Master of Science Thesis in Electrical Engineering
Department of Electrical Engineering, Linköping University, 2016

Automatic segmentation and classification of white blood cells from peripheral blood smear images taken from chickens

Erik H Kjellman

LINKÖPING UNIVERSITY

Master of Science Thesis in Electrical Engineering

Automatic segmentation and classification of white blood cells from peripheral blood smear images taken from chickens

Erik H Kjellman

LiTH-ISY-EX--YY/NNNN--SE

Supervisor: **Gustav Häger**
ISY, Linköpings universitet

Examiner: **Michael Felsberg**
ISY, Linköpings universitet

CVL
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Sammanfattning

Förhållandet mellan två olika sorters vita blodceller, nämligen heterofiler och lymfocyter, är ett användbart mått på stressnivån hos kycklingar. Detta förhållande räknas ut för hand idag, genom att manuellt räkna ett antal blodceller i blodutstryk. Det är ett väldigt tidskrävande arbete, och även känsligt för misstag.

Detta examensarbete undersöker möjligheten att automatiskt segmentera och klassificera de vita blodcellerna i blodutstryk från kycklingar för att räkna ut detta förhållande. Detta görs genom maskininlärning, genom att använda så kallade Convolutional Neural Networks, faltande neuronnät.

Detta arbete genomförs i samarbete med AVIAN Behavioural Genomics and Physiology group på Linköpings Universitet, som tillhandahåller blodutstryk från sina kycklingar, och expertis från sina människor.

Resultatet visar att processen att räkna ut förhållandet mellan heterofiler och lymfocyter kan göras semi-automatiskt eller helautomatiskt, beroende på kvaliteten på bilderna och åldern på individerna.

Abstract

The ratio between two different types of blood cells, i.e. heterophils and lymphocytes, is a useful measure to gauge the stress level of domestic chickens. This ratio is calculated by hand today, by manually counting blood cells in peripheral blood smear images. This is a very laborious and time consuming task, and prone to human error. This process should be possible to automate.

The aim of this thesis is to investigate automatic segmentation and classification of white blood cells in blood smear images taken from chickens in order to calculate the previously mentioned ratio. This is done through machine learning, by using Convolutional Neural Networks.

This thesis was produced at the AVIAN Behavioural Genomics and Physiology group at Linköping University, which provided blood smear image data from their chickens as well as expertise from their humans.

The results show that the process of calculating the ratio can be made semi-automatic or fully automatic, depending on the quality of the images and age of the individual chickens.

Acknowledgments

I want to thank the AVIAN Behavioural Genomics and Physiology group at Linköping University for giving the opportunity to make this thesis possible. Also, a big thanks to my supervisor and examiner at CVL, Gustav Häger and Michael Felsberg, respectively.

Finally, I would like to thank my friends and family for supporting me.

Linköping, November 2016
Erik H Kjellman

Contents

Notation

Glossary

Term	Meaning
CNN	Convolutional Neural Network – a type of neural network useful for image classification
CUDA	Nvidia's GPU toolkit, simplifying parallel computing
GPU	Graphical Processing Unit, the highly parallel processor on a graphics card
HL	Heterophil to lymphocyte ratio, i.e. the number of heterophils divided by the number of lymphocytes
PNG	Portable Network Graphics – a lossless image format
NDPI	The file format used for the blood smear images, essentially a proprietary extension of the TIFF format
ROI	Region Of Interest – a chosen region in an image used as input to the classification algorithm.

1 Introduction

1.1 Background

The ratio between Heterophils and Lymphocytes in chickens is a useful measure of their stress level [6]. As of now this ratio is calculated by hand, by using a digital slide scanner and counting the cells in the resulting image. This is a very laborious and time consuming task. This thesis analyses the possibility of automating this process, completely or partially, in conjunction with an interactive graphical user interface for manual correction.

The goal is that it will be a usable application for researchers in biology and similar fields, without degrees in engineering or other strictly technical fields.

1.2 Biological background

1.2.1 Red blood cells

The red blood cells in avian species have a cell nucleus, which mammalian red blood cells do not. This makes it harder to distinguish them from the white blood cells, which is one of the main reasons that cell recognition software for humans and other mammals cannot directly be applied to avian blood smear images.

