Classification of Cells Based on Scale-space Measures and Semi-supervised Machine Learning

S. Vohra[2], L. Antanas[2], L. Raedt[2] and D. Prodanov[1]

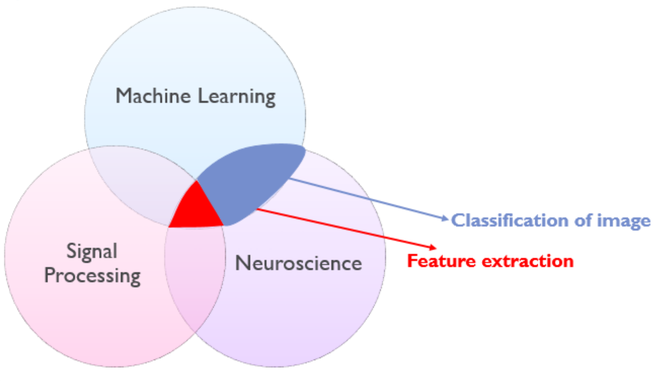
[1]EHS, Imec, Leuven, Belgium, [2] KU Leuven

Introduction

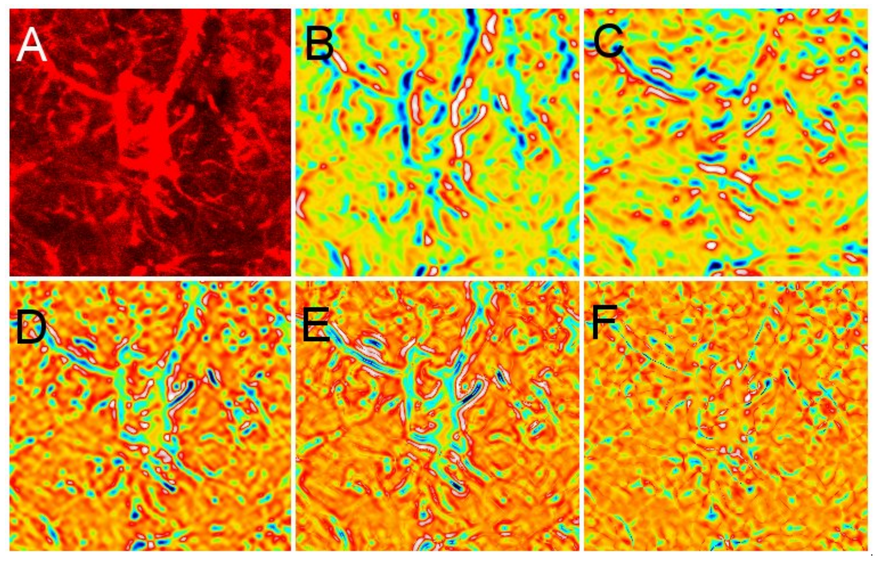
Semi-supervised active learning-based approaches appear to be very promising in a variety of applications from document classification to biological cell segmentation. Multi-scale approaches in image analysis can be used in variety of segmentation applications, for example scale spaces and wavelets. In this poster we present an approach combining scale space feature extraction and machine learning for the application of semi-automated classification of histological microscopic images. The approach can be further used towards automatic characterization of different cell phenotypes.

Problem Outline

Given: A set of Images
Find: Segmentation & Classification of Astrocytes



Features Extraction



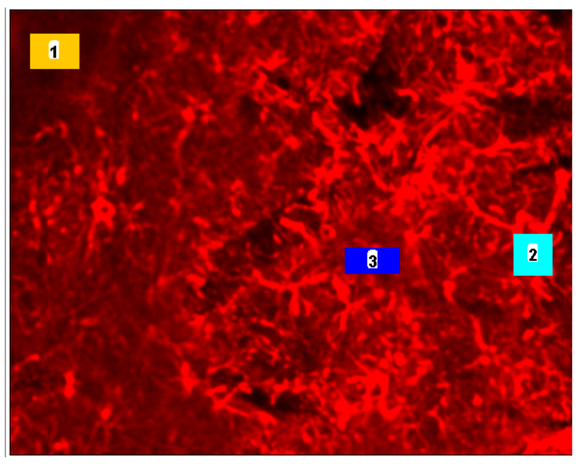
A: Astrocytes stained for GFAP. **B** - horizontal component of the gradient; **C** - vertical component of the gradient. Anisotropic decomposition of the LoG, i.e. Mexican hat (**C**) ; orthogonal (**D**) and tangential component (**E**).

