

Peripheral Blood Smear Analyses Using Deep Learning

Rabiah A. Al-Qudah

A Thesis
in
The Department
of
Computer Science & Software Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Computer Science and Software Engineering) at
Concordia University
Montréal, Québec, Canada

March 2022

© Rabiah A. Al-Qudah, 2022

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Rabiah A. Al-Qudah**

Entitled: **Peripheral Blood Smear Analyses Using Deep Learning**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science and Software Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Chun Wang

_____ External Examiner
Dr. Herbert Yang

_____ External to Program
Dr. Wei-Ping Zhu

_____ Examiner
Dr. Adam Krzyzak

_____ Examiner
Dr. Andrew Delong

_____ Thesis Supervisor
Dr. Ching Y. Suen

Approved by _____
Dr. Leila Kosseim, Graduate Program Director

June 20, 2022 _____
Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

Peripheral Blood Smear Analyses Using Deep Learning

Rabiah A. Al-Qudah, Ph.D.

Concordia University, 2022

Peripheral Blood Smear (PBS) analysis is a vital routine test carried out by hematologists to assess some aspects of humans' health status. PBS analysis is prone to human errors and utilizing computer-based analysis can greatly enhance this process in terms of accuracy and cost. Recent approaches in learning algorithms, such as deep learning, are data hungry, but due to the scarcity of labeled medical images, researchers had to find viable alternative solutions to increase the size of available datasets. Synthetic datasets provide a promising solution to data scarcity, however, the complexity of blood smears' natural structure adds an extra layer of challenge to its synthesizing process. In this thesis, we propose a methodology that utilizes Locality Sensitive Hashing (LSH) to create a novel balanced dataset of synthetic blood smears. This dataset, which was automatically annotated during the generation phase, covers 17 essential categories of blood cells. The dataset also got the approval of 5 experienced hematologists to meet the general standards of making thin blood smears.

Moreover, a platelet classifier and a WBC classifier were trained on the synthetic dataset. For classifying platelets, a hybrid approach of deep learning and image processing techniques is proposed. This approach improved the platelet classification accuracy and macro-average precision from 82.6% to 98.6% and 76.6% to 97.6% respectively. Moreover, for white blood cell classification, a novel scheme for training deep networks is proposed, namely, Enhanced Incremental Training, that automatically recognises and handles classes that confuse and negatively affect neural network predictions. To handle the confusable classes, we also propose a procedure called "training revert". Application of the proposed method has improved the classification accuracy and macro-average precision from 61.5% to 95% and 76.6% to 94.27% respectively.

In addition, the feasibility of using animal reticulocyte cells as a viable solution to compensate for the deficiency of human data is investigated. The integration of animal cells is

implemented by employing multiple deep classifiers that utilize transfer learning in different experimental setups in a procedure that mimics the protocol followed in experimental medical labs. Moreover, three measures are defined, namely, the pretraining boost, the dataset similarity boost, and the dataset size boost measures to compare the effectiveness of the utilized experimental setups. All the experiments of this work were conducted on a novel public human reticulocyte dataset and the best performing model achieved 98.9%, 98.9%, 98.6% average accuracy, average macro precision, and average macro F-score respectively.

Finally, this work provides a comprehensive framework for analysing two main blood smears that are still being conducted manually in labs. To automate the analysis process, a novel method for constructing synthetic whole-slide blood smear datasets is proposed. Moreover, to conduct the blood cell classification, which includes eighteen blood cell types and abnormalities, two novel techniques are proposed, namely: enhanced incremental training and animal to human cells transfer learning. The outcomes of this work were published in six reputable international conferences and journals such as the computers in biology and medicine and IEEE access journals.

Acknowledgments

Five years ago, I left my country and family to pursue my graduate studies in Canada. I still remember the bittersweet feelings I had; a mixture of excitement about starting a new chapter in my life, and the bitterness of having to say goodbye to my family and loved ones. Today, as I am ending this chapter, I find myself experiencing the same feelings; the joy of graduation and the bitterness of having to leave Dr. Suen's research lab. Being supervised by Dr. Suen was a unique and enriching experience, no words can express my gratitude to Dr. Suen for turning the challenging PhD journey into a joyful productive one.

To my beloved father, it breaks my heart that I am typing these lines while you are no longer with us, I was in a race with time since day one of my PhD, trying hard to graduate while you were still alive, but cancer was unfortunately faster than me. Your big smile through the pain and chemotherapy every time I passed a milestone in my PhD was my biggest motivation. To my mother, thank you for your continued support, prayers, and motivation.

To my husband, I cannot thank you enough for your patience and support. Finally, to my beautiful daughters, Yara and Hala, thank you for all the uplifting cards, scrapbooks, and stress handling techniques. Thank you for excusing me for being very busy all year long. As a mother, I really hope you learned from me that starting a family does not mean giving up on your dreams.

Contents

List of Figures	x
List of Tables	xii
1 Introduction	1
1.1 Motivation	1
1.2 Objectives and Contributions	2
1.3 Related Publications	4
1.3.1 Available Blood Smears and Cells Datasets	4
1.3.2 Malaria Detection	7
1.3.3 Blood Cell Detection and Classification	11
1.3.4 Leukemia Diagnosis	15
1.3.5 Reticulocyte Detection	17
1.4 Thesis Statement	21
1.5 Outline	22
2 Background	23
2.1 Blood Components	23
2.2 Phlebotomy and Blood Cell Separation	27
2.3 Disease Diagnosis from Blood	28
2.3.1 The Complete Blood Count Test	28

2.3.2	Peripheral Blood Smears	29
2.3.3	Clinical Laboratory Professionals Review and Diagnosis from CBC and PBS Data	31
2.3.4	Reticulocyte Smears	32
3	Methods	34
3.1	Study Scope	34
3.2	Generation of Synthetic Blood Smears Using Locality Sensitive Hashing . .	35
3.2.1	Stage 1: Data Pools Preparation	37
3.2.2	Stage 2: Blood Smears Generation and Annotation	37
3.2.3	Medical Assessment	45
3.3	General PBS Analysis Using Deep Learning	46
3.3.1	Phase 1: Classification of Main Blood Cell Types using YOLO Deep Network	47
3.3.2	Phase 2: Classification of Blood Cell Subtypes	48
3.3.3	Classifying platelets	49
3.3.4	Classifying WBCs	53
3.3.5	Considering Determinism	63
3.3.6	Evaluation	64
3.3.7	Experimenting Enhanced Incremental Training with Non-medical Datasets	64
3.4	Detection of Reticulocytes	66
3.4.1	Datasets	67
3.4.2	Can Animal Cells be a Possible Solution for Medical Data Scarcity?	70
3.4.3	Transfer Learning Between Similar Datasets: Feline to Human Reticulocyte Transfer Learning Experiments	72

3.4.4	Transfer Learning Between Dissimilar Datasets: ImageNet to Human Reticulocyte Transfer Learning Experiments	73
3.4.5	No Transfer Learning Experiments	74
3.4.6	Dataset Size and Balance Effect	74
3.4.7	Technical Setup	74
3.4.8	Evaluation Metrics	76
4	Results	78
4.1	Results of Synthetic Blood Smears Generation Using Locality Sensitive Hashing	78
4.1.1	Synthetic Dataset Generation Results	78
4.1.2	Medical Assessment Results	80
4.2	Results of Phase 1: Classification of Main Blood Cell Types using YOLO Deep Network	82
4.3	Phase 2: Results of Blood Cell Subtypes Classification	83
4.3.1	Platelets classification results	83
4.3.2	WBC classification Results	85
4.3.3	Comparison with existing works	94
4.3.4	Results of Experimenting Enhanced Incremental Training with Non-medical Datasets	95
4.4	Results of Reticulocytes Detection	97
4.4.1	Boost Measures Results	98
5	Limitations	102
6	Conclusion and Future Work	103
7	Publications	107

Bibliography	109
Appendix A WBC types	121
A.1 Neutrophils	121
A.2 Eosinophils	122
A.3 Basophil	122
A.4 Monocytes	122
A.5 Lymphocytes	123

List of Figures

1	Light microscopic images	4
2	Whole-slide images	4
3	RBCs abnormal morphologies [2]	24
4	Different blood samples used in different blood tests	30
5	Constructing blood smear approaches as black box. <i>A</i> is an enlarged annotated sample result from the approach described in 3.2.2, <i>B</i> and <i>C</i> are enlarged sample results from the approach described in 3.2.2, <i>D</i> is an enlarged sample result from the approach described 3.2.2.	36
6	General workflow of the proposed blood smear analysis framework	46
7	Activated platelets and normal platelets	50
8	Platelet classification approaches	50
9	The enhanced incremental training framework	54
10	Instances from separable and visually similar blood cell types, with microscope at magnification 300x	56
11	Samples from the humming bird dataset	65
12	Visual resemblance in the monkeys and canines categories	66
13	Human reticulocyte samples from the CENPARMI human reticulocyte dataset	68
14	Human RBC samples from the CENPARMI human reticulocyte dataset . .	68
15	Samples of the background objects class from the CENPARMI human reticulocyte dataset	68

16	Instance counts of the feline and human reticulocyte datasets	69
17	Different whole-slide instances where aggregate reticulocytes vary in size .	70
18	Different whole-slide instances where some cells are not annotated	70
19	The proposed experiments	72
20	Synthetic dataset main types statistics	79
21	Synthetic dataset Subtypes Statistics	79
22	Instances from real and synthetic blood smears	80
23	The results of the questionnaire	81
24	Sample results from the second phase of classification.	84
25	The workflow of training the four neural networks using enhanced incre- mental training and the accuracy results obtained from each stage using the validation set	87
26	The workflow of training using the proposed training scheme	87
27	The normalized confusion matrix of the top model trained using the one phase WBC classification method	91
28	The normalized confusion matrix of the top model trained using the pro- posed method	91
29	Min Max Average plots of the K-fold experiments using the CENPARMI dataset	99
30	Min Max Average plots of the K-fold experiments using the balanced ver- sion of CENPARMI dataset	99

List of Tables

1	Summary of the reviewed literature	18
2	Medical conditions and disorders the proposed system can infer	32
3	Types of blood cells	34
4	Statistics of the main blood cell types	39
5	The properties of the fully connected neural network utilized for platelet and WBC classification	52
6	WBC categories	58
7	Transfer learning formulation	71
8	The properties of the fully connected neural network utilized in the target networks	75
9	Blood cell classification results using YOLO	83
10	Confusion matrix resulted from the one-phase classification approach . . .	84
11	Confusion matrix resulted from the two-phase classification approach . . .	85
12	One phase WBC Classification Results	86
13	Classes included in each stage	88
14	Test results of the WBC classifiers trained using the proposed method . . .	89
15	Comparison of the results of WBC and platelet classification using our pro- posed system with existing works in the literature in terms of Accuracy . .	94
16	Results of 5-fold human reticulocyte classifier training	97
17	Pretraining and similarity boost measure results	98

18	The average loss and mean absolute deviation of the conducted experiments	100
19	Dataset size boost measure	101

Chapter 1

Introduction

1.1 Motivation

Blood tests are very important diagnostic tools. Despite the advancement in haematology, Peripheral Blood Smear (PBS) remains as a very important diagnostic test to both haematologists and clinicians; the literature reveals that approximately 70% of clinical decisions are supported by laboratory medicine [10]. Moreover, PBSs play a vital role in the diagnosis and monitoring of disease progression and therapeutic response. However, PBSs are still being manually analyzed by lab specialists using microscopes. Computer researchers have already developed many methods that automate some basic aspects of blood analysis [108, 19], whereas, only a handful of works conducted a deeper level of blood analyses to classify the morphological abnormalities and the blood cell subtypes. The automation of PBS examination will not only save hematologists time, money and reduce errors, but will also protect and save lives of front-line workers, especially during pandemics.

Deep Neural Networks (DNN) rely on large volumes of data. Obtaining large datasets can be a real challenge for researchers from different research areas. Acquiring medical

datasets can be even more challenging due to privacy constraints on patients' data. Moreover, annotating this type of data is a costly procedure that can only be performed by medical experts. In the context of blood smear analysis some extra domain-specific challenges arise. The first challenge is that some blood cell subtypes are rare in occurrence, for example, the authors in [124], were only able to collect three cases of reactive plasmacytosis in three years. Hence, having a sufficient number of blood smears containing such rare types for training and testing a deep network might take many years.

Additionally, the complexity of preparing a balanced dataset of blood smears comes from its natural structure; each blood smear contains hundreds of blood cells from different types, which are naturally distributed in an imbalanced manner. In [122], the authors demonstrated that Red Blood Cells (RBCs) occurred approximately seven times more than White Blood Cells (WBCs) in the training set. This challenge implies that traditional augmentation techniques might not help as it will only amplify the imbalance issue.

