Automated White Blood Cell Classification

Team CELLery: Megan Kehoe, Yusi Ou, Jeff Wu, Isabella Yang

RICE ece redefining limits

Department of Bioengineering

Motivation

- Classification white blood cells (WBCs) by morphology is of interest in fields such as pathology and oncology
- Numbers of can be indicative of certain diseases
- Accurate cell counting is a long and tedious process
- Automatic isolation and classification algorithm would decrease the time needed and also improve accuracy and precision

Background

Five subtypes present in blood smear images, plus red blood cells and platelets

- Morphological Differences
- Textural Differences
- Easy for humans to classify
- However, feature segmentation and recognition is a very difficult machine learning problem

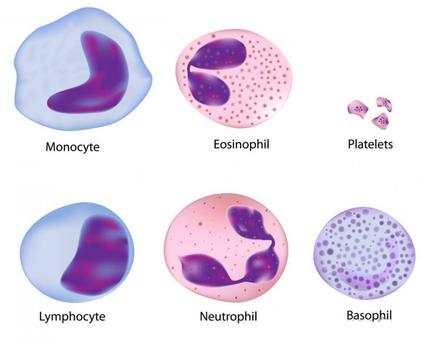

Figure 1. The five subtypes of WBCs differ morphologically and texturally. Platelets often appear in the blood smear images. [1]

Cell and Nucleus Segmentation

Bright field blood cells images are processed in Matlab R2014a

1. Original image cropped to single WBC (30 images of each class)
2. Segmented by auto-thresholding in green color channel
3. Nucleus and cytoplasm isolation is performed
4. Obtained 300 images total (150 binary, 150 gray)

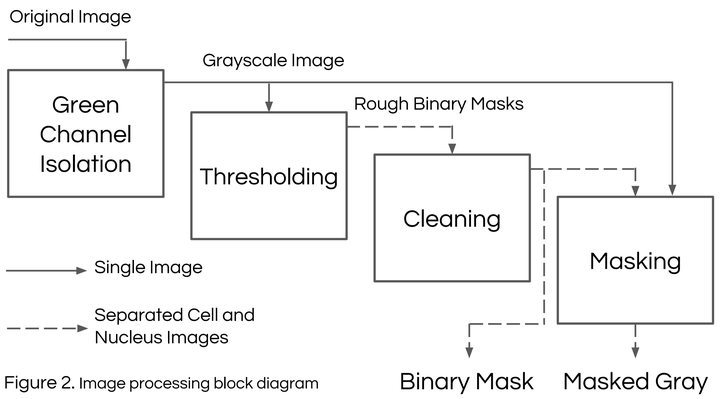

Figure 2. Image processing block diagram

Processing Images

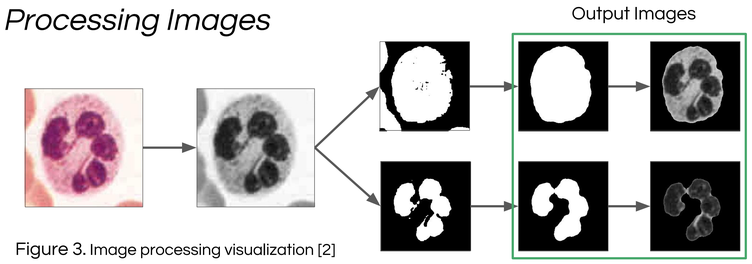

Figure 3. Image processing visualization [2]

Subtype Classification

Feature Extraction: 12 Total

Grayscale Features	Binary Features
Homogeneity (of cell and nucleus)	Compactness
Contrast (of cell and nucleus)	Ratio of area of nucleus to area of cell
Entropy (of cell and nucleus)	Ratio of area to perimeter (of cell and nucleus)
	Circularity (of cell and nucleus)

Support Vector Machine (SVM)

- Creates a boundary between points in the feature space to classify new test examples
- 1 Vs. 1 and 1 Vs. All SVM using LibSVM library
- Gaussian kernel
- Parameters C and γ chosen by cross-validation

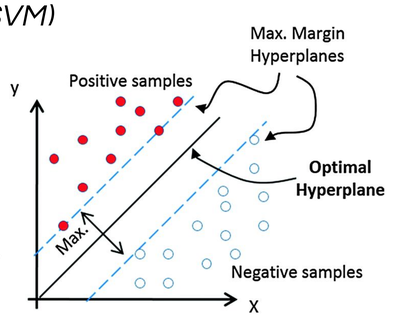

Figure 4. Example of SVM Classifier [3]

Neural Network

- MATLAB Neural Networks Pattern Recognition Toolbox
- Creates back-propagation neural network
- Automatic separation of training, validation and test images

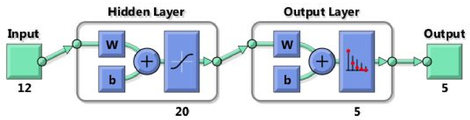

Figure 5. Implemented neural network: 12 inputs (features), 20 hidden layers, 5 classes

Results and Data Analysis

SVM

Figure 6A. Confusion matrix for 1 Vs. 1 classification on test data. 1 = basophil, 2 = eosinophil, 3 = lymphocyte, 4 = monocyte, 5 = neutrophil.

Figure 6B. Confusion matrix for 1 Vs. All classification on test data.

- 1 Vs. 1: **33.3% accuracy**, 1 Vs. All: **46.7% accuracy**
- Best accuracy for basophils and neutrophils; 0% for monocytes

Neural Network

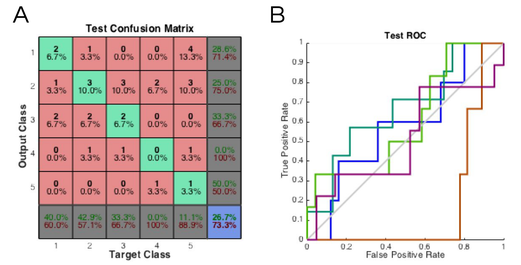

Figure 7A. Confusion matrix for the test data. Results vary from class to class but are overall low.

Figure 7B. The Receiver Operating Characteristic (*ROC*) *curve* is a plot of the true positive rate vs. the false positive rate with varying threshold. A perfect test would show points in the upper-left corner

- No significant correlation on ROC curve, **26.7% accuracy**
- Less accurate than SVM, similar distribution between classes

Discussion and Conclusion

- 1 Vs. All SVM had the highest accuracy (46.7%)
- Limitations
  - Small data set consisting of only bright field images
  - Low Image resolution and poor consistency of image contrast
  - Segmentation of overlapping cells
  - All monocytes misclassified
- Future
  - Obtain larger data set (>100 images for each class)
  - Explore more features using higher resolution images
  - Improve segmentation using color spectrum to further tune feature matrix

References

1. White Blood Cell Types. Retrieved December 9, 2015, from http://images.wisegeek.com/white-blood-cell-types.jpg
2. Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V., Beksac, M., & Beksac, S. (n.d.). Feature extraction and classification of blood cells for an automated differential blood count system. *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222).*
3. Fernandez-Lozano, C., Fernández-Blanco, E., Dave, K., Pedreira, N., Gestal, M., Dorado, J., & Munteanu, C. (n.d.). Improving enzyme regulatory protein classification by means of SVM-RFE feature selection. *Mol. BioSyst. Molecular BioSystems*, 1063-1063.