

# VIP IPA: Nuclei Detection and Counting

Authors: Tikhon Pachin, Yu-Hsuan Lin, Annapoorna Prabhu, William Stevens, Shang-Hsuan Lee  
Mentors: Prof. Edward J.Delp, Prof. Carla Zoltowski

PURDUE  
UNIVERSITY

## INTRODUCTION

Nuclei and cell segmentation on microscopy images have been a topic of interest for over 10 years. There are multiple different algorithms that have been developed to target the issue. Segmentation is an important step in cancer prediction and detection as well as disease classifications. Last semester we focused on Watershed segmentation with some preliminary image processing.

This semester the work was split into two major steps: segmentation optimization and infection detection. Segmentation optimization focused on improvement of the team's algorithm. Infection detection step analyzed and developed a convolutional neural network to distinguish healthy and infected cells.

The ultimate output of the algorithm is a collection of individual cell images extracted from the original image and classified either as healthy or infected.

## DATA

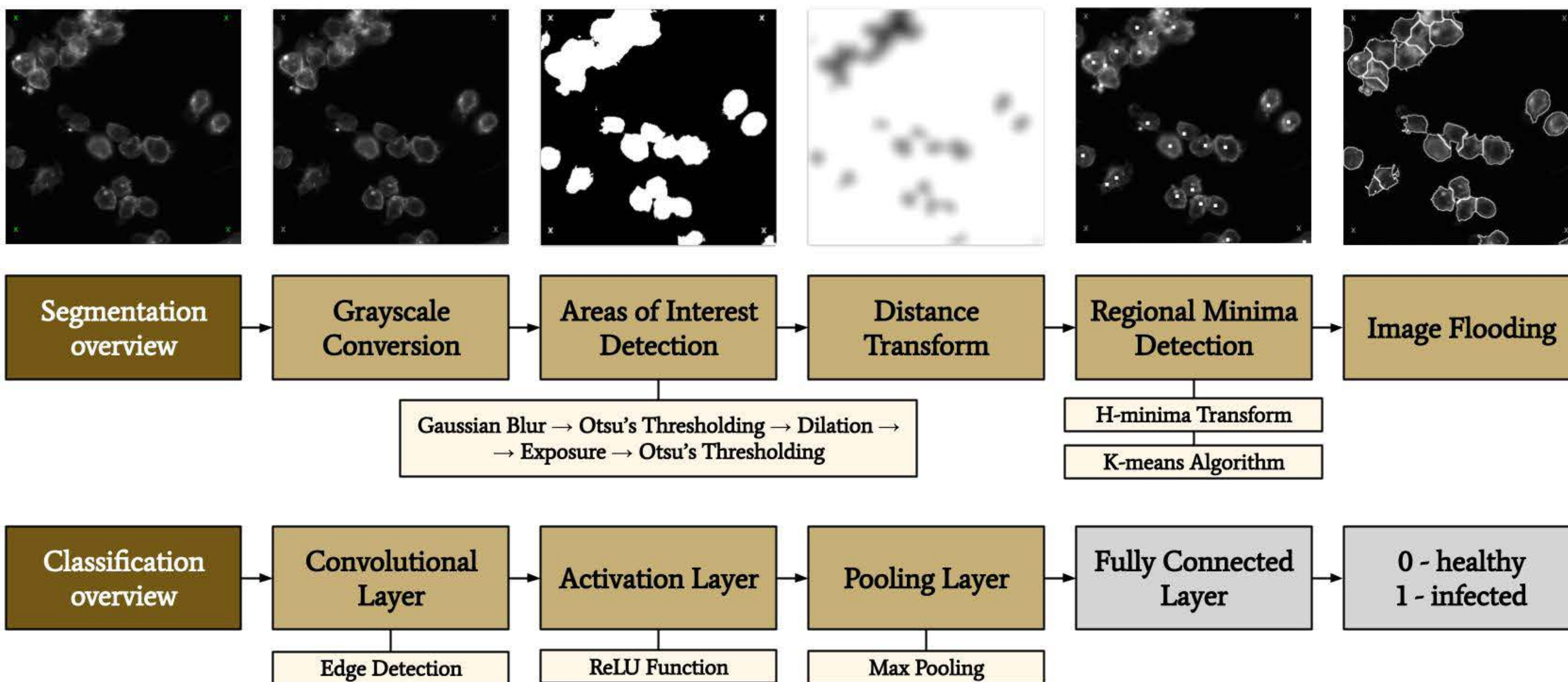
For our project we used Broad Bioimage Benchmark Collection resources that provide open source and ready to use microscopy images for testing purposes.

It was not possible to find cancer research imagery, so we focused our efforts on malaria detection and classification. The procedure of the algorithm would remain the same. The only difference is the training data set fed into the convolutional neural network.