Moreover, most datasets employed in this research area are private which limits results reproducibility and comparability. Finally, the available public datasets are only annotated for some blood cell types which is not sufficient to fully comprehend and analyse blood smears.

This thesis contributes in overcoming two main challenges in the context of blood smear analyses, the automation of blood smears analyses and blood image data scarcity.

1.2 Objectives and Contributions

The main goal of this thesis is to automate the analyses of two types of blood smears, however, data scarcity is a major obstacle to the automation process. Hence, the problem of data scarcity is tackled by two solutions, synthetic data generation and Transfer Learning (TL). The main contributions of this thesis can be summarized as:

1. Provide a thorough literature review on the subject of PBS analysis, along with the main current challenges and possible future directions.
2. Employ Locality Sensitive Hashing (LSH) with Random Projections as a synthetic image generation method.
3. Create a dataset of synthetic whole blood smears. This dataset, which got the approval of five medical experts, is automatically annotated during the data generation process. To our knowledge, this is the first comprehensive synthetic dataset of this kind.
4. Propose a novel training method called "Enhanced Incremental Training", that trains the neural network in stages, while assessing confusable classes that negatively affect the network performance. Moreover, to handle the confusable classes, we propose a procedure called "training revert". A set of experiments are presented in this thesis to verify the effectiveness of the proposed method.
5. Propose a classifier that is trained on a synthetic dataset, and is capable of classifying the subtypes and morphological abnormalities annotated in the synthetic dataset. Classifying those classes aids in diagnosing more than 20 medical conditions and diseases. To our knowledge, this is the first classifier designed to work on a synthetic dataset rather than the usual datasets.
6. Train the deep classifier to categorize platelets into 3 main categories, which is rarely discussed in the computer science literature. Moreover, to our best knowledge this is the first work to consider the activated platelets in the classification.
7. Study the feasibility of utilizing animal cells as a viable solution to human medical data scarcity. To our knowledge, this is the first work that studies this aspect.
8. Propose a public labeled human reticulocyte dataset for research purposes.

9. Propose a human reticulocyte classifier, that classifies blood cells into 3 main classes; RBCs, reticulocytes and other random background objects.
10. Conduct a set of experiments to study the dataset size and balance trade-off.

1.3 Related Publications

In this section, available blood smears and cells datasets are listed. Moreover, the most recent studies on the topic of blood analyses are summarized.

1.3.1 Available Blood Smears and Cells Datasets

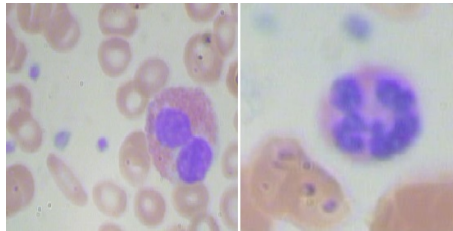


Figure 1: Light microscopic images

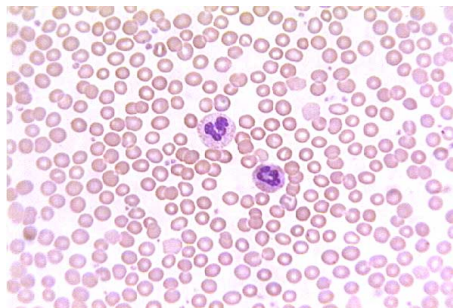


Figure 2: Whole-slide images

In the context of PBS analysis research, two types of datasets can be found: light microscopic image datasets, also known as, blood cell datasets, and whole-slide datasets, as shown in Figures 1 and 2, respectively. The first type is being extensively investigated

compared to the second one despite the fact that the whole-slide sets pose more realistic and challenging scenarios, as the entities in the blood sample appear microscopic, touching and crowded instead of the simple scenario represented in the light microscopic images, where only one object of interest appears. Some public and synthetic datasets are listed in the next two sections.

Synthetic Blood Smear Datasets

Synthetic datasets are widely used as a solution to data scarcity [42, 47]. Multiple techniques have been used to generate such datasets. The work in [23], generated blood smears by pasting blood cells on blood smear canvases by applying a Markov random process followed by a pix2pixHD network. A shortcoming of this synthetic dataset is that it only considers RBCs, which is a unrealistic scenario.

Non-synthetic Blood Smear and Blood Cell Datasets

Some public blood smear and blood cell datasets are:

1. Acute Lymphoblastic Leukemia Image Database for Image Processing (ALL-IDB) dataset [69]: this dataset is composed of 108 images collected in September 2005. It contains approximately 39000 entities with an image resolution of 2592x1944. The dataset contains records for both healthy and sick people and it only supports annotation for blast cells. ALL-IDB1 dataset consists of whole-slide images, whereas ALL-IDB2 consists of light microscopic images. This dataset has been mainly utilised for Leukemia detection.
2. Blood Cell Count and Detection (BCCD) dataset [115]: this dataset consists of 364 images. The dataset is annotated in Visual Object Challenge (VOC) format for RBCs, WBCs, and Platelets. This dataset can be utilised for classifying and counting blood cells.

3. Leukocyte Images for Segmentation and Classification (LISC) dataset [102]: this dataset consists of 250 light microscopic images of the five main WBC types: Neutrophil, Lymphocyte, Monocyte, Eosinophil, and Basophil. The dataset also contains 126 whole-slide blood smears but without annotation. This dataset can be utilised for automating the differential blood test.
4. National Institute of Health (NIH) dataset [95, 4]: this dataset was published in 2018. It consists of light microscopic cell images from thin blood smear slides in which images were collected at Chittagong Medical College Hospital, Bangladesh. The dataset contains 27,558 light microscopic cell images, with equal instances of Plasmodium parasitized and uninfected segmented red blood cell images. This dataset has been widely used in Malaria related research.

Some general conclusions can be drawn from this section are:

1. A promising solution to the data imbalance and scarcity issues in the context of PBSs can generate more synthetic data. However, Only a handful of works have provided synthetic data generation solutions in this context.
2. Most public datasets are composed of light microscopic images, or cropped blood smear images that only show small cropped portions of blood smears. Hence, more whole-slide blood smear datasets are needed as it poses more realistic and challenging scenarios.
3. The majority of the public datasets are only annotated for the main blood cell types (RBCs, WBCs, Platelets) or only the main WBC or RBC subtypes. On the other hand, datasets annotated for morphological abnormalities are scarce. This has resulted in limiting most of the work done in this context to only consider the main blood cell types and subtypes. If annotations are made to cover more blood cell subtypes and morphological abnormalities, then computer researchers will be able to

tackle more areas of blood analysis.

Computer researchers have mainly focused on four distinct directions in the context of blood smear analysis:

1. Malaria Detection.
2. Blood Cell Detection and Classification.
3. Leukemia Diagnosis.
4. Reticulocyte Detection

In sections 1.3.2, 1.3.3, and 1.3.4, the recent state-of-the-art techniques used for each of these directions are presented in greater detail.

1.3.2 Malaria Detection

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female *Anopheles* mosquitoes. The estimated number of malaria deaths stood at 435,000 in 2017. Many studies used the NIH dataset to train their neural networks [68, 96, 61, 93, 95, 118].

In [68], a Convolution Neural Network (CNN) architecture that is comprised of 12 layers was trained on the NIH dataset after applying a set of augmentation operations such as horizontal flip, vertical flip, width shift, height shift, fill mode, zoom range, and rotational range. The dataset was preprocessed by applying normalization, gamma correction, and logarithmic correction. This model achieved an accuracy of 98.23% and a F1 score of 97.74%.

The NIH dataset instances were mean normalized in [96] and several augmentation techniques including rotations, translation, shearing, zooming, and flipping were performed.

A CNN that consists of 3 blocks was trained for Malaria detection. Moreover Visual Geometry Group-19 (VGG-19), SqueezeNet, InceptionResNet-V2 were customized by truncating them at their deepest convolutional layer and adding a Global Average Pooling (GAP) and dense layers. Several combinations of the listed models were ensembled by taking the average of the predictions, the VGG-19 and SqueezeNet combination outperformed the individual models and other ensembles in all performance metrics with an accuracy of 99.51%.

The work in [61] utilized the same NIH set to train a CNN of 3 convolutional layers, one hidden layer, input, flatten and output layers. This shallow CNN achieved a good accuracy of 95% with no augmentation or preprocessing. In [94], the NIH dataset was preprocessed by stain normalization, Min-Max Normalization, and Standardization. Many augmentation techniques were applied: horizontal and vertical flips, Gaussian blur, rotation, horizontal and vertical shifting, darkening and lightening, ZCA whitening, and feature wise standardization, and Change of color space and Gaussian Blur. The dataset size was extended to 137,940 after augmentation. This work proposed a CNN that consists of 8 convolution layers. A VGG16 deep network was also customized by removing the pre-trained fully convolution layers, and adding a dense layer, a dropout layer, and a fully connected layer. A third architecture called CNNE_x-SVM was trained by emitting the customised VGG16 features to an Support Vector Machine (SVM). Finally, all models were ensembled by taking a weighted average of all predictions. The customised VGG16 achieved an accuracy of 97.6%, and the ensemble one achieved an accuracy of 97.7%.

The authors in [95] applied a multi-scale Laplacian of Gaussian (LoG) filter to detect RBC centroids. The detected cells were then segmented. Morphology opening operation was then applied to remove artifacts. F1-score of 0.952 was achieved in the detection phase. In addition, for cell classification a custom CNN of three convolutional layers and two fully

connected layers and other pre-trained CNNs (AlexNet, VGG-16, Xception, DenseNet-121, and ResNet-50) were employed. ResNet outperformed the other networks with 95.7% accuracy. The accuracy was improved to 95.9% by evaluating the optimal layer for feature extraction, as the final layer is not necessarily the optimal one.

Finally, in [118], an autoencoder was employed as a classifier. Multiple augmentation techniques were performed on the dataset, such as shift, zoom, and rotations. The performance of the autoencoder achieved an accuracy of 99.23%. Moreover, the model was tested to work on smartphones without the need for internet access. Some general conclusions can be drawn from [68], [96], [61], [94], [118], and [95] since it was all trained using the same dataset is that ensemble Learning when accompanied with preprocessing and augmentation can be a very effective and outperforms individual deep models.

Other works considered different datasets for training deep networks for Malaria detection [142, 93, 21, 116, 76, 83, 27]. For example, the authors in [142] developed an Android smartphone application using a dataset of 1819 whole slide thick smear images from 150 patients. The parasite in this work was detected using a pipeline that starts by applying an intensity-based Iterative Global Minimum Screening (IGMS) procedure to reduce the size of the initial search space and limit the number of regions of interest which are fed to a CNN consisting of seven convolutional layers. The classification accuracy was 93.46%.

The work in [93] annotated over 92k objects of the four major malaria species. Two CNNs were proposed for classification in this work, one for the cell quantitation and a second one for species identification. The first CNN had 3 convolutional layers, followed by two Inception modules and one fully connected layer. Where the late stage branch CNN was a fully convolutional neural network with 7 convolutional layers.

The authors of [21] combined two datasets, one collected by the authors and another public dataset from the Institute for Molecular Medicine Finland (FIMM). The overall collection has a total of 1030 images of infected cells, and 1520 images of non-infected ones.

The VGG network proposed in [21] was customized by removing the last three layers. The features obtained from the network were then fed to an SVM, and the model was called VGG19-SVM. Accuracy of 93.13% and F-score 91.66% were achieved.

In [116], a dataset of 1000 instances, multi-wavelength was utilized to increase the sample size, 45°, and 135° rotations were also applied. The authors opted to utilize AlexNet, VGG-16, ResNet50, GoogLeNet, and a customized CNN network of 5 convolutional layers and 2 fully connected layers. ResNet outperformed with an accuracy of 97.6% in classifying the test set as healthy or infected.

A CNN framework that was able to perform the extended depth of field images from z-stacks of thick blood films for automated malaria diagnosis was presented in [76]. Two deep architectures were proposed, EDoF-CNN-3D, and EDoF-CNN-Max. In EDoF-CNN-3D, the encoder part of the network was modified by replacing the two-dimensional convolutions with three-dimensional ones. The output tensor was flattened on the z-axis before the residual layers of the network by average pooling. On the other hand, the EDoF-CNN-Max combined the idea of the Siamese networks and wavelet-based networks. Each focal plane was passed through the encoder part of the network, and the maximum of the activation values were selected before going through the residual layers. The detection recall of the EDoF-CNN-3D method was 73%.

The work in [83] used another whole-slide dataset which consists of 800 infected cells and 2000 healthy cells. Two augmentation techniques were applied on half of the dataset to produce two augmented sub-datasets: image interpolation in the spatial domain, and image interpolation in the feature domain.

