White Blood Cell Similarity - Machines vs Humans

Anna Palmqvist Sjövall
Lund University
Lund, Sweden
anna.sjovall.293@student.lu.se

Oscar Odestål
Lund University
Lund, Sweden
oscar.odestal.173@student.lu.se

Who should we trust when the humans and the machines have different answers? Is the computer better at finding patterns in medical data than human experts? These are interesting questions that arise when working with machine learning in the medical field. The answer? Build systems that provide human experts with knowledge gathered from analyzing hundreds of thousands of images. Using this knowledge together with the human's outstanding deduction capability will save lives.

White blood cells (WBCs) circulate the human body and act as our defense mechanism against diseases. An abnormal blood sample might indicate that the patient suffers from a disease such as leukemia or lymphoma. To correctly identify these diseases, the WBCs need to be classified consistently and correctly.

CellaVision delivers automated processes of the otherwise manual microscopy. The systems can pre-classify the image of a cell, i.e. give a suggestion of which class a WBC image belongs to. Also, CellaVision shows six example cells in each class, *static reference cells*, handpicked to be typical for that class by a technician. The technician can use both the pre-classification and its results graphically presented, as well as static reference cells before making the final classification themselves. These static reference cells do not account for the specific image or system at hand. So, the reference cells are not the most visually similar and can be from systems with different optics.

The problem is that the human experts that perform the final classification are not consistent. Giving a cell image to five experts for classification could result in five different answers. Thus classifications are not reliable enough. What if there existed a classification standard for the experts? We try to move in that direction by building a program that will reevaluate the current classes by promoting data exploration.

We can place blood cell images on a two-dimensional map that encourages further exploration of the cell classes. We have found a way to present medical technologists with cells that look very similar to help with classification. We call these cells *adaptive reference cells* which are shown in Figure 1. Here we see eight other images of cells that are the most visually similar from the data set. These eight images have all been classified as class 13 by other experts. It is thus reasonable that the image in the middle also belongs to class 13.

A better classification will lead to a more accurate and consistent diagnosis. Hopefully, the classification process will be faster which will increase the test capacity and lead to faster treatment. If the reevaluated classes are accepted by the medical field, it would be beneficial to use these classes as the standard. Using these classes which are developed using data exploration would result in a class division based entirely on cell similarity. We believe that such a division will make the classification more reliable.

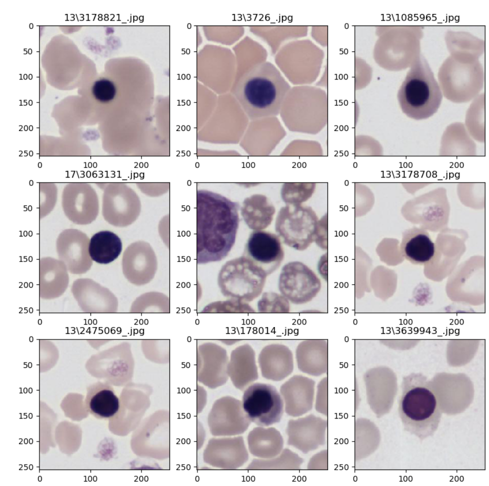

Fig. 1. The cells surrounding the middle image is its adaptive reference cells. The first number represent the cell class. The middle image is classified as 13.

This work could both be used to improve the current classification and encourage further investigation of the current classes.

We handed CellaVision's expert 200 images, known to be hard to classify, and asked to re-classify them. Surprisingly (or not), it turned out that the expert's classification overlapped by only 50% compared to the current classification. This is an interesting fact that further strengthens the need for our work.