Morphology-Based Cell Classification

NSERC USRA Report - Summer 2016

Supervisor: Dr. Leah Keshet

Darrick Lee

1 Introduction and Motivation

This summer, I worked with Professor Leah Keshet and Dhananjay Bhaskar on a project to automatically classify cells based on their shape (morphology) using methods from machine learning. There were two main applications that we were interested in. One application was to investigate the appearance of leader cells, which have been observed to "pull" a group of cells to facilitate collective cell migration (Fig. 1 left). We sought to distinguish leader cells from the rest of the cells by understanding the differences in their shapes. Our analysis was performed on experimental data provided by Dr. Calvin Roskelley. Another application was to study the relationship between the underlying biochemistry of the cell and cell shape. Specifically, we wanted to understand how different proportions of Rac and Rho GTPases within a cell would affect its morphology [1]. For this application, we studied data simulated from a Cellular Potts model (Fig. 1 right).

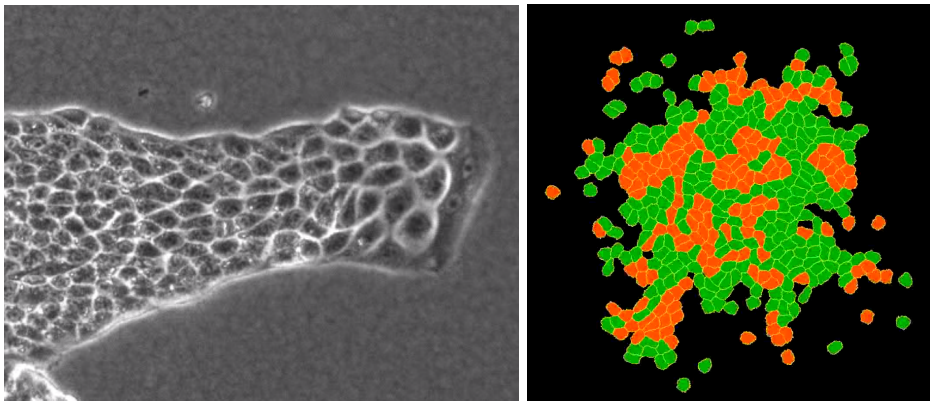

Figure 1: (Left) A screenshot of a video showing collective cell migration, with a large leader cell at the front of the group. Video was taken from [3]. (Right) A screenshot of a Cellular Potts model simulation, where colors provide the classification of the cells. The simulation was produced by Dhananjay Bhaskar.

For both applications, the main idea is to input some representation of these cells into an algorithm, which then predicts the correct classification (leader vs. non-leader or high-Rac vs. high-Rho). This was done by sampling a variety of unsupervised and supervised machine learning techniques. Broadly speaking, unsupervised learning algorithms are those which attempt to classify unlabelled data, meaning the algorithm has no prior knowledge of which classifications exist. Alternatively, supervised learning algorithms are given a set of labelled training data, and the algorithm attempts to find patterns in the data which corresponds to their labels.

2 Unsupervised Learning

The unsupervised learning methods used in this project were large clustering algorithms. The main idea is to calculate a set of features for each cell which are input into a clustering algorithm. The clustering algorithm then attempts to find distinct clusters of these data points in the feature space. The set of features we were interested in were morphological features such as area, perimeter, curvature, and parameters derived from fitting shapes to the cell boundary (Fig. 2). Examples of algorithms used are k-means clustering and agglomerative clustering.

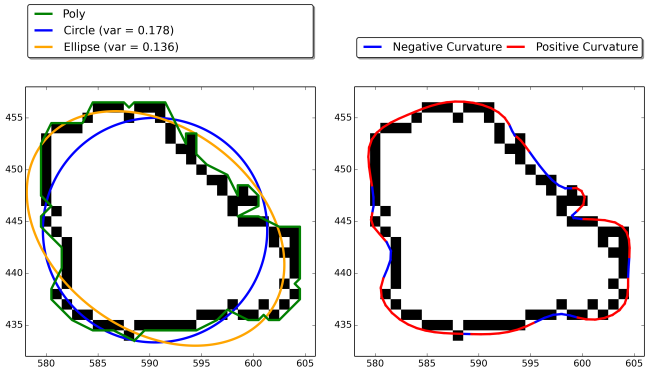

Figure 2: (Left) A sample cell with its polygon, circle and ellipse fits. (Right) A sample cell with its spline fit colored according to the sign of its curvature.

3 Supervised Learning

The main supervised learning technique used for this project was neural networks [2]. One type of neural network architecture is the fully connected network, where every neuron in one layer is connected to every neuron in the next layer (Fig. 3). The general idea is that a set of features are given to the input layer (left side of the figure), data flows through the network via a set of tunable functions, and the predicted classification is output at the final layer. These networks are trained using labelled data and the network uses optimization techniques to tune function parameters in the internal layers. For our purposes, we can use the features extracted for the unsupervised learning algorithms. Lastly, we attempted to use convolutional networks in which the image is input into the network, and the network attempts to learn features on its own.

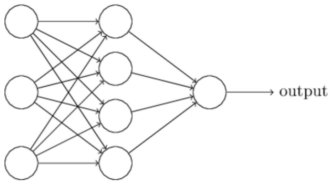

Figure 3: An example of a simple neural network. Image taken from [2]

References

[1] Chris Bakal, John Aach, George Church, and Norbert Perrimon. Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology. *Science*, 316(5832):1753–1756, June 2007.

[2] Michael A. Nielsen. Neural Networks and Deep Learning. 2015.

[3] Naoya Yamaguchi, Takeomi Mizutani, Kazushige Kawabata, and Hisashi Haga. Leader cells regulate collective cell migration via Rac activation in the downstream signaling of integrin 1 and PI3k. *Scientific Reports*, 5:7656, 2015.