The authors in [27] used 4100 whole-slide peripheral blood smear images to train a Deep Belief Network (DBN) to classify the objects to either parasites or non-parasites. The objects were extracted from peripheral blood smear images using the level set method. A concatenated feature of color (histogram-based features and color coherence vector) and

texture (Haralick features, LBP features, and gray level run length matrix feature) were used to initialize the visible layer of the 4 hidden layer DBN. The deep network achieved a F-score of 89.66%.

The work in [85], trained the proposed framework on a blood cell images dataset of size 1182. The proposed framework utilised a functional link artificial neural network (FLANN) and a sparse stacked autoencoder. The proposed model scored an accuracy of 89.10%.

Finally, in [15], a dataset from Kaggle which comprises of instances that are classified as either, parasitized or uninfected is utilized to train a 16-layer CNN. The f1 score of the proposed CNN was 96%

1.3.3 Blood Cell Detection and Classification

The work in [90] proposed an architecture for microcytic hypochromia. The target features were a combination of blood smear image features extracted by AlexNet deep convolutional neural network and clinical features from (RBC count, Haemoglobin concentration (HB), RBC distribution width (RDW)). Samples were collected especially for this research from twenty patients. Both Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA) algorithms were used to reduce the feature set with minimal loss of information. k-Nearest Neighbors (k-NN), SVM, and Neural Network (NN), were employed for the classification phase. Each network was trained with three different feature sets: the clinical features, image features and fused features. The NN and the SVM classifier scored 99% accuracy at testing when trained with the fused features, which shows the superiority of the proposed fusion model.

The works in ([134], [91]) classify WBCs not only to their main types but also to some morphological abnormalities. In [134], a total of 14,700 annotated whole-slide images that include 11 categories of leukocytes were considered. Cell recognition was performed using

Single Shot Detector (SSD) and YOLOv3, Different variations of SSD and YOLOv3 were examined, a 0.931 Mean Average Precision (MAP) and an accuracy of 90.09% score were reported for the SSD 300×300 , and the highest MAP scored with YOLOv3 320×320 was 0.92.

In [91], the authors collected a private dataset that contains a total of 92480 leukocytes belonging to 40 categories, with one object of interest in each instance. To handle the dataset imbalance, many augmentation techniques were applied such as, horizontal and vertical flips, and adding random noises and color changes to the original images. The architecture of the proposed deep residual neural network consists of 7 convolutional layers, 2 fully connected layers and three residual blocks to improve its performance. The authors examined 7 different schemes by using different activation functions to train the network. The average classification accuracy was 76.84%.

The work in [51] studied blood-cell classification in medical hyperspectral imaging (MHSI). It utilized four different architectures: SVM, VGG16, CNN without Gabor wavelet, CNN with Gabor wavelet and a combination of modulated Gabor wavelet and CNN kernels, named as MGCNN. In MGCNN, each convolutional layer performs a dot product between multi-scale and orientation Gabor operators and the initial CNN kernels, to transform the convolutional kernels into the frequency domain in order to extract the features. Three datasets were utilized for testing; Bloodcells1-3, Bloodcells2-2, and a white blood cells dataset. The highest Overall Accuracy (OA) was achieved using the proposed model with a score of 94.03%, 94.40% and 97.65% on the first, second and third datasets respectively.

In [122], a subset of the All-IDB1 dataset was used, as the authors selected 42 images and performed a pixel wise annotation on them. The training set size was increased from 29 to 145 images by performing random reflection and translation augmentation techniques. The class weighting technique was used to handle the dataset imbalance caused by having RBCs appear seven times more than WBCs. The set was then fed to a SegNet for semantic

segmentation purposes. The highest class accuracy was the WBC's accuracy with a score of 94%.

In [133], a dataset of WBC images available on GitHub was used. Augmentation techniques such as random rotation, scaling, reflection, and shearing were performed. Gaussian noise was also applied to a subset of the training and testing set to train the network on poor quality images. The resulting 12,500 instances set was then fed to 3 sets of experiments that are made up of 10, 20, 30 CNNs where each CNN is constructed by generating random numbers of convolution blocks and layer sizes from preset ranges. The feature maps of each experiment were then concatenated and emitted to a PatternNet deep network to ensemble the strongest features to contribute in the final decision. The 30-CNNs experiment outperformed the 10 and 20 CNNs experiments with an accuracy score of 99.37%.

The input images were acquired from Pinterest online open source haematology database in [14]. RBCs were cropped from the blood smear to generate the dataset of normal, acanthocyte, sickle cell, teardrop and elliptocyte cell. The authors utilized SVM and AlexNet deep network. SVM model outperformed the AlexNet model, the authors referred this to the small dataset size. It is noted that there is a noticeable difference between the results of the models, for example, the SVM model achieved 100% accuracy in classifying Achantocyte where the deep learning model achieved 0%.

The work in [136] combines Fourier Ptychographic Microscopy (FPM) and an improved version of YOLO networks for WCB detection. In order to improve the detection of the microscopic WBCs, the feature maps of the last three layers were concatenated and passed to a final convolution layer. The proposed model was trained and tested on a 1000 whole slide image set.

The authors in [72] try to address the problem caused by the lack of some deep networks ability to fully exploit the long-term dependence relationship between certain key

features of images and image labels. A combination of a CNN and Recurrent Neural Network (RNN) was employed to deepen the understanding of the image context. A dataset comprising 12,444 augmented and rotated images of blood cells were collected from Kaggle and BCCD public datasets. The proposed network consists of Pre-trained convolutional neural network layer, RNN layer, Merge layer, and fully connected layer with Softmax output. The proposed model achieved an accuracy of 90.79%.

In [79], a dataset of blood cell images was augmented by performing rotation, reflection and translation, three pipelines one for each blood cell were implemented to perform the classification using CNNs and UNet deep networks. The 64000 blood cell dataset used in [98] is a combination of all-idb, DPDx, ASH image bank and other images available on Google.

The authors of [98] proposed a two-stage solution, In the first phase, a contour aware CNN was used for the segmentation of individual cells. In order to classify WBCs into five subtypes, features were extracted by a CNN and forwarded to ELM for classification. Several features were extracted such as centroid, medial axis ratio, and cell deform ratio, next, the extracted features were forwarded to Extreme Learning Machine (ELM) for classification. Overall RBC classification accuracy was 90.10%, the highest WBC subtype accuracy was 98.68%, which was scored on the Monocyte class .

The authors in [26] trained a capsule neural network on the LISC dataset to classify blood cells into the five main WBCs subtypes. In [137], not only a classification system was proposed, but also a different cell augmenting method was presented. The augmentation method was implemented by segmenting and pasting blood cells on different microscopic surrounding images. The model was able to achieve an accuracy score of 97.6%.

In [112], a modified version of the YOLOv3 network was trained on the BCCD dataset. The modifications on the YOLO structure included, using depthwise separable convolution,

and utilizing the Swish activation function instead of the Leaky ReLU function. The precision of the platelets, red blood cells, and white blood cells classes were 90.25%, 80.41%, and 98.92% and the MAP was 89.8%.

Finally, the authors in [84], utilized the BCCD and Kaggle datasets. The canonical correlation analysis approach was applied on the training dataset of blood cell images before training. A CNN was combined with a LSTM to classify the images into: Eosinophil, Lymphocyte, Monocyte, Neutrophil. The proposed method achieved a F1 score of 96.2%.

1.3.4 Leukemia Diagnosis

Leukemia is a fatal malignancy and has two main types: acute and chronic. Moreover, there are two subtypes of each leukemia main type depending on the size and shape of the WBC: lymphoid and myeloid. Acute leukemia is usually diagnosed after having clinical signs and symptoms that need to be confirmed by laboratory investigations. Complete blood count for WBCs, RBCs, platelets, and a peripheral blood smear are the initial tests. In many cases they are not enough to confirm the diagnosis, which is why the clinical practice is to do a bone marrow smear and biopsy. A bone marrow specimen will have a smear and a biopsy. This specimen is usually good enough to confirm the diagnosis of leukemia, but more testing is mandatory for subtyping the leukemia into lymphoid or myeloid, and then subclassifying each subtype. In bone marrow specimen, number and shape of WBCs are key points in diagnosing leukemia.

The works in [6] and [110] perform Leukemia diagnosis and classify the result into its subtypes. In [6], two public leukemia datasets, ALL-IDB and ASH Image Bank, were used to train the network to classify the samples into one of the four main Leukemia types. The number of samples increased to 8 times for both datasets by applying shifting, rotation, and flipping. A CNN of 2 convolution layers, a Flatten layer, followed by a fully connected layer was proposed in this work. The accuracy of classification obtained was 81.74%.

The works in [89, 75, 129, 123], and [130] aim to train a well generalized model to detect Leukemia. In [89] segmented white blood cell images of the C-NMC dataset were augmented by performing horizontal and vertical flips, and random translations. A Squeeze-and-Excitation-ResNeXt50 network achieved a weighted F1-score of 88.91%. In [75] only one object of interest appeared in each instance. All images were converted to grayscale and the cell region was then binarized using the threshold estimated by Otsu's method followed by the erosion operation. The authors opted to train the model using a ResNet with two fully connected layers and utilize the bagging ensemble training strategy. The model achieved an F1-score of 0.84.

The work in [129] trained three deep architectures, AlexNet, CaffeNet, Vgg-f, to generate features. The feature space was then reduced by applying the gain ratio algorithm, before emitting the features to an SVM classifier. An accuracy score of 100% was achieved when the classification was performed by concatenating and reducing the features obtained by all models. In [123], a CNN architecture comprising of 5 convolutional layers and 2 layers (fully connected and softmax) was trained on ALL-IDB1 dataset after applying many augmentation operations: histogram equalization, translation, reflection, rotation, shearing, conversion to grayscale, and blurring. The proposed method achieved an accuracy of 96.6%.

The authors in [130], extracted feature maps from the All-IDB1 dataset using AlexNet, CaffeNet, Vgg-f deep networks before being classified using SVM, Multilayer Perceptron (MLP) and Random Forest (RF). The authors also experimented the concatenation of feature maps obtained by all deep networks, the feature space was reduced by utilizing the PCA technique and the majority voting rule to combine the outcomes obtained by each classifier. An accuracy score of 100% was achieved.

The authors in [120] utilised the ALL-IDB2 and the C-NMC datasets to train and test a novel framework for efficient feature selection. The proposed approach combined a VGG

network with an improved version of the salp swarm algorithm (SSA). The VGG network was used as a feature extractor, where the SSA algorithm was employed for feature selection.

Three datasets, namely, ALL-IDB1, SMC-IDB, and IUMS-IDB were used for Leukemia detection in [36]. Each image instance was preprocessed by first applying an RGB to HSV conversion, followed by thresholding and Boolean mask generation. Next, a blob detection step based on the scale-normalized LoG (Laplacian of Gaussian) was executed. The resulting image was then segmented by applying multiple image processing filters and techniques. Next, an AlexNet network was employed for feature extraction. Finally the blood cells were classified using SVMs. The proposed framework scored an accuracy of 94.1%. A drawback of this framework, is the very long pipeline.

Finally, in [106], the authors utilized both ALL-IDB and LISC datasets. First, a Generative Adversarial Network (GAN) was utilized to increase the training instances. Next, a Darknet-53 and ShuffleNet were trained on the dataset. The classification results were 100% for the ALL-IDB and 99.70% for the LISC dataset.

1.3.5 Reticulocyte Detection

Xu and collaborators [140] proposed a pipeline to classify 8 types of RBCs; discocytes, echinocytes, elongated, granular, oval, reticulocytes, sickle and stomatocyte. First, RBCs were extracted in order to isolate the regions of interest (ROI), next, touching RBCs were separated by applying an improved random walk method based on automatic seed generation. In the second step, a mask-based RBC patch-size normalization method was utilized to normalize the variant size of segmented cell patches into uniform size. Third, a CNN was employed to perform the RBC classification. In addition to classify RBCs the authors performed shape factor analysis for each RBC type to further quantify specific RBC shape parameters. The proposed method scored an average AUC value of 0.94.

Vinicki and collaborators [67] proposed a public feline reticulocyte dataset. Moreover, an SSD network was utilized for the purpose of classifying the cells into RBCs, aggregate reticulocytes, and punctate reticulocytes. The proposed model achieved an accuracy of 98.7% in classifying aggregate reticulocytes. The authors also tested the model on a dataset created with a smartphone camera, the model was only able to classify 88.5% of the total reticulocytes correctly.

Finally, in [132], a new human reticulocyte dataset of 784 images was proposed. A Faster R-CNN was utilized and the test recall was 99.9% and 97.7% for RBCs and reticulocytes respectively.

Table 1 summarizes all the literature findings mentioned in this section.

