A Machine Learning Approach to Cell Classification

MoHan Zhang[a], Cindy Tan[b] and Dhananjay Bhaskar[b]

Supervisors: Dr. Leah Edelstein-Keshet[b], and Dr. Calvin Roskelley[c]

[a]Department of Mathematics, University of British Columbia

[b]Faculty of Applied Science, University of British Columbia

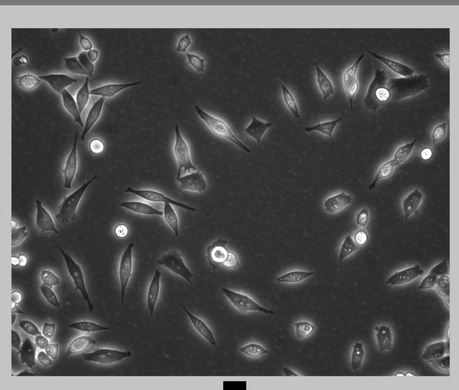
[c]Department of Cellular and Physiological Sciences, University of British Columbia

Introduction

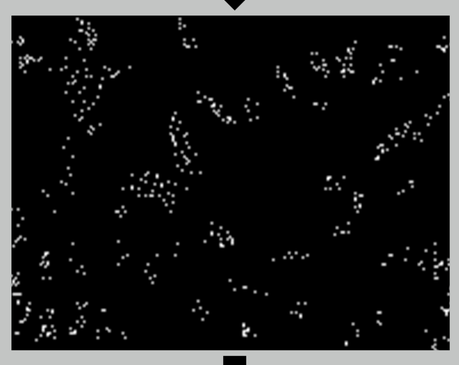
The precise regulatory mechanism that governs cell shape, size and polarity is not well understood. To facilitate a systematic investigation of cell morphology, we have developed tools to identify cells from live imaging data, quantify cell geometry and automatically classify cells using unsupervised machine learning. This poster illustrates our methodology.

Step 1: Image Processing

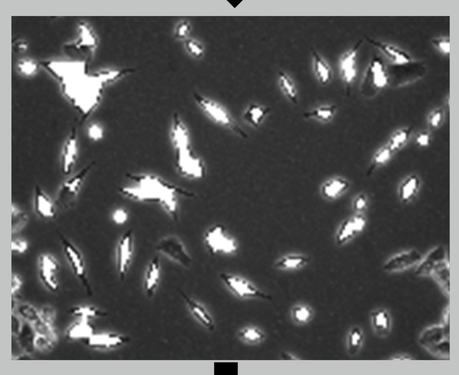
MIA PaCa-2 Cell Line
In vitro Model for Pancreatic Carcinoma
Phase-Contrast Microscopy



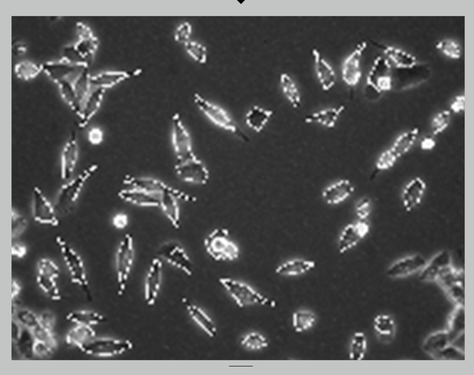
Edge Detection
Sobel–Feldman Derivative Filter
Grayscale Binarization (Thresholding)



Foreground Binary Mask
Mathematical Morphology
Erosion, Dilation, Opening and Closing



Segmentation
Distance Transform
Marker-Based Watershed Algorithm



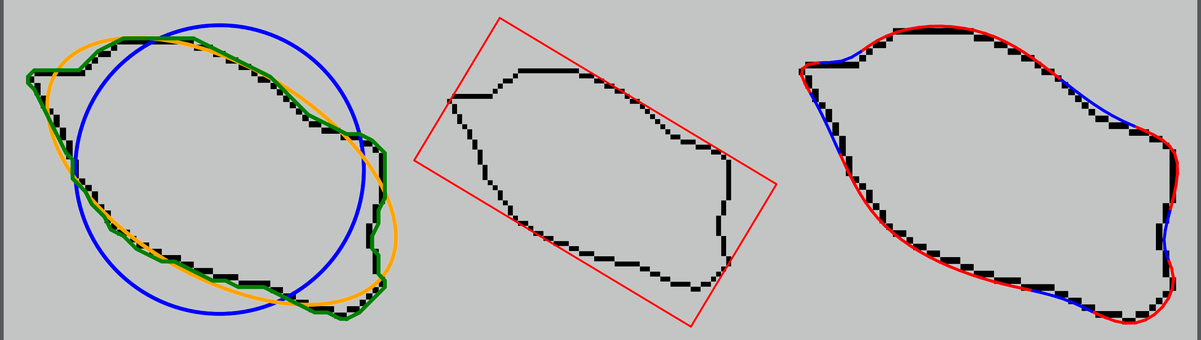
We obtained 149 correct segmentations from 20 images. 63 cells exhibiting circular, elliptical, elongated and protrusive morphology were manually selected for feature extraction.