Data source: <https://bbbc.broadinstitute.org/>

Segmentation testing data: <https://bbbc.broadinstitute.org/BBBC002>

Classification testing data: <https://bbbc.broadinstitute.org/BBBC041>



## METHODS

### Segmentation methodology:

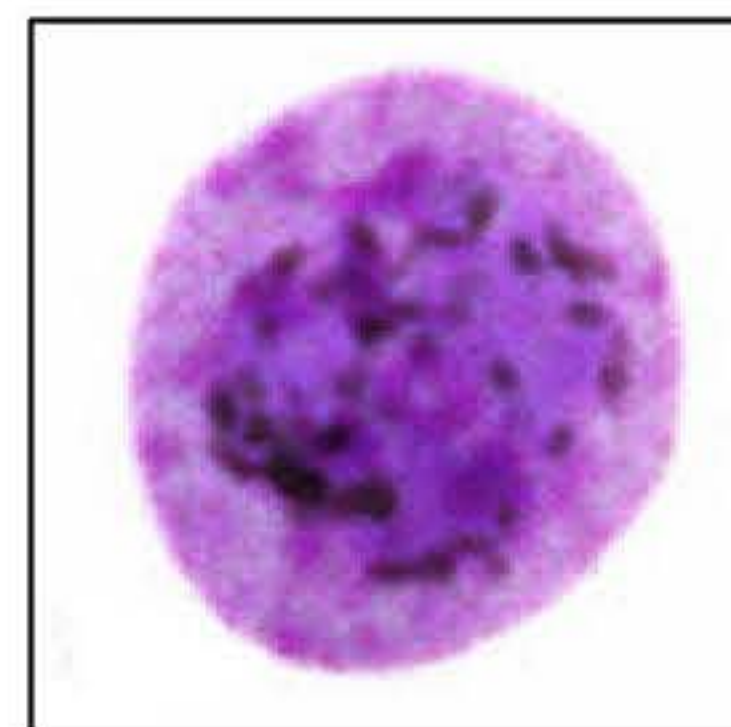
Cell segmentation is performed through **Watershed segmentation**. It translates the original image into a topographical map and “fills it” with water. The watersheds are placed at the water collision points, representing cell borders. The following steps were experimentally determined for image processing:

- **Grayscale conversion:** ITU-R BT.709:  $Y = 0.2126R + 0.7152G + 0.0722B$ .
- **Areas of interest detection algorithm:** a sequence of filters specifically designed to eliminate holes within each cell if such are present. Holes are a common issue that throws off the topographical map extraction.
- **Distance transform:** represents an image as a topographical map. The center of the cell is the “highest” point since it is farthest from the background.
- **Regional minima detection algorithm:** points of flooding detection algorithm. Utilizes K-means algorithm for best position detection. It also requires the number of cells estimation to be fed into the K-means search.
- **Flooding algorithm:** increases the water level in each basin until collision with the background or another body of water. Overflow protection in this step is essential to avoid accidental spillage between different basins.

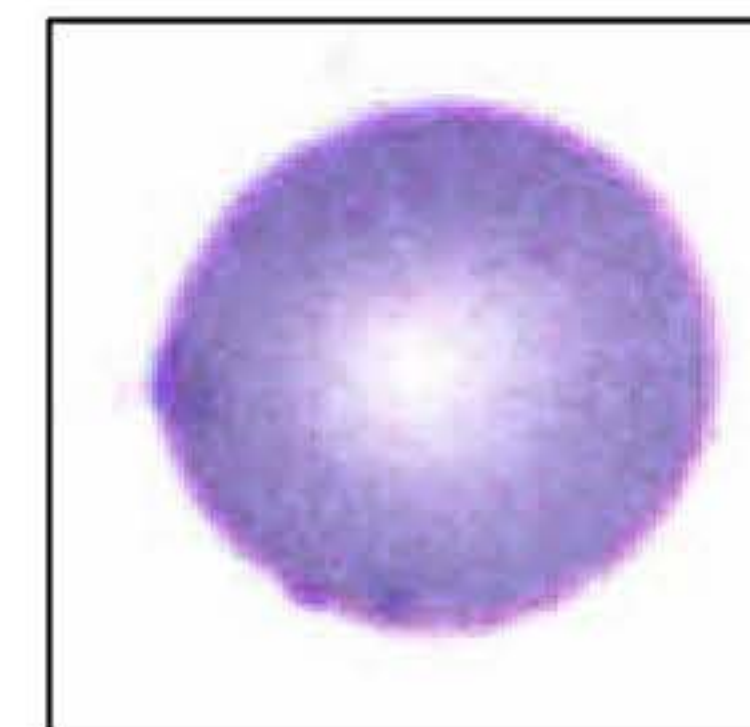
### Classification methodology:

Cell classification is performed with the help of Convolutional Neural Networks (CNN). The following is the preliminary design proposition:

- **Convolutional layer:** feature extraction through simple edge detection.
- **Activation layer:** amplification of the extracted features.
- **Pooling layer:** downsampling of the extracted features.
- **Fully connected layer:** certainty level assignment if the cell is infected or not.



1 - Infected cell



0 - Healthy cell

## FINAL RESULTS

The team has managed to complete the segmentation process with the results visible above. Individual cells are segmented with a very high accuracy. Non-heavily clustered cell segmentation is visually acceptable. Heavily clustered cells are still suffering from over-segmentation. Regional minima are occasionally misplaced.

The areas of interest detection algorithm has proven to be accurate with cell hole removal. The regional minima detection algorithm needs further optimization, however, the current implementation is already acceptable for further use with individual and non-heavily clustered cells. Flooding and overflow protection have proven to be successful due to lack of basin merges.

CNN design has been proposed and approved, however, it is hard to judge its accuracy before the implementation is complete and fully tested.

The next steps are CNN training, evaluation and segmentation optimization.