Table 1: Summary of the reviewed literature

Malaria Detection			
Reference	Method	Dataset	Results
[15], (2021)	CNN	Kaggle	F1 score: 96%
[118], (2020)	Autoencoder	NIH	Accuracy: 99.23%
[85], (2020)	FLANN and sparse stacked autoencoder	Private dataset	Accuracy: 89.10%
[68], (2019)	CNN	NIH	Accuracy: 98.23% F1 score: 97.74%
[96], (2019)	CNN, VGG19, SqueezeNet, InceptionResNet-V2	NIH	Accuracy: 99.51%
[61], (2019)	CNN	NIH	Accuracy: 95%
[94], (2019)	CNN, VGG16, SVM	NIH	Accuracy: 97.7%

[95], (2018)	CNN, AlexNet, VGG-16, Xception, DenseNet-121, ResNet-50	NIH	Accuracy: 95.9%
[142], (2019)	CNN	Private dataset	Accuracy: 93.46%
[93], (2019)	CNN	Private dataset	Class accuracy: Ring: 94.8% Late: 96.6%
[21], (2019)	VGG, SVM	Private dataset + FIMM	Accuracy: 93.13% F-score 91.66%
[116], (2019)	AlexNet, VGG-16, ResNet50, GoogLeNet, CNN	Private dataset	Accuracy: 97.6%
[76], (2019)	CNN	Private dataset	Recall: 73%
[83], (2018)	CNN	Private dataset	Accuracy: 99%
[27], (2017)	Deep Belief Network (DBN)	Private dataset	F-score: 89.66%
Blood Cell Detection and Classification			
Reference	Method	Dataset	Results
[112], (2021)	YOLOv3	BCCD dataset	MAP 89.8%
[84], (2021)	CNN, LSTM	BCCD, Kaggle	MAP 96.2%
[26], (2020)	Caspule neural network	LISC dataset	Accuracy: 96.86%
[90], (2019)	AlexNet, SVM	Private dataset	Accuracy: 99%
[134], (2019)	YOLO, SSD	Private dataset	Accuracy: 90.09%

[137], (2019)	CNN	Private dataset	Accuracy: 97.6%
[91], (2018)	Residual Network	Private dataset	Accuracy: 76.84%
[51], (2019)	SVM, VGG16, CNN, CNN with Gabor wavelet	Private dataset	Accuracy: 97.65%
[122], (2017)	SegNet	ALL-IDB	Accuracy: 89.45%
[133], (2018)	CNN, PatterNet	Private dataset	Accuracy: 99.37%
[14], (2018)	SVM, AlexNet	Pinterest haematology dataset	Acanthocyte Accuracy: 100%
[136], (2018)	YOLO	Private dataset	Precision: 100%
[72], (2018)	CNN, RNN	Kaggle and BCCD	Accuracy: 90.79%
[79], (2018)	CNN, U-Net	Private dataset	Specificity: 99.11% Sensitivity: 100%
[98], (2017)	contour aware CNN, ELM	ALL-IDB, DPDx, ASH,Google.	Class accuracy: RBC: 94.71% WBC: 98.68%
Leukemia Detection			
Reference	Method	Dataset	Results
[106], (2021)	Darknet-53, ShuffleNet	ALLIDB, LISC	Accuracy: 100%
[89], (2019)	Squeeze-and- Excitation-ResNeXt50	Augmented, C-NMC dataset	F1-score: 88.91%
[36], (2020)	AlexNet	ALLIDB, SMC-IDB, IUMS-IDB	Accuracy: 94.1%

[89], (2019)	Squeeze-and-Excitation-ResNeXt50	Augmented, C-NMC dataset	F1-score: 88.91%
[6], (2019)	CNN	ALL-IDB and ASH Image Bank	Accuracy: 81.74%
[89], (2019)	Squeeze-and-Excitation-ResNeXt50	Augmented, C-NMC dataset	F1-score: 88.91%
[75], (2019)	ResNet	Private dataset	F1-score: 0.84
[129], (2018)	AlexNet, CaffeNet, VGG-f, SVM	Private dataset	Accuracy: 100 %
[123], (2018)	CNN	ALL-IDB1	Accuracy: 96.6%
[130], (2017)	AlexNet, CaffeNet, VGG-f, SVM	All-IDB 1	Accuracy: 100%
Reticulocyte Detection			
Reference	Method	Dataset	Results
[132], (2021)	R-CNN	Public Dataset	Recall: 97.7%
[67], (2021)	SSD	Public Feline Dataset	Class accuracy: 98.7%
[140], (2017)	CNN	Private Dataset	AUC: 94%

1.4 Thesis Statement

Blood smear analysis is an important diagnostic tool for medical doctors. Recent advancements in machine and deep learning have paved the way for researchers to utilize these learning networks for blood smear analysis. On the other hand, such learning algorithms are data hungry, and data-driven. Hence, having a balanced, expressive dataset is a key

factor in having reliable, well generalized models. Hence, the problem of medical data scarcity has heavily affected the advancement in automated blood analyses. In this thesis, the problem of blood data scarcity is investigated, moreover, two solutions are provided. In addition, an automated blood smear analysis system using deep learning is proposed.

1.5 Outline

The rest of the thesis is organized as follows: the most recent literature on the thesis topic is presented in Chapter 2. The proposed methods are outlined in Chapter 3. The results are described in Chapter 4, the discussion on the results is given in Chapter 5. Finally, Chapter 6 presents the conclusion and future work.

Chapter 2

Background

The human body is made up of 11 organ systems that collaborate to manage all the essential body functions. The cardiovascular organ system consists of 3 main parts; a closed circuit of vessels, the heart and blood [117]. Blood is the fluid that constantly circulates through the vessels. Blood is used to transport oxygen and nutrients to the body tissues and carbon dioxide to the lungs. Moreover, it carries waste products to the kidneys and liver [41].

2.1 Blood Components

Blood consists mainly of four components: RBCs, WBCs, plasma and platelets [103]. In the following subsections each of these components will be explored in detail.

Red Blood Cells (RBCs)

RBCs, also referred to as erythrocytes, are biconcave flexible disks containing haemoglobin which is responsible for its redness. They are anucleate entities with a diameter of 6 to 8 micrometers [103], and are responsible for carrying oxygen to the entire body and bringing carbon dioxide back from the entire body to the lungs to be released through breathing.

RBCs are produced from pluripotent haemopoietic cells [78]. These stem cells reside

initially in the bone marrow before they undergo a series of maturation events which eventually result in fully functional RBCs. Epo [50] is the primary hormone that manages the process of producing erythrocytes. The Pronormoblast is the first identifiable form of RBCs followed by the Basophilic normoblast, Basophilic Normoblast, Polychromatic Normoblast, Orthochromic Normoblast and reticulocyte in the bone marrow. Next, Reticulocytes enter the circulating blood and mature into functioned RBCs [7]. The average life span of those mature RBCs is 110 days [86].












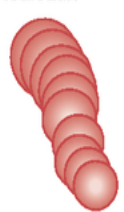














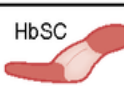
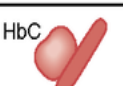
RED BLOOD CELL MORPHOLOGY					
Size variation	Hemoglobin distribution	Shape variation		Inclusions	Red cell distribution
Normal 	Hypochromia  1+	Target cell 	Acanthocyte 	Pappenheimer bodies (siderotic granules) 	Agglutination 
Microcyte 	 2+	Spherocyte 	Helmet cell (fragmented cell) 	Cabot's ring 	Rouleaux 
Macrocyte 	 3+	Ovalocyte 	Schistocyte (fragmented cell) 	Basophilic stippling (coarse) 	
Oval macrocyte 	 4+	Stomatocyte 	Tear drop 	Howell-Jolly 	
Hypochromic macrocyte 	Polychromasia  (Reticulocyte)	Sickle cell 	Burr cell 	Crystal formation  HbSC  HbC	

Figure 3: RBCs abnormal morphologies [2]

In certain health conditions, the body produces abnormal RBCs, Figure 3 illustrates some of the commonly seen abnormalities. These abnormalities can be summarized as:

1. Abnormal cell distribution: RBCs should be slightly separated and barely touching when spread on a slide, but when the cells are not distributed in this manner then an abnormality is identified. For example, when cells appear in stacks (rouleaux)

or chains then the presence of some health conditions like multiple myeloma and connective tissue disorder can be concluded [58].

2. Abnormal cell variation: some indices are commonly used in labs to describe RBCs, those indices include Average RBC size (MCV), Hemoglobin amount per RBC (MCH), the amount of hemoglobin relative to the size of the cell per red blood cell (MCHC). When the cells' MCV does not fall in the normal range, then an abnormality is identified. For example, Macrocytic is identified when cells with a diameter of approximately 9 μm or larger, having a mean cell volume (MCV) of greater than 100 μm are present in a blood sample. Macrocytic can be associated with B12 deficiency, or Hypothyroidism or smoking [8].
3. Abnormal shape variations: tear drop cells, burr cells, and sickle cells are examples of abnormal RBC shape variations. In sickle cell anemia for example, RBCs are shaped like crescent moons.
4. Abnormal hemoglobin distributions: this type of abnormality occurs when hemoglobin (Hb) concentration is not normal. When the Hb concentration is less than the minimum normal level then the cells are called hypochromic and many health conditions might be identified in this case, such as, iron deficiency anemia.

RBCs are the most common blood cells, for instance, the number of WBCs in adult males ranges from 4.5 to 11.5 thousand in 1 microlitre, where the number of RBCs in adult males ranges from 4.6 to 6 million in 1 microlitre [103].

Plasma

Plasma is considered as the largest component of the human blood. Plasma is made up of 92% water, 7% vital proteins, and 1% mineral salts, sugars, fats, hormones and vitamins [38]. Centrifugation is the process of separating plasma from blood, where the blood is set

to spin in a centrifuge with an anticoagulant. Plasma serves multiple important functions in human bodies:

1. Waste removal from cellular functions that help to produce energy. In addition, plasma transports the waste to the organs responsible for excretion (e.g. liver, kidneys).
2. Regulation of body temperature.
3. Transportation of nutrients like lipids, and vitamins from the digestive system to the rest of the body.

Convalescent plasma is the plasma extracted from the blood of patients recovered from a disease, and is used as a therapy to help other patients recover. For example, in 2020 the U.S. Food and Drug Administration (FDA) authorized this therapy for people with coronavirus disease 2019 (COVID-19). People suffering from COVID-19 can be given convalescent plasma of recovered patients to boost their ability to fight the virus [127].

White Blood Cells

WBCs, also known as leukocytes, are a heterogeneous group of nucleated blood cells. In healthy individuals, the count of these cells varies between 4000 and 10,000 per microliter [131]. Leukocytes are in charge of protecting the body against both foreign invaders and infectious diseases. The process of producing WBCs starting in the bone marrow and ending with mature cells circulating in the blood is called Leukopoiesis [29]. The 5 main types of WBCs are: Neutrophils, Monocytes, Eosinophils, Basophils, which result from the Myelopoiesis process, and Lymphocytes which results from Lymphopoiesis. More details about WBC types can be found in Appendix A.

Platelets

Platelets, also known as thrombocytes, are very small blood cells (i.e. the average diameter size is about 2 to 4 μm). Platelets help the body build clots to stop bleeding. This is because platelets send signals to attract more platelets and grow sticky edges to adhere to one another.

Platelets are activated once an activation cause is triggered (such as injury) forming very irregular edges and shaping a star-like appearance [125]. Finally, giant platelets are described as platelets that are abnormally large, i.e., as large as a normal red blood cell. The presence of this type is associated with some disorders, such as, Bernard-Soulier disease.

2.2 Phlebotomy and Blood Cell Separation

Phlebotomy, also known as blood draw or venipuncture, is the process of using a needle to take blood from a vein, usually in the arm. It is an important tool for diagnosing many medical conditions. This process has to be implemented by a well trained phlebotomist and the World Health Organization (WHO) elements of quality assurance in Phlebotomy have to be followed [5]. The collected blood sample is kept in a special tube that prevents clotting.

Finally, the blood sample goes through a separation process, which is implemented by placing the tube in a lab device called centrifuge. The centrifuge spins the blood sample for a specific amount of time (typically 15 minutes) at a specific speed. This process places the RBCs at the bottom, topped by the platelets, then the white blood cells and finally, plasma are placed at the very top.

2.3 Disease Diagnosis from Blood

Medical doctors rely on blood tests to: diagnose patients for certain diseases and medical conditions, examine the functions of certain organs and see how the patient is responding to prescribed treatments and/or medications. In this section blood tests will be categorised based on the type of blood sample needed for the test.