Summary and Outlook

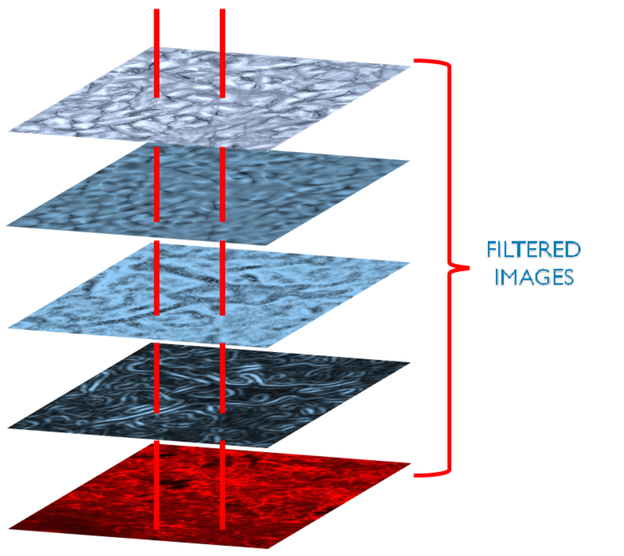
- We are able to segment 3 different regions representing cells of interest (astrocytes), tissue and tissue defects with more than 80 % success on holdout dataset.
- Results demonstrate that our method is capable of coping with the noise variability and is able to segment accurately astrocytic arbors.
- Selected filters represent features of interest.
- The approach needs only few images to learn the physical object model

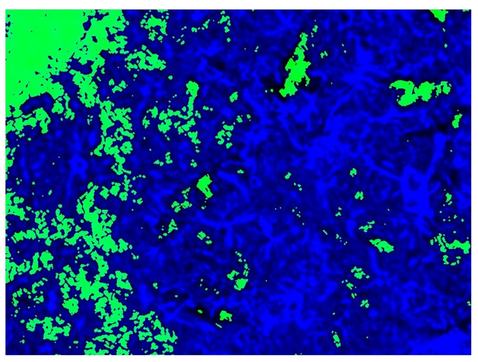
Future efforts will include implementing the Gibbs sampling and Mixture of Gaussians for finding the suitable patterns in astrocytic arbors.

