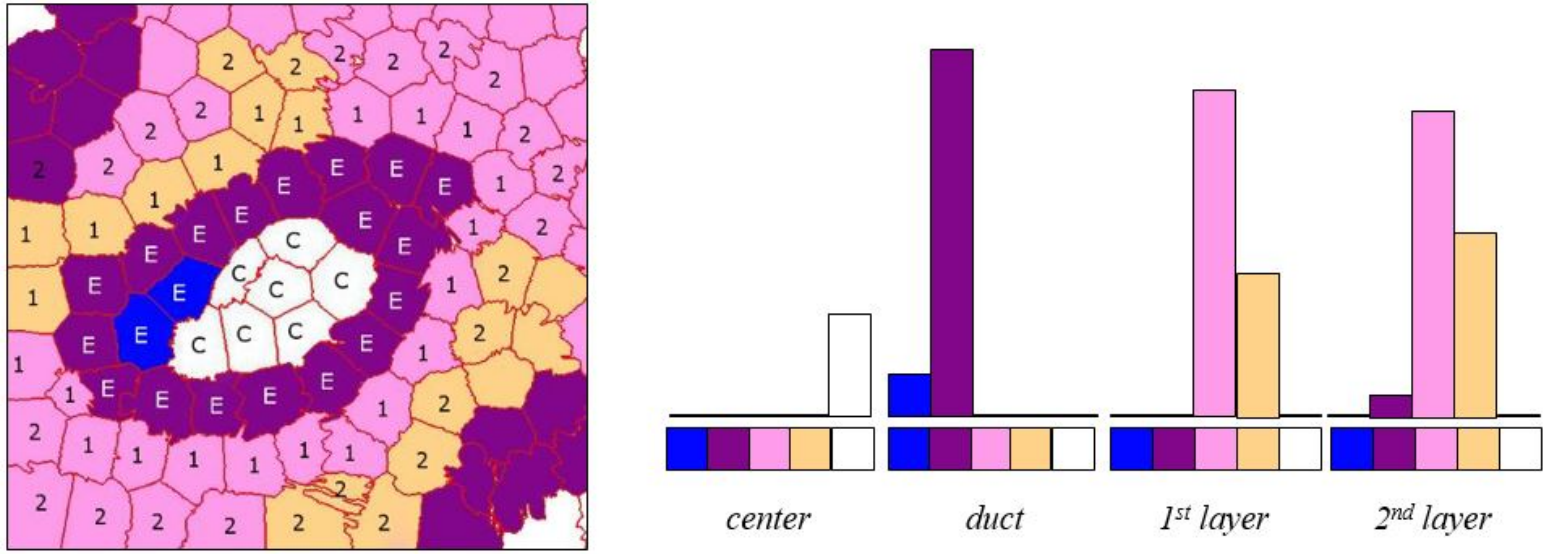
1. Tissue segmentation and classification
   - We used superpixel (Ren and Malik) clustering to initially identify important tissue types and their features. Both SVM feature-based and CNN pixel-based classifiers were developed.



2. Diagnostic Classification

- Multiple methods were compared from simple color/texture histograms to our "structure feature". Software will include all of them.

- Below, histograms of tissue types of 4 layers about a duct: white = background, blue = malignant epithelium, purple = bening epithelium, pink = normal stroma, and orange = desmoplastic stroma.



Aim 3: Dissemination

- We will make our specific classifiers that have been trained on our own data available for other users to run, and we will provide classifier code and scripts for researchers who want to develop their own classifiers with their own data or public data.
- We are working with the PIIP group (PI*s*: Anant Madabhushi and Anne Martel) to integrate our software into their platform and have it run under the Sedeen browser.

Table 3: Software tools we will provide.

ROI
ROI feature extractor
ROI pre-trained classifier
tools for training new ROI classifiers
ROI characteristic comparison
ROI analysis
ROI visualization
Tissue Segmentation
superpixel construction
superpixel clustering
supervised (CNN-based) segmentation classifier
tools for training new CNN-based segmentation classifier
superpixel segmentation visualization
Diagnostic Classification
superpixel frequency feature extractor
superpixel co-occurrence feature extractor
structure feature extractor
superpixel frequency and co-occurrence SVM classifier
structure feature classifier
tools for training new classifiers
diagnostic classification visualization

Collaborators

- Close Collaborators
  - Dr. Joann Elmore, UCLA. Co-PI
  - Dr. Donald Weaver, U of Vermont, consultant
  - Dr. Selim Aksoy, Bilkent University, long-time collaborator
  - Dr. Stevan Knezevich, melanoma pathologist in practice
- Others
  - Dr. Anant Madabhushi, Case Western University
  - Dr. Anne Martel, PathCore
  - Dr. Larry True, Genitourinary Pathology, UW
  - Dr. Mrinal Mandal, U of Alberta, works on melanoma pathology