2.3.1 The Complete Blood Count Test

The literature shows that about 70% of clinical and medical decisions are backed by laboratory medicine [10]. In this section, tests that are performed in medical labs using a blood sample, rather than a blood smear are discussed. Such tests can be classified into many categories but in this thesis, we will only focus on tests related to blood cell types, subtypes, and their morphological abnormalities.

In medical labs, Complete Blood Count (CBC) is one of the most frequently requested tests by medical doctors. The CBC test counts the numbers of RBCs, WBCs, and platelets. It is used to diagnose several blood diseases such as anemia, as well as other conditions that indirectly affect the blood (such as dehydration, bone marrow disorders, Hemoglobin abnormalities, inflammation, Thalassemia, and Sickle cell disease). The CBC test also measures more detailed aspects of each blood component. The following 3 subsections will explain more about the aspects that are measured from each of the main blood types.

RBC tests The hematocrit, hemoglobin and RBC indices are all measures that the lab specialist reports in CBC results. The hematocrit measures the percentage (%) of the volume that the RBCs occupy within the whole blood. In automated analyzers, the hematocrit (HcT) is typically calculated from both the RBC count and the measured MCV. The hemoglobin level measures the total amount of the oxygen-carrying protein in the blood.

The laboratory professional interprets the accuracy of the RBC count, hemoglobin, and

hematocrit values using a quick mathematical check referred to as the rule of three. Simply, $RBC\ count \times 3 = hemoglobin \times 3 = hematocrit(\%)$. In case of these values do not agree within 3% of the measured values, an instrument malfunction or a measurement error could have occurred, or it is an indication that the patient could have a pathology that requires further investigation.

Finally, RBC indices, which suggest how the RBCs will appear microscopically as well as provide significant diagnostic information, are measured.

WBC tests The WBC differential is an analysis and enumeration of the various subtypes of WBCs [45]. An altered concentration of one specific type of leukocyte most commonly causes an increase or decrease in the total WBC count. For this reason, an abnormal total WBC count should be followed by a WBC differential, also known as diff. The differential results are reported as the percentage of each cell type counted. To accurately interpret whether an increase or decrease in cell types exists, the absolute concentration of each cell type is calculated using the results of the WBC count and the differential.

Platelets Automated hematology instruments generate the platelet count, which is reported as billions of platelets per liter (number of 10^9 platelets*/L). The mean platelet volume (MPV) is similar to MCV for erythrocytes because it represents the average volume of individual platelets. Laboratory professional utilizes both platelet count and MPV to assess thrombopoiesis and pathologic conditions related to platelets. A decreased platelet count generally represents decreased thrombopoiesis, increased platelet destruction, or consumption. Reactive or malignant conditions can cause an increase in the platelet count.

2.3.2 Peripheral Blood Smears

A Peripheral Blood Smear (PBS), also known as a blood film, is the result of spreading and staining a thin layer of blood on a microscope slide.

A well-made and properly stained blood smear is required for accurate interpretation. The slide is examined macroscopically and microscopically by a lab specialist to ensure that the blood was spread and stained properly. At a macroscopic level, a smear should look pinkish purple in color and transition to a feathered edge. Whereas at a microscopic level, the blood cells should be evenly distributed and areas between cells have to be clear. Moreover, Erythrocytes should look orange-red, Neutrophilic granules should look pale purple, Eosinophilic granules should look red-orange, Basophilic granules should look purplish black, Lymphocytes' cytoplasm should look blue, Leukocytes' nuclei should look purple, and finally Chromatin and parachromatin should be distinct within the nucleus.

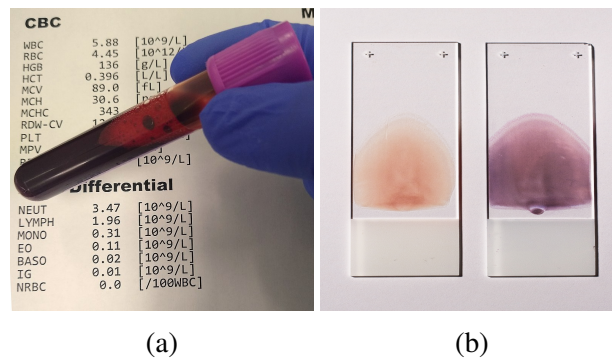


Figure 4: Different blood samples used in different blood tests

a: A blood sample collected in a tube for CBC with differential [138], b: A blood sample spread on a microscope slide for PBS analysis[34]

Despite the advancement in haematology, PBS remains as a very important diagnostic test to both haematologists and clinicians. In many cases, the CBC with differential test is not enough and more blood analyses are needed and a PBS is requested. Figure 4 shows the difference between the blood samples needed for CBC and PBS. The thorough examination of a peripheral blood smear can be used:

1. As a screening tool to identify illness.

2. For making the definitive diagnosis of certain hematologic and nonhematologic conditions.
3. To monitor the patient's response to therapy.

The peripheral blood smear evaluation includes an estimation of leukocyte and platelet count, the detection of abnormal cells and abnormal erythrocyte distribution, the review of erythrocyte and platelet morphology. A blood smear is mainly requested to:

1. Check the presence of immature, and abnormal cells.
2. Identify blood cells that are beyond the capabilities of the automated analyzers, i.e., morphological abnormalities.
3. Verify results obtained by automatic analyzers.

2.3.3 Clinical Laboratory Professionals Review and Diagnosis from CBC and PBS Data

The role of laboratory professionals is to analyze and interpret the data generated by the automated hematology analyzer and the manual peripheral blood smear review. The interpretation is essentially a correlation of the various components of the CBC in order to identify the likelihood of abnormal results, pathology, and discrepancies in the generated data.

Within the CBC report, the laboratory professional must be keenly aware of the critical limits that represent the critical low and high values for hematologic parameters. The normal values for hematologic parameters vary depending on age, gender, race, ethnicity, and geographic area. Thus, it is important that this information is available when reviewing data. For example, various parameters of the CBC are dramatically different in newborns

compared with adults, hence, the patient’s age must be considered when evaluating a patient’s blood picture. Hence, cells classified from the same blood smear are tested among those normal ranges in lab, to finally infer whether any of the medical conditions listed in Table 2 [70, 148] are present. The table lists 22 general medical conditions caused by the abnormal increase or decrease of WBC, RBC, or platelet types. Each of those listed medical conditions can reveal the existence of a group of illnesses, for example, Lymphocytopenia when accompanied with a low number of Monocyte indicates that the patient suffers from MonoMAC syndrome. Moreover, monocytosis is seen in chronic bacterial infections, inflammatory conditions such as Crohn’s disease, and malignancies such as chronic and acute myeloid leukaemia[10].

Table 2: Medical conditions and disorders the proposed system can infer

Blood cell type / abnormality	Abnormal high number	Abnormal low number
Lymphocytes	Lymphocytic, viral infection	Lymphocytopenia
Neutrophils	Neutrophilic leukocytosis, bacterial infection	Neutropenia
Monocytes	Monocytosis, Malaria, fungal infection	
Basophils	Basophilia, malignancy	
Eosinophils	Allergic reaction, fungal infection, Hypereosinophilic syndrome	
Platelet	Thrombocytosis, hypercoagulability	Thrombocytopenia, hypocoagulability
Giant Platelet	Macrothrombocytopenia	
Hypersegmented Neutrophils	Megaloblastic anemia	
Hyposegmented Neutrophils	Pelger-Huët anomaly	
Nucleated RBC	Leukemia, anemia	

2.3.4 Reticulocyte Smears

Reticulocytes are immature RBCs released by the bone marrow that circulate for one to two days in blood before becoming mature RBCs. Reticulocyte count is a routine test

that can be done by manual counting of reticulocytes and RBCs under oil immersion lens and calculate the percentage of reticulocytes per 100 RBCs. Automated hematology cell counters are also available, but the manual method is easy to perform and is preferred by most laboratories. Counting the number of reticulocytes in a blood sample can be a vital source of information for medical doctors to diagnose and assess patients' health condition [77]. Doctors request a reticulocyte count for the following reasons [1]:

1. Assess how well the bone marrow is functioning.
2. Diagnose and distinguish between types of anemia and assess how well the body is responding to iron deficiency treatments.
3. Monitor patients' health condition after undergoing chemotherapy, radiation therapy, or bone marrow transplant procedure.

The number of reticulocytes is normally less than 1% of the total number of RBCs. A higher percentage indicates an abnormal condition called reticulocytosis. Reticulocytosis can be a sign for many health conditions such as, hemolysis, hemorrhage, leukemia, pregnancy, recovery from vitamin B12, folate, or iron deficiency, and sickle cell anemia. On the other hand, an abnormal low number of reticulocytes is called reticulocytopenia. This condition can indicate bone marrow failure, chronic disease, folate deficiency, infection, iron deficiency, liver disease, malignancy, pernicious anemia and vitamin B12 deficiency.

Chapter 3

Methods

3.1 Study Scope

In this thesis, the automation of two types of blood smears is implemented; general PBSs (see section 2.3.2) and reticulocyte smears (see section 2.3.4).

Based on medical advice, the blood cell subtypes and morphological abnormalities of WBCs, RBCs and platelets that are listed in Table 3 are considered for general PBS automation. Throughout this study we will refer to WBCs, Nucleated RBCs and Platelets as the main blood cell types, while referring to the rest of the subtypes and abnormal morphologies as blood cell subtypes.

Table 3: Types of blood cells

Main blood cell type	Blood cell subtypes/ abnormalities
Platelets	Giant Platelets, Activated Platelets
WBCs	Atypical Lymphocyte, Atypical Monocyte, Band cell, Basket cell, Basophil, Eosinophil, Hyper-Segmented Neutrophils, Hypo-Segmented Neutrophils, Lymphocyte, Monocyte, Plasma cell, Reactive Lymphocyte, Segmented Neutrophils
RBCs	Nucleated RBCs

On the other hand, the second smear type covers two types of blood cells, reticulocytes and RBCs. The following two sections 3.2 and 3.3 discuss the automation of general PBSs, whereas section 3.4 presents the methods proposed for reticulocytes data scarcity and reticulocyte detection automation.

3.2 Generation of Synthetic Blood Smears Using Locality Sensitive Hashing

Deep neural networks are data hungry and data-driven. In this thesis, we opted to propose a method that produces a synthetic dataset of PBSs, due to the lack of public datasets that annotate blood cells to the main types and subtypes.

Researchers proposed different methods to generate synthetic data. The authors of [113] reviewed 200 publications that discuss synthetic image generation using Generative Adversarial Networks (GANs). GANs were utilized as tools to generate light microscopic synthetic images to increase the size of the training set [106, 49, 73].

Simulation to Reality (Sim-to-Real), Simulation to Simulation (Sim-to-Sim) are also techniques that include extracting labelled data from simulations to circumvent data scarcity. These methods are mostly used to train robots [55, 62, 88, 37].

In this section, a framework for constructing a synthetic balanced dataset of blood smears is proposed. Synthetic data have helped to improve the performance of NNs by providing sufficient instances that help NNs learn features of target classes. Synthetic instances do not necessarily need to look identical to real instances, but it must look realistic, hence, the quality of the smears produced by our framework does not matter as much as its ability to help the classification network better generalize [71].

The proposed framework aims to assemble images of segmented blood cells from all main and sub cell types as mentioned in Table 3 on blood smear canvases while keeping

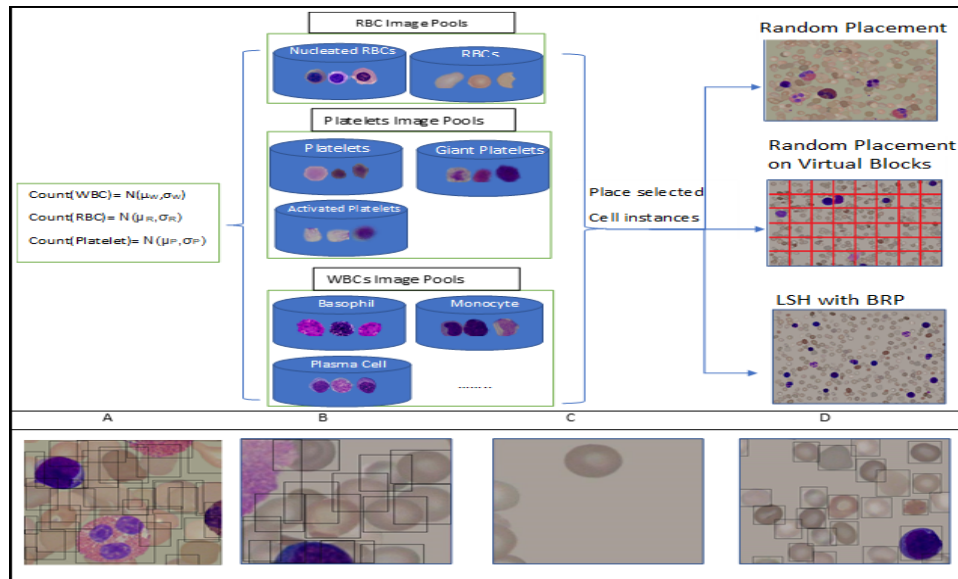


Figure 5: Constructing blood smear approaches as black box. *A* is an enlarged annotated sample result from the approach described in 3.2.2, *B* and *C* are enlarged sample results from the approach described in 3.2.2, *D* is an enlarged sample result from the approach described 3.2.2.

in mind the dataset balance issue and retaining the natural distribution of blood cells. The complexity of preparing a balanced dataset of blood smears comes from its natural structure; each blood smear contains hundreds of blood cells from different types, which are naturally distributed in an imbalanced manner. For example, in [122], the authors demonstrated that Red Blood Cells (RBCs) occurred approximately seven times more than White Blood Cells (WBCs) in the training set. This challenge implies that traditional augmentation techniques might not help as it will only amplify the imbalance issue.