1.2.2 White blood cells

White blood cells, or leukocytes, are cells of the immune system that are involved in protecting the body against infectious diseases, e.g. from viruses, bacteria or parasites. Both Lymphocytes and Heterophils are white blood cells.

Heterophils

Heterophils are a type of granulocyte that occurs in most avian species. Granulocytes are a category of leukocytes characterized by the presence of granules in their cytoplasm, that are more or less visible in the blood smear images. Heterophils in avian species are functionally equivalent to neutrophils in most mammal species.

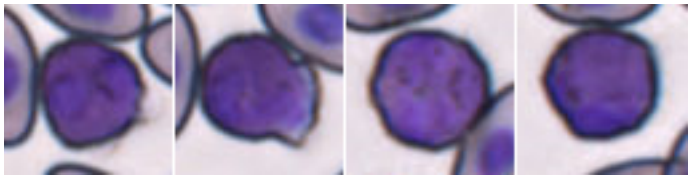


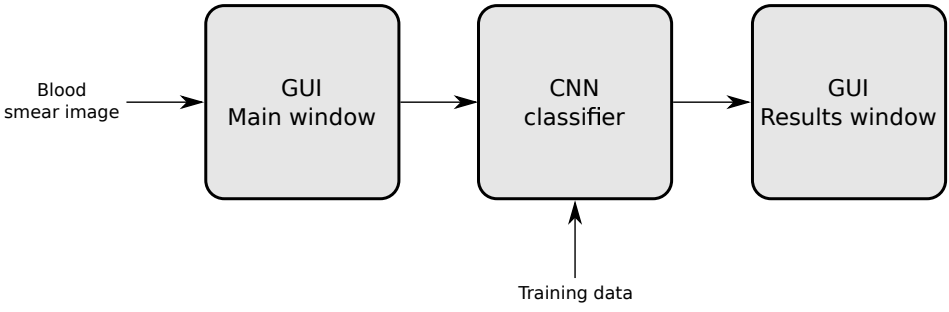
Figure 1.1: *Heterophils from a 9 week old chicken.*

Lymphocytes

Lymphocytes are smaller leukocytes with a large nucleus. They are sometimes confused with platelets, which are often smaller and darker than the lymphocytes. Platelets are also commonly referred to as thrombocytes and their function is to stop bleeding and clotting injured blood vessels [4].

Figure 1.2: *Lymphocytes from a 9 week old chicken.*

1.3 Problem description



Color images of blood cells from peripheral blood smears are taken with a digital microscope, and different types of blood cells are to be segmented and classified using a convolutional neural network (CNN). The most important blood cells are heterophils and lymphocytes, other white blood cells such as monocytes, eosinophils and basophils are not important for the task at hand, but it can be beneficial to detect these as well.

The images are given in the ndpi format, which is basically a proprietary extension of the TIFF file format, with different zoom levels of the blood smears. This application will only use the ones that are taken with the largest zoom available, in which a white blood cell occupies an area of approximately 50x50 pixels.

From the images a ground truth must be established. This is done by manually cropping out the individual cells and saving them as PNG images, with a number in the name corresponding to its class. Since the cells are seldom isolated in the image, it is expected that almost all individual cell images will contain parts of other cells around it. Since the CNN algorithm is much like a black box in terms of insight into the network from the user's perspective, thorough testing must be done with the trained network so as to verify the accuracy and flexibility.

1.4 Calculation of accuracy

The accuracy is calculated as precision, recall and F score. More on this in the Results chapter.

1.5 Limitations

There are several limitations to this thesis, the main ones will be mentioned here.

1.5.1 Data set

There are hundreds of images available, where each image is in the order of 10^{10} pixels. However, since labelling is time consuming only a few thousand white blood cells have been labelled in these images. In this thesis, around 10-20 of the available images have been used. The data set only contain blood smear images from chickens of the age of 9 and 12 weeks, which probably limits the robustness of the application, since the cells vary in size and shape during the chickens lifetimes.
The color and quality of the blood smear images can vary when scanning and/or staining them, so to keep the algorithm as unbiased as possible to specific images, the cut out learning images are taken in equal amounts from the selected part of the available images.

1.5.2 Loss of depth

When manually analysing a blood smear image in a microscope, it is possible to set the focus to different depths of a cell. This can make it easier to correctly classify the type of the cell, since some features are more apparent at other depths. Since the cells are essentially photographed at a fixed depth, any other information about the cell that could be found on another focus plane will be lost. However, since the modern method of counting the cells by hand most often use the same digital blood smear images as those that are used in this thesis, and seldom involves manually counting the cells through a microscope, the loss of depth is thus not limited to our method.