Semi-supervised Machine Learning

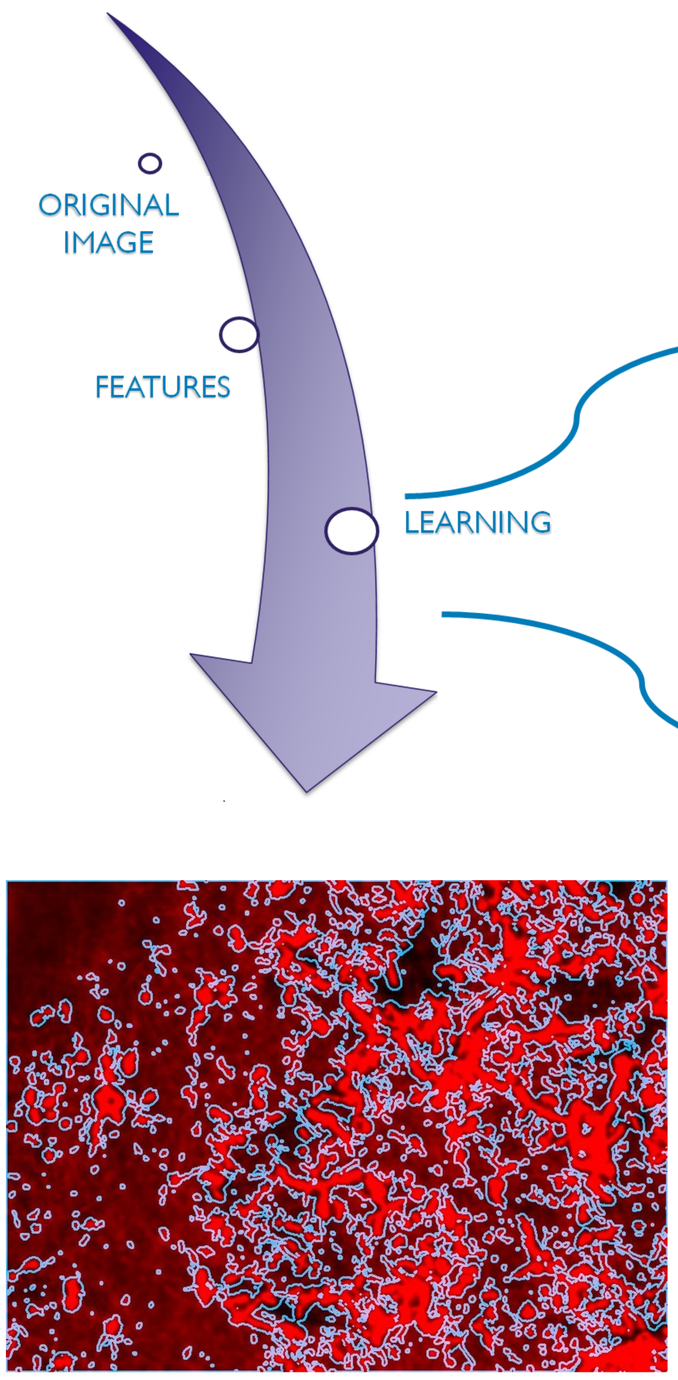


1 - TISSUE DEFECT
2 - ASTROCYTES
3 – EXTRACELLULAR MATRIX

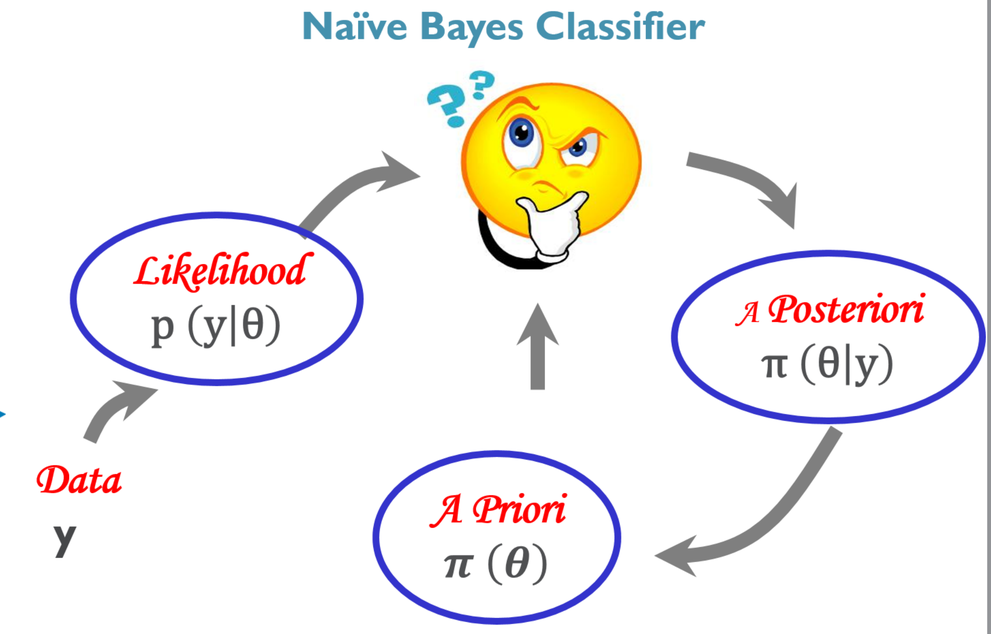


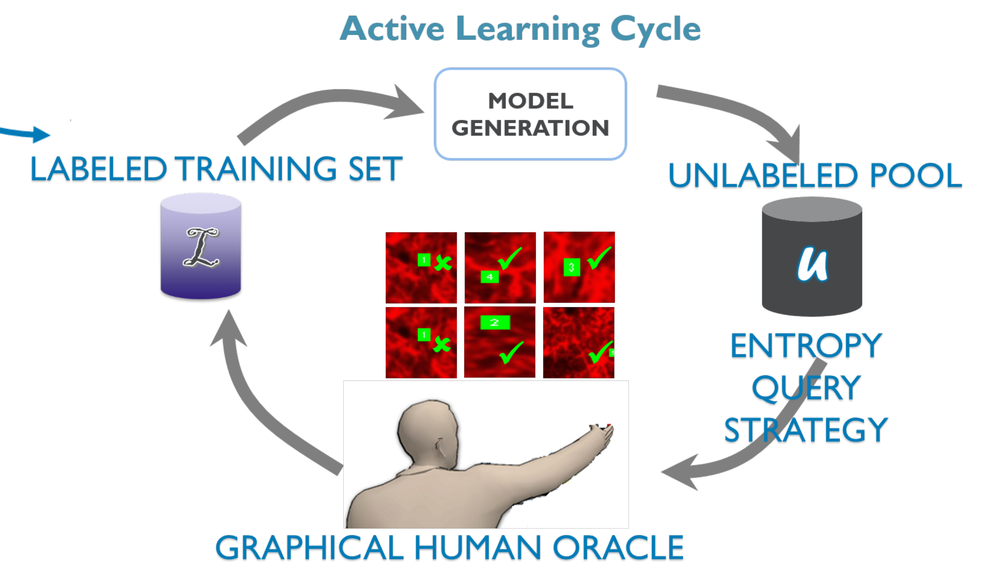


TISSUE DEFECT SEGMENTATION



ASTROCYTES SEGMENTED IMAGE





We computed Shannon's entropy H of different feature distributions to measure the uncertainty of candidate queries in the pool. We are selecting 5-10 samples in each iteration and labeling is done by the "human oracle".

$$H(X) = -\sum_{i=1}^{n} P_\theta(y|x) log P_\theta(y|x)$$