This approach, as illustrated in Figure 5, runs in two stages; data pools preparation stage, and blood smears generation and annotation stage. The expected results of the proposed approach are:

1. A dataset of blood smear images.
2. A set of annotation files for the dataset instances that annotates blood cells to three main classes (Nucleated RBC, WBC, Platelet).

3. A set of annotation files for the dataset instances that annotates blood cells to the 17 subtypes of blood cells.

The following subsections provide more details about each stage.

3.2.1 Stage 1: Data Pools Preparation

In this stage, we aim to create an image pool for each main and sub blood cell type, hence 18 image pools will be processed and ready by the end of this stage; 17 blood cell subtypes and RBCs. The RBCs class is added as a pool because it still shows in all blood smears even if it is not annotated. Each of the 18 pools contains images of segmented cells of the pool cell type. All images were collected from public resources and processed by removing the background. In addition, all images were augmented by multiple rotations. Some restrictions need to be considered at this stage are:

1. All images must be subject to the same microscope magnification, because the size of the blood cell can be a major factor in distinguishing and classifying some morphological abnormalities.
2. All instances must be treated with the same stain for consistency purposes.
3. Each pool must be representative, well generalised, and comprise all possible appearances of the cell type. In other words, for each cell type we aim to collect distinct images that cover all possible features more than we aim for a large number of instances.

3.2.2 Stage 2: Blood Smears Generation and Annotation

In this stage, cell images from stage one are assembled on blood smear canvases to form thin blood smear instances. This procedure can be a bit complicated because each blood

smear contains hundreds of cells that have to be ordered in a natural realistic way. The following factors need to be considered at this stage:

1. No restrictions limit the selection of blood cell subtypes to appear in a blood smear, as the presence, absence or deficiency of each cell type or subtype represents certain types of syndromes or medical diagnoses that is independent from all other syndromes or medical diagnoses that can be concluded from other cell types appear in the same blood smear.
2. Total number of the main blood cell types must be carefully selected to represent realistic blood smears. Hence, RBC, Platelet and WBC cells are assumed to follow Normal (Gaussian) Distributions. The Normal distributions of WBCs and Platelets and their parameters were derived from the All-IDB dataset [69] statistics, however, the RBCs' Gaussian Distribution was assumed to follow the one in [23].

The Gaussian Distribution is a probability distribution that is typically used to model normal phenomena and is described by the probability density function (PDF). A PDF describes the probability of a value (x) of an experiment to fall within a particular range of values, it is mathematically represented by the following formula:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (1)$$

where, the mean denoted by μ is the arithmetic average of the data, and σ is the standard deviation that is calculated by:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N} \quad (2)$$

where, N is the number of sample observations. Table 4 lists μ and σ for each main cell type.

Table 4: Statistics of the main blood cell types

Blood cell type	Mean	Standard Deviation
Platelets	4.3	4.8
WBCs	8.6	10.5
RBCs	669	149

At this point an efficient approach is needed to place cell images from different pools on a blood smear background canvas. For this purpose, we first select three random numbers; a random number from each Gaussian Distribution. Second, subtypes of each main blood cell type are uniformly selected. Next, instances from each selected subtype pool are selected. When all subtypes and cell instances are selected from image pools, an efficient approach is needed to place these cells on canvas. In the following subsections we present and discuss three different placement strategies to paste the selected instances on blood smear canvas in a realistic fashion.

Naive approach 1: Random placement

In this approach random paste coordinates are selected for each cell. This approach executes fast and is easy to implement, however, it does not guarantee the spread of cells on the smear canvas. As shown in Figure 5-A, this method forms cell clumps, and some cells might override others which leads to wrong annotations. Figure 5-A shows how bounding boxes are heavily intersected.

Naive approach 2: Random placement on virtual blocks

In this approach the blood smear background canvas is divided into $M \times N$ virtual blocks. For each cell a block is uniformly chosen, then random paste coordinates are selected inside the chosen virtual block. A drawback of this approach is that some blocks might be selected more than others which will also form cell clumps. Figure 5-B illustrates a crowded block

where 5-C illustrates an almost empty block.

Nearest neighbour mining approach

The main shortcoming resulted from the previous naive approaches is the formation of clumps in the generated blood smears. To avoid placing cells on canvas with high probability of occlusion we need to choose a paste location that does not overlap with any other surrounding cell. Implementing this in a brute-force manner can guarantee the accuracy of the results but the processing time grows linearly with the number of cells on canvas. On the other hand, choosing a potential paste point and estimating its nearest neighbors and reject those points that will cause occlusion with neighboring objects can reduce the number of comparisons and the complexity. Locality Sensitive Hashing (LSH) is a nearest neighbour retrieval algorithm that can be utilised for this purpose because it is a generic hashing technique that intends to preserve the local relations of the data.

In our problem, we have a set of cell objects to be pasted on a canvas, each object will be represented by its top left coordinates, called paste points, hence our space is a 2D Euclidean space. A dictionary keeps the width and height of the paste point's corresponding cell. Our goal is to retrieve the nearest neighbor points to each potential paste point, check if the dimensions of the paste point will overlap with its neighbors or not within a certain threshold, and finally decide to accept this new point or reject it. An effective approach to implement our goal on the mentioned Euclidean space can be dividing the space by a set of projections and hash near points into the same bucket. This approach is called Random Projections.

The core idea behind random projections is given in the Johnson-Lindenstrauss lemma,[57] which states that if points in a vector space are of sufficiently high dimension, then they may be projected into a suitable lower-dimensional space in a way which approximately preserves the distances between the points. This can be represented as:

Let \mathbb{X} be a space of objects [22], to which dataset and query objects belong. Let D be a distance measure defined on \mathbb{X} . Let H be a family of hash functions $h : \mathbb{X} \rightarrow \mathbb{Z}$, where \mathbb{Z} is the set of integers.

Let R_1, R_2, P_1 , and P_2 be real numbers. For any points X_1 and X_2 in \mathbb{X} that are close to each other, there is a high probability P_1 that they fall into the same bucket:

$$P_H[h(X_1) = h(X_2)] \geq P_1 \text{ for } D(X_1, X_2) \leq R_1 \quad (3)$$

Moreover, for any points X_1 and X_2 in \mathbb{X} that are far apart, there is a low probability $P_2 < P_1$ that they fall into the same bucket:

$$P_H[h(X_1) = h(X_2)] \leq P_2 \text{ for } D(X_1, X_2) \geq cR_1 = R_2 \quad (4)$$

Let L denote the number of random projections, then the space will be partitioned using L hyperplanes by selecting pr_1, \dots, pr_L vectors at random from a Gaussian distribution. Then each dataset and query objects are hashed using equation 5, where m is any potential paste point, and i is the random projection index.

$$[h(m)]_i = \begin{cases} 0 & pr^T m \leq 0 \\ 1 & pr^T m > 0 \end{cases}, i = 1 \dots L \quad (5)$$

In this work, random binary projections method is implemented, hence, all values will be projected to either 0 or 1. Object m will then be stored in a hash table with its hash value, $h(m)$, as its key. Every time a potential paste point is queried against the LSH engine, it will be hashed using equation 5, and all previous points in the same bucket will be returned as possible neighbors. Each of the returned neighbors will be checked against a distance criteria, if all neighbor points are farther than a certain threshold, then the potential paste

point will pass and the object will be pasted. Else, if at least one neighbor point is closer than the same threshold, then the point will be rejected and a new paste point will be selected and queried against the LSH engine. Since we are dealing with cells as bounding boxes, a good distance measure will be measuring the intersection over union ratio, also known as Jaccard similarity between the potential paste point and its neighbors. Jaccard Similarity can be defined as:

$$J(p, n_i) = |p \cap n_i| / |p \cup n_i| \quad (6)$$

where n_i denotes a neighbor point, and p denotes the potential paste point. The Jaccard similarity value between each neighbor and the paste point will be checked by equation 7 for the validity of the paste point.

$$Evaluation[p] = \begin{cases} 0 & \exists n_i, J(p, n_i) > T \\ 1 & Else \end{cases}, n_i \in N \quad (7)$$

Algorithm 1 demonstrates more details of the proposed approach. The following parameters are defined in Algorithm 1:

- R_w, R_R, R_P are the random counts of WBCs, RBCs, and Platelets respectively. Each of these counts represents the number of cells that will appear in the being generated blood smear instance.
- LSH : is the locality sensitive hashing engine, that is initialized once instantiated to D dimensions and L projections.

Algorithm 1 Creating a blood smear dataset of size DS

```
1: for  $i \leftarrow 1$  to  $DS$  do
2:    $R_W \leftarrow \mathcal{N}(\sigma_W, \mu_W)$ 
3:    $R_R \leftarrow \mathcal{N}(\sigma_R, \mu_R)$ 
4:    $R_P \leftarrow \mathcal{N}(\sigma_P, \mu_P)$ 
5:    $LSH \leftarrow D(Dimensions), L(Projections)$ 
6:   for  $R \leftarrow R_W, R_R, R_P$  do
7:      $SubTypesCount \leftarrow Rand(1, Max_{types})$ 
8:      $Max \leftarrow 1$ 
9:     for  $S \leftarrow 1$  to  $SubTypesCount$  do
10:       $TypePercentage \leftarrow Rand(Min, Max)$ 
11:       $Max \leftarrow Max - TypePercentage$ 
12:      for  $Count \leftarrow 1$  to  $TypePercentage * R$  do
13:         $Cell \leftarrow Rand(Pool(S))$ 
14:         $RejectPoint \leftarrow 1$ 
15:        while  $RejectPoint$  do
16:           $PastePoint \leftarrow Rand(Canvas)$ 
17:           $Neighbors \leftarrow LSH.NN(PastePoint)$ 
18:          for  $N \leftarrow Neighbors$  do
19:            if  $Jaccard(N, PastePoint) \geq T$  then
20:               $RejectPoint \leftarrow 1$ 
21:               $Break;$ 
22:            else
23:               $RejectPoint \leftarrow 0$ 
24:            end if
25:          end for
26:          if  $RejectPoint = 0$  then
27:             $Paste(PastePoint, Cell, Canvas)$ 
28:             $Anotate(PastePoint, MainTypesFile)$ 
29:             $Anotate(PastePoint, SubTypesFile)$ 
30:          end if
31:        end while
32:      end for
33:    end for
34:  end for
35:   $SaveAnnotation(MainTypesFile_i)$ 
36:   $SaveAnnotation(SubTypesFile_i)$ 
37:   $SaveImage(Canvas_i)$ 
38: end for
```

- *Max*: for each of the main cell types (WBCs, Platelets, Nucleated RBCs), its corresponding subtypes have to appear in percentages that sum up to 1. The *Max* parameter is initialized to 1, and is later decreased as the algorithm progresses.

Some highlights from Algorithm 1 are:

- In lines 6 to 9: for each of the main blood cell types, subtypes are randomly chosen and stored in the *SubTypesCount* parameter. These subtypes and the percentages of which each of them will contribute in the blood smear instance are then randomly selected.
- In lines 11 to 15: cell instances are selected from corresponding pools, one cell instance at a time, and a random paste point is then selected within the blood smear canvas. Next, the LSH engine retrieves all potential neighbor points.
- In lines 16 to 22: each neighbor is checked against the potential paste point using the Jaccard similarity metric. If any neighbor overlaps with the potential point past the allowed percentage then the potential paste point will be rejected.
- In lines 35 and 36: the annotations of the generated blood smear are saved.