1.5.3 Image artefacts

Getting a perfect image from a blood smear image is virtually impossible. The concentration of blood cells in the image varies significantly, in some parts the cells may be clumped together in groups of several hundred cells, in other parts the cells are spread very far apart. There are also many damaged cells from the smearing process, and strands of hair and dust is quite common. When analysing the images manually, the most common method is to choose a few regions of interest (ROI) where the amount of artefacts is low, and the concentration of cells is at an acceptable level.

The approach used in this thesis has been to automatically cut a blood smear image into squares with a size of 2048 by 2048 pixels, saving them as TIFF images, sorting them by size in kilobytes, which crudely sorts them after cell density, since images with few cells will be more efficiently compressed. Finally, a few ROIs were chosen manually from somewhere in the middle of this list, where the cell density was not too high or too low, and the amount of artefacts are minimal.

1.5.4 Ambiguity in classes

The different types of blood cells can sometimes look very similar to each other, making it difficult even for trained humans to accurately classify these cells[1]. This owes to individual differences in the chickens. The size and color of the cells can vary greatly even in the same blood smear image. There are always a number of cells that got damaged in various degrees when producing the blood smear images, which makes the classification even harder. These cells are often not taken into account when counting them manually though.

This has been the hardest problem to solve in this thesis, and remains somewhat unsolved. The lymphocyte cells especially can be very similar to partially broken red blood cells, and a trade-off had to be done between either often missing some lymphocytes while not falsely classifying red blood cells as lymphocytes, or less often missing lymphocytes while falsely classifying red blood cells as lymphocytes. The resulting algorithm leans more towards falsely classifying damaged red blood cells as lymphocytes, as it is more important not to miss lymphocytes.

This has been achieved by removing some of the damaged red blood cells from the training images, so that the classifier is not as strict when it comes to classifying them.

1.5.5 Calculation of accuracy

The previously mentioned limitations and issues makes the calculation of the accuracy hard to determine. Therefore, some choices had to be made regarding issues in determining cell types. If the classifier fails to mark a heterophil or lymphocyte that could be seen as damaged, but is present in the ground truth image, it is not taken into account when determining the precision. If it marks a heterophil or lymphocyte that could be seen as damaged, but is not present in the ground truth image, it is not taken into account when determining the recall. Examples of these are given in the Results chapter.

1.5.6 Human blood smear comparison

The problem of counting leukocytes in human blood smears is largely solved[2][3], but for avian species the problem is significantly more challenging to solve, mainly because their red blood cells have nuclei, which the human counterpart does not. A comparison of human and avian blood smear images can be seen in Figure 1.3. It is readily apparent in this image that the human blood smear images are easier to automatically classify, since the blood cells in the human image have simpler shapes and textures. Figure 1.4 shows comparisons between lymphocytes and especially hard to classify red blood cells.

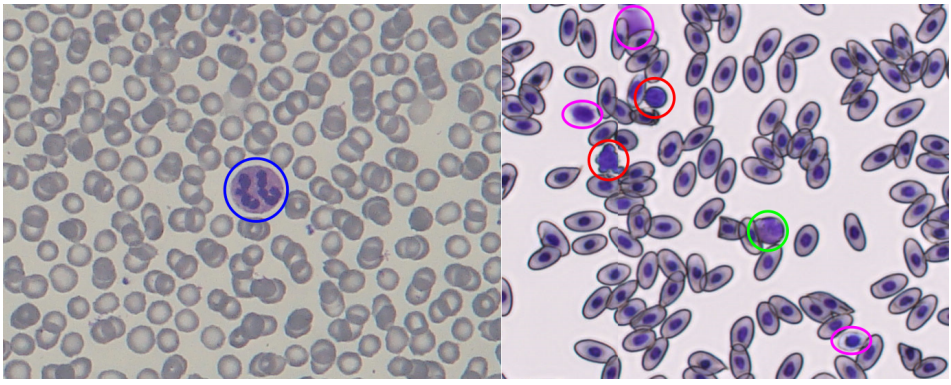


Figure 1.3: *Comparison of human (to the left) and avian (to the right) blood smears. The unmarked cells in both images are red blood cells. The blue in the left image is a neutrophil, the red, green and pink in the right are lymphocytes, heterophils, and various damaged red blood cells respectively.*