Step 2: Feature Extraction

Consider an arbitrary geometry $f(\theta) = 0$ parametrized by $\theta = (\theta_1, ..., \theta_M)^T$. To fit this geometry to a set of boundary points $(x_i, y_i)_{i=1}^N$ (assuming $N > M$), we solve the following optimization problem:

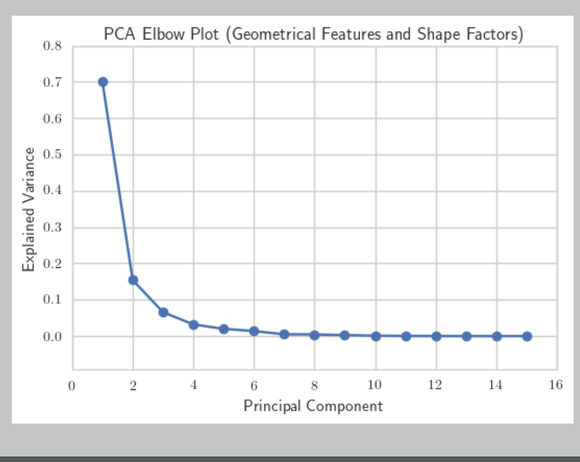
$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} r_i^2(\theta)$$

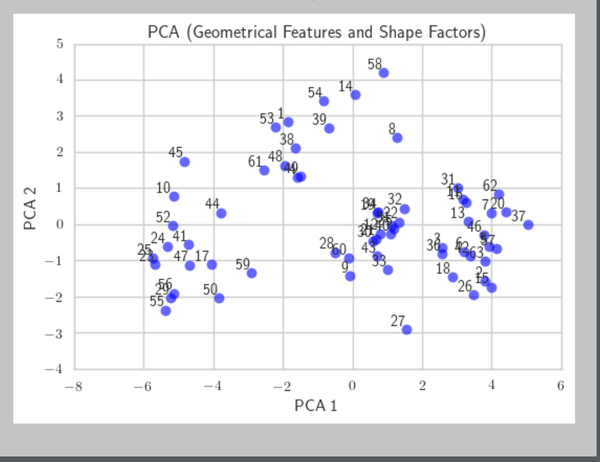
where r_i is the orthogonal distance between boundary point (x_i, y_i) and shape $f(\theta) = 0$.



Step 3: Principal Component Analysis (PCA)

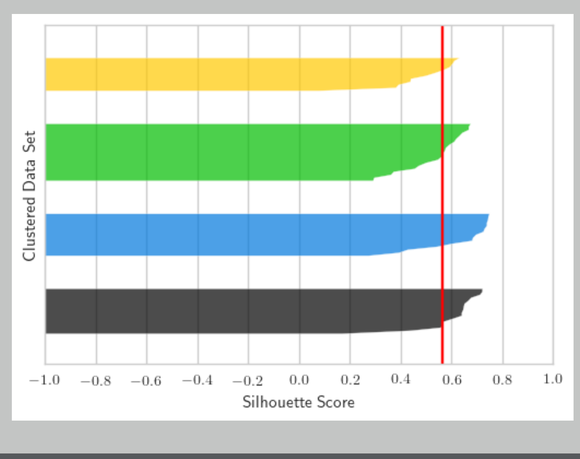
We exploit correlation between features to project high dimensional feature vector to two dimensional space where points can be easily clustered:

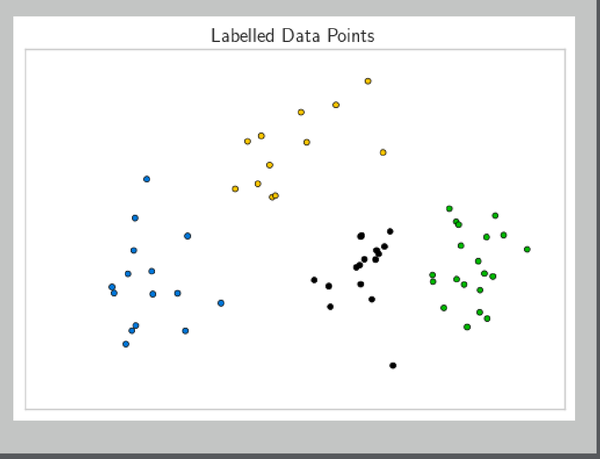




Step 4: Cluster Identification

Silhouette score analysis identified four clusters using K-Means algorithm:





Step 5: Validation

We observe that cells are correctly classified by morphology:

Left: Elliptical cells corresponding to black cluster labels

Right: Elongated cells corresponding to green cluster labels

Left: Protrusive cells corresponding to yellow cluster labels

Right: Circular cells corresponding to blue cluster labels

Future Work

- Compute additional boundary features and quantify cell shape symmetry
- Implement new methods to identify clusters in higher dimensions
- Incorporate motion-based features from time-lapse microscopy

Source code: https://github.com/dbhaskar92/Research-Scripts

For bibliography please contact: dbhaskar92@math.ubc.ca