Two annotation files are automatically generated for each generated synthetic blood smear, hence two sets of annotations will be provided for the dataset:

1. The first set of annotations targets multi-step classification pipelines, which classify the main types of blood cells, then further classify the regions of interest into its corresponding subtypes in a second phase. Since Nucleated RBCs is the only subcategory of RBCs, this annotation set will classify cells into: WBCs, Platelets, and Nucleated RBCs.
2. The second set of annotations targets one-step classifiers, where the system classifies whole blood smears into all 16 subtypes. Platelets category is also annotated in this

set based on the advice of a medical expert. Hence, this set of annotations includes 17 categories.

3.2.3 Medical Assessment

A questionnaire of 12 questions about the dataset was created to ensure that the dataset meets the medical standards of thin blood smears. The questionnaire was filled by 5 experienced hematologists. A blood smear was displayed in each question, followed by some questions to evaluate the following aspects about the dataset instances:

1. **Factor 1:** The general quality of the smears in terms of the total numbers of RBCs, WBCs and Platelets.
2. **Factor 2:** The correctness of the first set of annotation, by asking the respondents to verify some labels.
3. **Factor 3:** The correctness of the second set of annotation.
4. **Factor 4:** The quality of blood cell subtype choices.
5. **Factor 5:** The level of overlap and occlusion between the blood cells on the synthetic blood smears to assess the quality of the locations produced by our proposed algorithm.

The questionnaire was initially reviewed and verified by a designated hematologist for quality assurance purposes. The average years of experience of the participating hematologists is 6 years. The expected answers were set before starting this experiment and will be denoted by the ground truth throughout this thesis. The questionnaire can be found here [92].

3.3 General PBS Analysis Using Deep Learning

In this section, the framework illustrated in Figure 6 is implemented in order to classify blood cells that appear in general blood smears. In the first phase a deep network detects and classifies blood cells from whole-slide smears into Nucleated RBCs, WBCs, platelets and other objects. The detected objects are then cropped and passed to the second phase. The nucleated RBCs are not further classified. On the other hand, there are 13 WBC subtypes and 3 platelet subtypes. Hence in the second phase, the platelet instances are passed to a classifier and an image processing module. Moreover, the cells that were classified as WBCs in the first phase are emitted to a deep-classifier-based module called the incremental training module. It is worth mentioning that the plasma WBC cells were excluded in phase 2 due to the lack of test instances, therefore, the number of WBC subtypes that will be considered in the incremental training phase are 12.

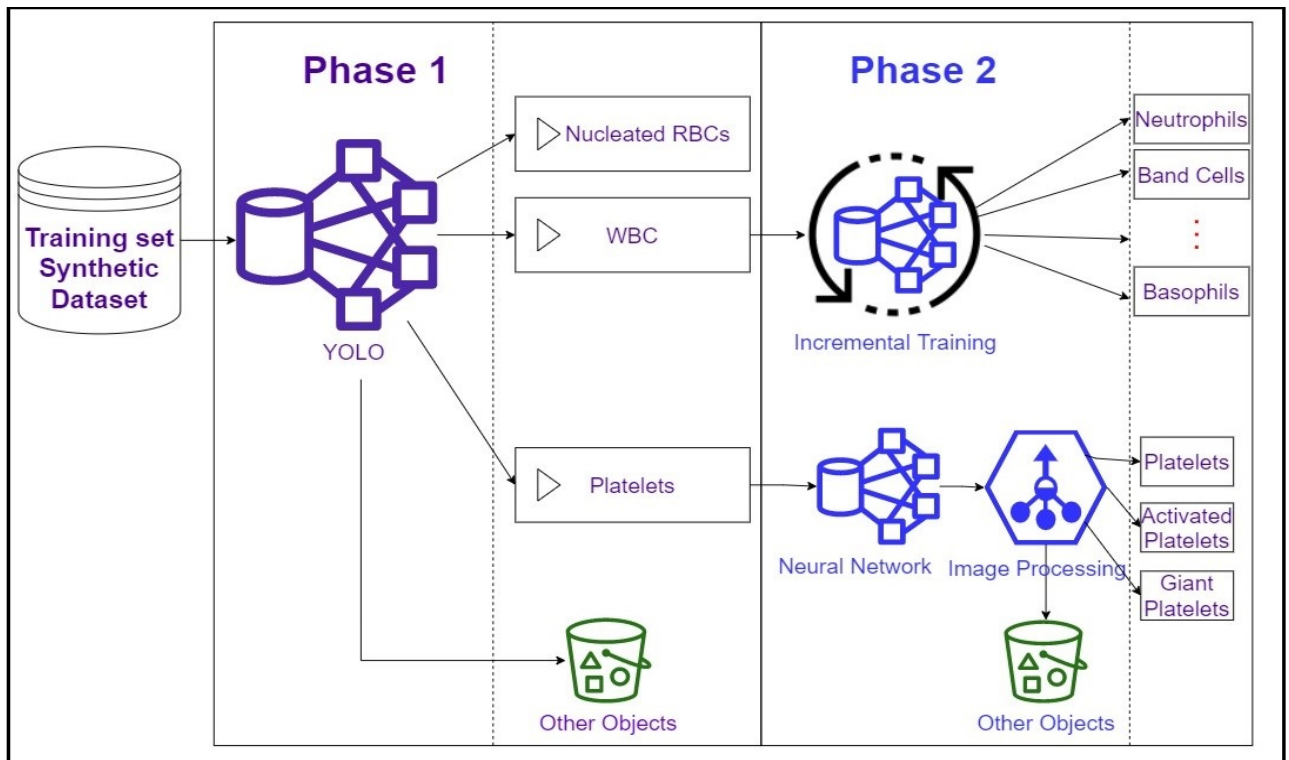


Figure 6: General workflow of the proposed blood smear analysis framework

3.3.1 Phase 1: Classification of Main Blood Cell Types using YOLO Deep Network

In this phase, the synthetic blood smears along with the first set of annotations; the main types annotations, are fed to two different YOLO deep networks. All computations mentioned in this section were made on the supercomputer Helios from Laval University, managed by Calcul Québec and Compute Canada.

YOLO deep network [99], is a real time object detection and classification network that performs object detection and classification in one scan. YOLO is a convolutional neural network which divides the input image into an $N \times N$ grid. Each grid cell predicts bounding boxes and confidence scores for those boxes, where confidence scores mirror how confident the network is that the bounding box contains an object. Each bounding box consists of 5 attributes: $(x, y, h, w, \text{confidence})$ where, x and y are the coordinates of the center of the bounding box, h and w are the height and the width of the bounding box. Moreover, each grid cell predicts C conditional class probabilities to classify the object that is located in the grid cell. A deep network loss function aims to minimize the network error and in YOLO, it is calculated as the combination of localisation and classification errors :

$$\begin{aligned}
 & \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\
 & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{noobj} (C_i - \hat{C}_i)^2 + \lambda_{coord} \sum_i \mathbb{1}_i^{obj} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
 \end{aligned} \tag{8}$$

where $\mathbb{1}_i^{obj}$ denotes if object appears in cell i and $\mathbb{1}_{ij}^{obj}$ denotes that the j th bounding box predictor in cell i is responsible for that prediction. A relatively small NN of 13 convolutional layers, namely, tiny YOLOv3 network [100] and a larger network that consists of 23 convolutional layers, namely YOLOv2 [99] are utilized for the experiments of this phase. Moreover, in order to further improve the result, the network input resolution is randomly

resized every 10 batches during training, this regime which was proposed in [99] works like data augmentation and helps the network learn to better generalise. The networks ran with a momentum value of 0.9, learning rate of 0.001, and network input resolution of 800 as an initial resolution. All experiments were trained with the synthetic dataset and validated and tested with the ALL-IDB1 Dataset.

In this phase, objects that are detected with a confidence score less than 10% are classified as "other objects" and are saved in a separate directory in order to be further examined by the lab technician. The threshold 10% was empirically set.

3.3.2 Phase 2: Classification of Blood Cell Subtypes

In image classification applications of deep learning, a neural network is trained on a set of images until an acceptable error rate is achieved. In the case of supervised learning where all instances are labelled, like in this study, a training dataset DS of size N can be represented as $\{(x_n, y_n)\}$ for $n= 1, \dots, N$, where $x_n \in \mathbb{R}^d$ is the instance feature vector. If there exist M possible output labels, then the labels set can be expressed as $C=\{c_1, \dots, c_M\}$, and each instance label which is a subset of the possible M output classes can be expressed as $y_n \in C$. In this work, a set of deep neural networks are utilised to classify 15 blood cell categories, hence, the feature vector of x_n is extracted automatically by the network and the size of the possible output classes set C is 15. When training a classifier we aim to have a discriminant function $f(x \rightarrow \beta_m)$ that maps each instance x_n to M class-specific parameters for each class β_1, \dots, β_M . Such functions are used to classify each test sample x as the class label that scores the highest β parameter, as shown in equation 9. In this work the softmax function will be utilized to calculate the β values.

$$arg \max_m [f(x \rightarrow \beta_m)] \quad (9)$$

To tackle the complex nature of blood cell classification problem, the classification process will be divided into 2 parts: the first part deals with platelets classification, and the second part deals with WBCs classification. These approaches are described in detail in the next two subsections.

The experiments of phase 2 are all trained with the synthetic dataset. For testing and validation, a hematologist with 9 years of experience reannotated the ALL-IDB1 dataset to cover the classes considered in this study. The dataset consists of 459 platelet instances and 906 WBC instances. In order to increase the size of the dataset that will be utilized for validation and testing, cropped cell images available on the web were collected. Search keywords like "microscopic band cell" and "microscopic white blood cells" were typed on the Google search engine. After a careful observation, images of good visual quality and that were taken with a magnification within the range of 300-500 were considered. One hundred fifty platelet instances and 1,100 WBC instances were collected. The blood cells from the All-IDB dataset and the images collected from the web were combined together for validation and testing.

3.3.3 Classifying platelets

The main challenge in classifying platelets arises from its very small size, which makes it difficult for the network to recognise and extract accurate and distinguishing features [32]. In this work, three categories of platelets are considered; normal platelets (a.k.a platelets), giant platelets, and activated platelets.

The main visual difference between platelets and giant platelets is the length of their diameters. On the other hand, activated platelets look different than the other two categories, because cells that belong to this category have very irregular edges. Pseudopods form on the surface of activated platelets, forming a star-like appearance once an activation cause is triggered (such as endothelial damage) [125]. Figure 7 illustrates examples of platelets and

activated platelets.

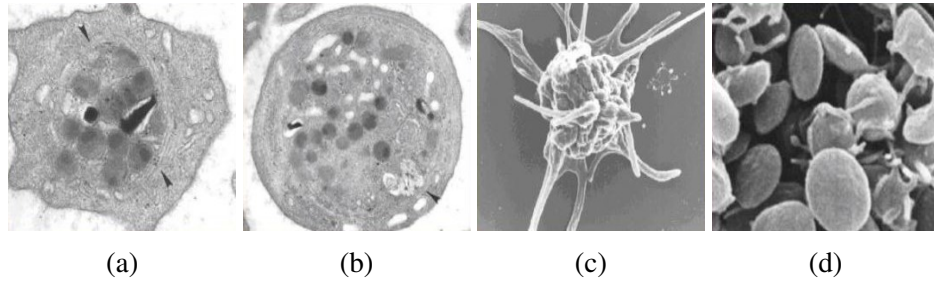


Figure 7: Activated platelets and normal platelets

(a) and (c): Activated platelet, (b) and (d): Normal platelet

In order to perform this classification task, a one-phase classification technique is enhanced and a two-phase classifier is proposed. In the next two subsections the neural networks that were utilized for classifying platelets and the proposed classification methods will be discussed. Figure 8 depicts the steps of both classification approaches.

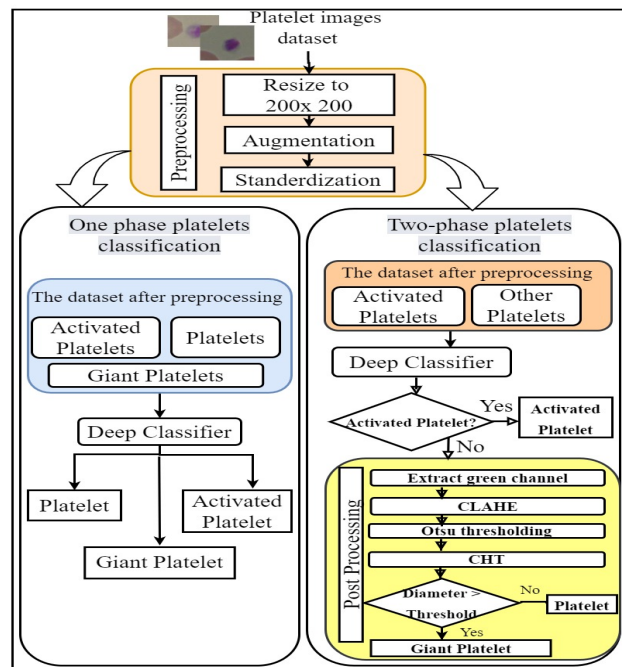


Figure 8: Platelet classification approaches

One-phase classification

The choice of deep classifier can directly affect the accuracy of the classification results, more complex contexts need more complex functions to be learned by the neural network. The literature reveals that deeper networks have performed well in classifying blood cells [16], hence in this work neural networks with a relatively high number of layers will be considered.