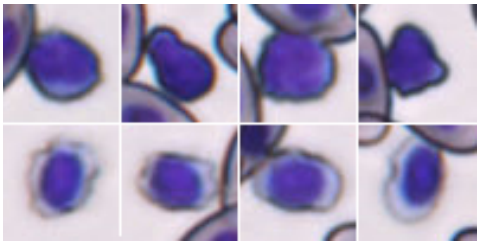


Figure 1.4: *The top four images show lymphocytes, and the bottom four undecided or broken red blood cells. These are hard even for an expert to correctly classify.*

2

Theory

2.1 Machine learning

Machine learning has a wide range of applications. It has proven especially useful for image classification and detection. While neural networks has a long history, applying them to detection problems is a more recent development. This is due to an increase in available computing power.

2.1.1 Neural Networks

2.1.2 Convolutional Neural Networks

In recent years, a kind of neural network has proven to be especially effective at image and video recognition, i.e. Convolutional Neural Networks (CNN). Instead of manually choosing which features to focus on when training a neural network, the CNN algorithm can take a whole image as input and dimensionally reduces this data in order to decrease the number of computations needed. This is done by convolving the image with trainable convolution kernels. The upside to this is that no potentially useful information must be removed when manually pre-processing the image. Convolutional neural networks are inspired by the organization of the animal visual cortex, and are variations of multilayered neural networks.

2.1.3 Max pooling

Max pooling is a method used to decrease computational cost and prevent overfitting a network. The aim is to downsample an input by applying a max filter to sub-regions of the original representation. Figure 2.1 gives a simple example of how this is done.

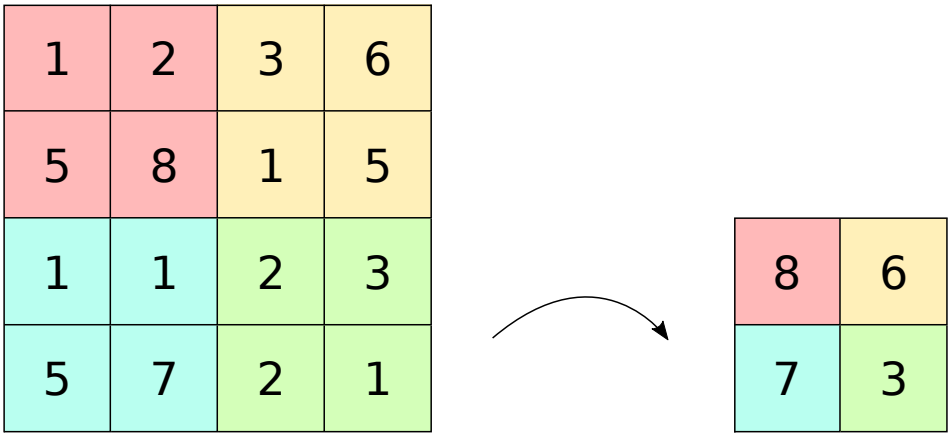


Figure 2.1: *Simple example showing how max pooling works, with a stride of 2.*

2.1.4 Dropout

Dropout is a technique that reduces over-fitting. At each training stage, or epoch, individual nodes are either removed, i.e. "dropped out" with a probability $1-p$ or kept with a probability p. This is essentially the same as using 2^n neural networks and combining their outputs, where n is the number of training epochs.

2.2 Problem outline

3

Method

3.1 Data sets

The data provided from the AVIAN Behavioural Genomics and Physiology group consists of around 100 blood smear images. Only around 10-20 of these available images have been used, owing to the fact that labeling these is a very time consuming task.

3.2 Detection and segmentation

3.2.1 Image preprocessing

The CNN network takes input as four-dimensional arrays of the size nx64x64x3, where the three last dimensions are the width, height and number of channels of the image, and the first denoting how many images there are. For example, an input of 100x64x64x3 means that there are 100 RGB images with the width and height of 64 pixels each. This means that the selected ROIs must be divided into 64 by 64 pixel RGB images in order to be classified by the CNN network. In order to give the best results however, these ROIs should be divided into overlapping squares, with a stride of 8 pixels in both the x and y directions, a value that was empirically arrived at.

3.2.2 Heatmap

The output from the CNN network for every image is a list of three values, denoting the probability of the image being either a heterophil, a lymphocyte or something other, e.g. a red blood cell or artefacts. Thus, the sum of these values

is always equal to one. A heatmap is then generated, with green signifying heterophils, and red lymphocytes. The CNN classifier can sometimes give weaker results on some parts of a cell, for instance if an image part only contains the inner parts of the cell without the cell wall. This problem is mitigated by dilating and blurring the heatmap before thresholding it. Finally, this post-processed image is then used as input to OpenCV's function findContours in conjunction with boundingRect, which returns the bounding box coordinates for the cells.
It is then a trivial task to crop these cells out of the ROI image to show in the Results window. Figure 3.1 shows an example of this process.
The heatmap for the

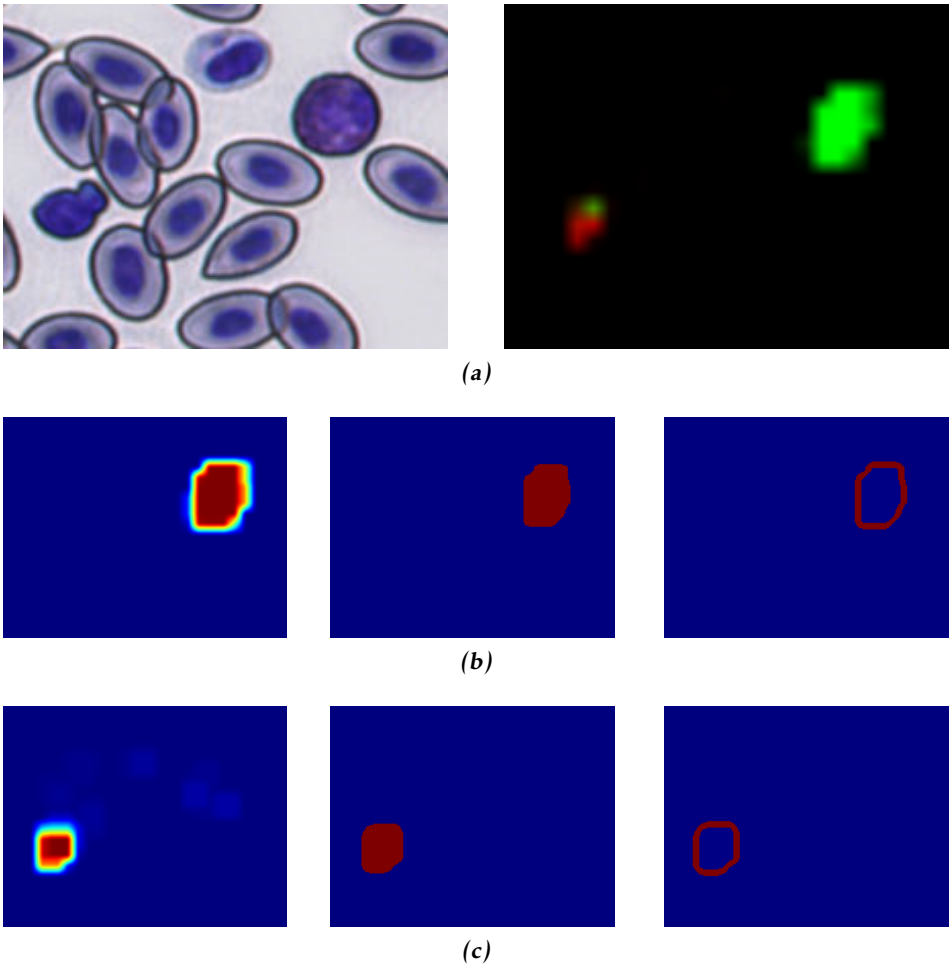


Figure 3.1: Figure (a) depicts the original ROI image to the left and its heatmap to the right. Figure (b) depicts the blurred and dilated heatmap for the heterophil cell to the left with thresholded counterpart in the middle, and its contour to the right. Figure (c) depicts the same as (b) but for the lymphocyte.

3.3 Data gathering

The GNU Image Manipulation Program, or GIMP, is used for extracting images of cells for training the neural network. However, the blood smear images are too large to handle, so before extracting these cell images the blood smear image is split into smaller files with a size of 2048 by 2048 pixels. Some of these images are chosen in accordance with the method described in 1.5.3. These smaller images are then opened in The GIMP, where a new layer is added, and squares of

different colors representing the different cell types are drawn on top of the cells. When all white blood cells have been covered with a square of the corresponding color, the layer with the squares is saved as a PNG image. A simple python script then reads both of these images, and uses the coordinates of the squares to cut out the cells from the cell image and saves them as a sequence of PNG images with the naming convention celltype_number.png, e.g. lymphocyte_1.png, lymphocyte_2.png, ..., lymphocyte_n.png.

3.4 Implementation

3.4.1 Neural network structure

The neural network consists of a number of layers in the following order:

1. An RGB input layer of 64 by 64 pixels
2. A convolutional layer with 32 filters, 5 by 5 pixels in size.
3. A max pooling layer of stride 2 in both x and y.
4. A convolutional layer with 32 filters, 5 by 5 pixels in size.
5. A max pooling layer of stride 2 in both x and y.
6. A convolutional layer with 32 filters, 5 by 5 pixels in size.
7. A max pooling layer of stride 2 in both x and y.
8. A fully connected layer of 256 units, with 50% dropout.
9. A fully connected layer of 3 units, with 50% dropout.

All layers except the last use the Rectified Linear Unit (ReLU) as an activation function, which has been shown to perform better than for example sigmoid activation functions CITATION HERE. The ReLU is defined as $\phi(x) = max(0, x)$. The last layer uses a softmax activation function, which gives a differentiable approximation of the non-differentiable ReLU function. It is defined as $\phi(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}}$

The structure of the neural network is shown in Figure 3.2.

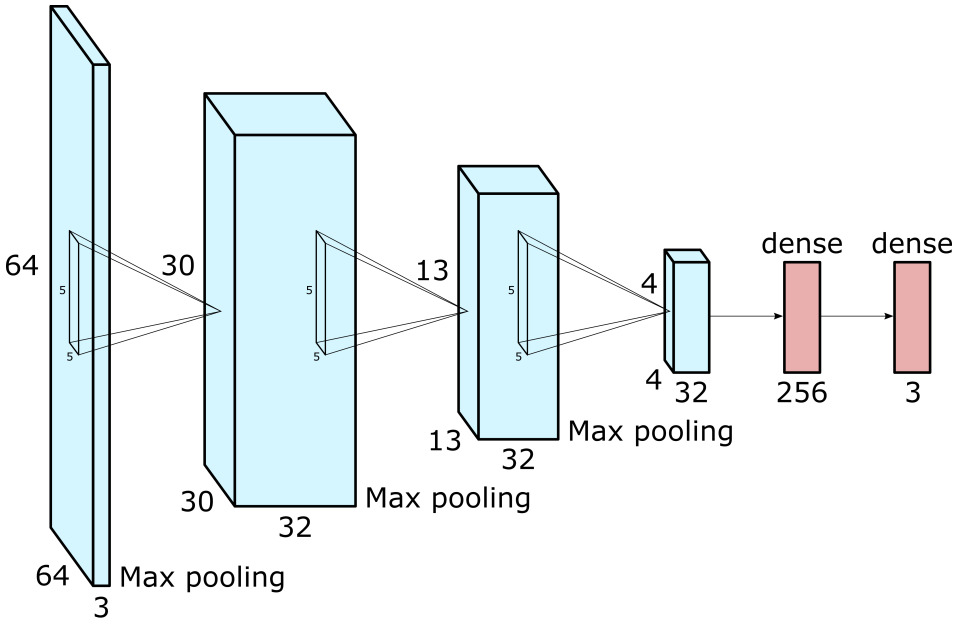


Figure 3.2: *Neural network structure.*

3.4.2 Framework

The implementation is done in Theano and Lasagne for the Convolutional Neural Network in Python, with additional libraries simplifying and greatly reducing computation time, most notably NumPy for efficient array computations.

3.4.3 Preprocessing

3.5 Graphical User Interface

Since this thesis aims to create a semi-automatic method for obtaining an HL ratio, an easy to use GUI must be included. This section details the design and implementation of this interface.

3.5.1 Design

The GUI consists of two main windows, where the first consists of an image viewer that supports zooming in and out of the blood smear image, and the drawing of ROIs. When the user is satisfied with the selected ROIs and clicks on the button labeled "Run", the main algorithm that finds the heterophils and lymphocytes is run, which may take several minutes up to hours, depending on the size of the selected ROIs. When the algorithm is done, the second main window pops

up, which shows a grid of nine cells, with left and right buttons to show the next and previous cells found. In this window, it is also possible to show only one type of cell, or choose between the ROIs created. In the upper left corner, the HL ratio for the currently chosen ROI(s) is shown. By right-clicking on the cells in the grid, it is possible to remove or move cells to another category, for example if a lymphocyte is wrongly classified as a heterophil. This feature can be seen in Figure 3.5.

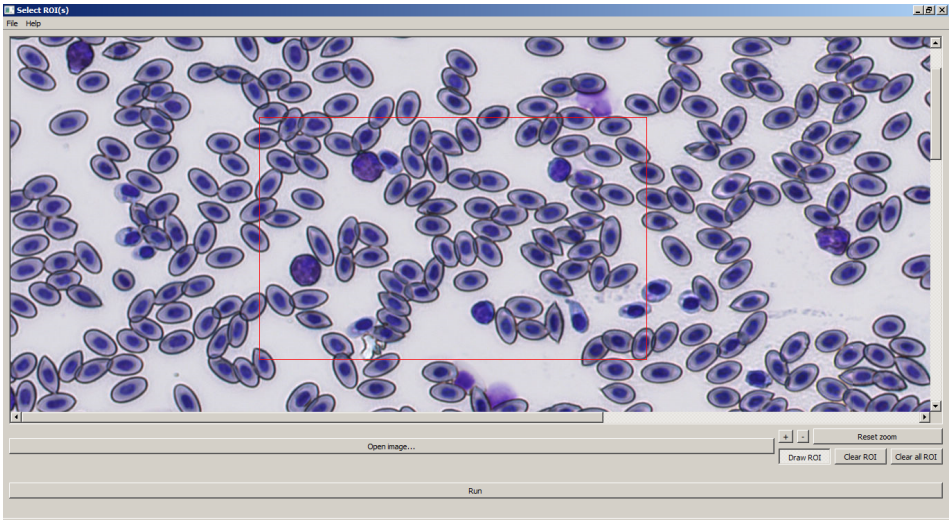


Figure 3.3: *Simple example showing how the main window of the GUI works, with a minimal ROI selected.*

3.5.2 Implementation

The interface was implemented in Qt Creator, which is a cross-platform integrated development environment (IDE) that includes an integrated GUI layout and forms designer. This greatly reduced the implementation time since GUI design can otherwise be a tedious and time consuming activity.

3.6 Evaluation

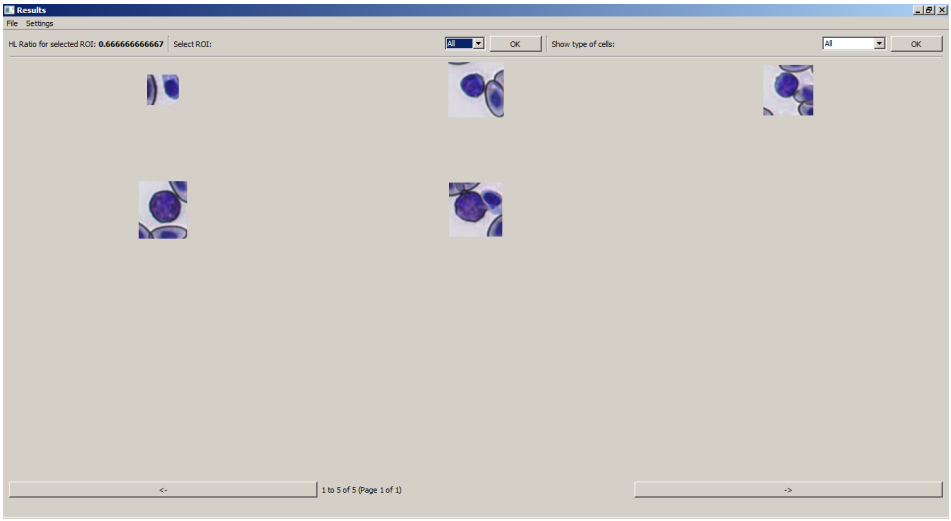


Figure 3.4: *Results from the previous ROI, where the first cell is a probable wrongly classified red blood cell.*

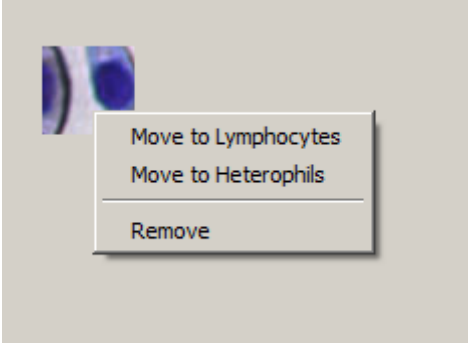


Figure 3.5: *Right-click menu showing the different options for removing or moving cells to another category.*

4

Results

4.1 Overview

The results show that it is beneficial to use this method as opposed to doing it manually. It is not perfect, but with human intervention a lot of time can be saved.

Age in weeks	Number of cells	Precision	Recall	F score
9	119	84.6%	83.0%	83.8%
12	121	79.6%	81.1%	80.3%

4.2 Calculation of accuracy

The accuracy is calculated as precision, recall and F score. Precision and recall are given as:

$$\text{Precision} = \frac{tp}{tp + fp}$$
$$\text{Recall} = \frac{tp}{tp + fn}$$

where fp and tp and fn stand for false positive, true positive and false negative respectively, and F score is the harmonic mean of these measures, i.e.:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

5

Conclusion

5.1 Conclusion

This thesis project has proven to be more challenging than previously thought, for several reasons. The main reason is that the lymphocyte cells are simply too easy to be mistaken for broken red blood cells and platelets. The heterophils however are much easier to distinguish from other cells, which is reflected in the results. Although this semi-automatic method can be used to significantly lessen the burden on the researcher(s) tasked to count these cells.

A better method of counting the white blood cells of chickens is to use a flow cytometer[5], although costly instruments prohibit this method from being widely used.

5.2 Future Work

There are many ways in which this method could be improved upon, the main ones are described in this section.

5.2.1 Data gathering

In order to generalize the algorithm further, much more data from chickens of different ages should be gathered. In this thesis, only chickens of the ages of 9 and 12 weeks were used.

Bibliography

[1] Ryan L. Bohls, Roger Smith, Pam J. Ferro, Nova J. Silvy, Zhongxia Li, and Ellen W. Collisson. The use of flow cytometry to discriminate avian lymphocytes from contaminating thrombocytes. *Developmental & Comparative Immunology*, 30(9):843 – 850, 2006. ISSN 0145-305X. doi: http://doi.org/10.1016/j.dci.2005.10.008. URL http://www.sciencedirect.com/science/article/pii/S0145305X05002132. Cited on page 4.

[2] M. Hamghalam and A. Ayatollahi. Automatic counting of leukocytes in giemsa-stained images of peripheral blood smear. In *2009 International Conference on Digital Image Processing*, pages 13–16, March 2009. doi: 10.1109/ICDIP.2009.9. Cited on page 5.

[3] PS Hiremath, Parashuram Bannigidad, and Sai Geeta. Automated identification and classification of white blood cells (leukocytes) in digital microscopic images. *IJCA special issue on "recent trends in image processing and pattern recognition" RTIPPR*, pages 59–63, 2010. Cited on page 5.

[4] K. Laki. Our ancient heritage in blood clotting and some of its consequences. *Annals of the New York Academy of Sciences*, 202(1):297–307, 1972. ISSN 1749-6632. doi: 10.1111/j.1749-6632.1972.tb16342.x. URL http://dx.doi.org/10.1111/j.1749-6632.1972.tb16342.x. Cited on page 2.

[5] Christian Seliger, Beatrice Schaerer, Marina Kohn, Helene Pendl, Steffen Weigend, Bernd Kaspers, and Sonja Härtle. A rapid high-precision flow cytometry based technique for total white blood cell counting in chickens. *Veterinary Immunology and Immunopathology*, 145(1–2):86 – 99, 2012. ISSN 0165-2427. doi: https://doi.org/10.1016/j.vetimm.2011.10.010. URL http://www.sciencedirect.com/science/article/pii/S016524271100417X. Cited on page 19.

[6] H. S. Siegel W. B. Gross. Evaluation of the heterophil/lymphocyte ratio as a measure of stress in chickens. *American Association of Avian Pathologists*, 27(4):972–979, 1983. Cited on page 1.