For the purpose of classifying platelets the VGG16, VGG19, ResNet50, InceptionV3, and DenseNet121 networks available on the Keras applications library [64] were experimented, but only the top two networks were considered. All the mentioned networks except the VGG19 and ResNet50 scored validation accuracies less than 70% when trained on the platelet images. Hence only the VGG19 which consists of 19 layers and the ResNet50 which consists of 50 layers were considered. It is worth mentioning that the reason behind the superiority of the results achieved by the VGG19 and the ResNet networks compared to the results obtained from the other neural networks is considered beyond the scope of this study. Before feeding the images to the VGG19 and the ResNet50 networks, the platelet instances were resized to 200 x 200. Next, the resulting images were augmented by applying rotation, and horizontal and vertical flips. The augmentation techniques and the resizing factor were imperically set. Finally, the images were converted from RGB to BGR, then each color channel was standardized with respect to the ImageNet dataset, without scaling, using the Keras "preprocess input" function. The work of M. Shanker and collaborators discusses the benefits of data standardization [114]. Both neural networks were initialized to the ImageNet pretrained weights, and customised by removing the pretrained fully connected layers, and freezing all the remaining layers. Next, an extra fully connected layer and a softmax layer were attached to the networks, finally, the batch size hyperparameter was tuned and the value 128 was considered. Table 5 summarizes the properties of the fully connected network that was utilized for this task.

Table 5: The properties of the fully connected neural network utilized for platelet and WBC classification

Framework	Keras 2.4.3
Loss function	Cross Entropy
Learning rate	.01
weight decay	1e-6
momentum	.9
Optimization algorithm	mini-batch Gradient Descent
Activation function	ReLU, Softmax
fully connected layer size	1000

Two-phase classification

This approach is proposed to improve the results obtained from the first approach. Here, the same ResNet and a VGG19 networks that were utilized in the one-step approach were trained to classify platelets into two classes only; activated platelets and "other platelets". The random undersampling technique was applied in order to reduce the giant and normal platelets training instances to one class while retaining the training dataset balance. Platelets that are classified as "other platelets" are further processed in the second phase by applying a set of image processing techniques. Each instance from the "other platelets" class is first processed by extracting the green channel, because most contours of blood cells appear continuous and contrasted against the background in the green channel [18]. Next, Contrast Limited Adaptive Histogram Equalization (CLAHE) [87] and Otsu thresholding [82] are applied. All the above mentioned image processing techniques were adopted based on their success in the context of blood cell classification [66, 13, 80, 20].

Finally, Hough Circle Transform (HCT) [141] is applied to outline the platelet cell with a circle and measure its diameter. Based on the diameter size, the instance is classified as either normal or giant. The size ranges for normal and giant platelets were decided with the help of a hematologist, the hematologist chose the largest and smallest normal platelet and giant platelet cells that appeared in the ALL-IDB datasets, and the diameter of each of those chosen cells was measured, next, those measurements were used as limits for the

ranges utilized in this study. This step imitates the way lab technicians classify platelets in a lab. A lab technician typically, estimates the diameter of platelet cells with bare eye under a microscope to decide whether the platelet cell is giant or normal. All objects that do not fall in the specified diameter range are moved to the "other objects" directory in order to be further examined by the lab technician.

3.3.4 Classifying WBCs

In this section, 12 WBC categories and morphological abnormalities are considered for classification. This relatively high number of classes, in addition to the sensitivity of making wrong classifications in medical diagnoses, adds an extra layer of complexity to this task. Many classes included in this study have high visual resemblance [126, 101, 120], which might affect the classification accuracy. Typically, a one-phase classifier is used for this task, however, to accommodate for the complexity of WBCs classification, the one-phase technique is extended to a two-phase classifier:

1. **One-phase WBC classification:** In this approach all WBC instances from all 12 classes are classified by a single neural network. For this task, VGG16, VGG19, ResNet50, InceptionV3, and DenseNet121 networks available on the Keras applications library were experimented, but only the VGG19 was considered for this approach and the enhanced incremental training approach because it achieved the best results. It is worth mentioning that the reason behind the superiority of the results achieved by the VGG19 network compared to the results obtained from the other neural networks is considered beyond the scope of this study. The WBC instances were first resized to 200 x 200. Next, the resulting images were augmented by applying horizontal and vertical flips. The augmentation techniques and the resizing factor were empirically set. Finally, the images were standardized using the Keras "preprocess input" function. The VGG19 network was customised by removing the

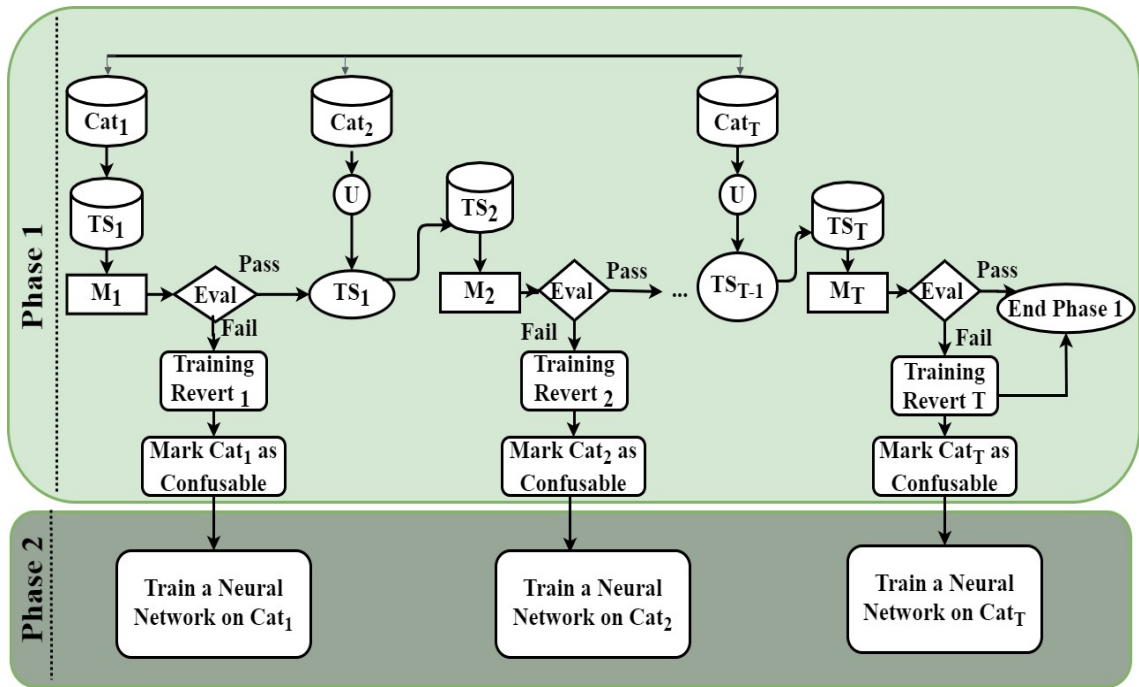


Figure 9: The enhanced incremental training framework

pretrained fully connected layers, and freezing all the remaining layers. Next, an extra fully connected layer and a softmax layer were attached to the network. The fully connected network utilized in this approach is similar to the one used for platelet classification, which is summarized in Table 5.

- 2. Deep neural networks pipeline trained using enhanced incremental training:** In this approach WBCs are classified in 2 phases. In the first phase, all non confusable classes are classified, and classes considered as confusable are passed to the second phase classifiers. For the purpose of figuring out the confusable classes, we opted to propose an enhanced version of incremental deep neural networks training [121], this approach will be explained in detail in this section.

Enhanced incremental training

Deep networks are black boxes [9], and knowing what exactly causes wrong predictions can be difficult, or even impossible [56]. Moreover, in any classification task, as the number of classes considered in a dataset increases, it becomes more infeasible to analyze and look for the reason behind high rates of wrong predictions from the confusion matrix. Hence, researchers tend to set some hypothesis on what might cause the low correct predictions rate, and then handle such hypothesis to measure and test its effect on the neural network performance.

In this work the term **confusable classes** is used frequently. Confusable classes are classes that include instances that have many visual attributes and features in common, which cause a higher level of confusion in the network performance and harm the overall classification results [145, 105]. Visual resemblance between different classes can make the classification task more complicated, as feature vectors for their instances come out very similar and lead to wrong estimations and results by the function mentioned in equation 9, and hence an increased rate of wrong classifications. In this study, the number of WBC classes is relatively high, some of these classes are separable and exhibit insignificant visual similarities (e.g., Basket cells and Eosinophils). On the contrary, many other classes are confusable (e.g., Lymphocyte and Atypical Lymphocyte), Figure 10 shows cells from separable and confusable classes. Hence, our hypothesis in this study is that the high visual resemblance between some WBC classes is the reason behind the high rate of wrong predictions.

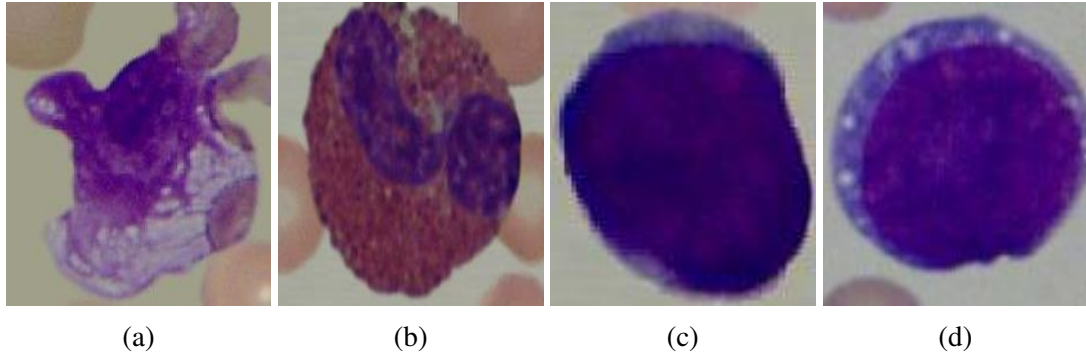


Figure 10: Instances from separable and visually similar blood cell types, with microscope at magnification 300x

(1) Separable instances: a: Basket cell b: Eosinophil (2) Confusable instances: c: Lymphocyte d: Atypical Lymphocyte [69].

Deep networks are typically trained on instances from all classes at the same time in a single training stage. But in order to test our hypothesis, the incremental training method proposed in [121] will be enhanced and utilized in a novel way in this work to recognize and handle confusable classes.

In incremental training, a training dataset TS is divided into S disjoint subsets of size L , i.e., each subset contains all the training instances of L classes, this implies that:

$$\sum_{i=1}^S L_i = |C| \quad (10)$$

where $|C|$ is the size of the set of all possible classes or output labels, S is the number of subsets, and L_i is the number of classes included in the i_{th} subset. The network is initially trained on the first subset, then the resulting weights are passed to the next stage of training, where another subset is added to the training set. A new training process is initiated every time a new subset is added. Eventually, the network training will be completed after S training stages. The regime of passing weights between the training stages acts like transfer learning, hence the learned knowledge is accumulated among the stages. We formalised incremental training in Algorithm 2.

Algorithm 2 Incremental Training

Parameters:

 i : training stage index ts_i : the training set at the i th stage W_i : the weights resulted from training the network at the i th stage s_i : the i th subset1: $S \leftarrow |C|/L$ 2: $s \leftarrow \{s_0, s_1, \dots, s_{S-1}\}$ 3: $W_{-1} \leftarrow InitialWeights$ 4: $ts_{-1} \leftarrow \emptyset$ 5: $i \leftarrow 0$ 6: **while** $i < S$ **do**7: $ts_i = ts_{i-1} \cup s_i$ 8: $NeuralNetwork.ReshapeOutput(CurrentShape + L)$ 9: $W_i \leftarrow NeuralNetwork.Train(ts_i, W_{i-1})$ 10: $i \leftarrow i + 1$ 11: **end while**

The concept of incremental learning has been proposed in the literature as a way of solving the problem of adapting neural networks to learn to classify instances from classes that were not in the initial training set [44]. For instance, in [104], a deep network that hierarchically grows to adapt to new classes that were not introduced during the training phase was proposed. In this work, the deep network expanded in a tree-like mode, where classes were categorised, and each category was represented by a branch in the Tree-CNN. Every time a new class was introduced, only the corresponding branch was retrained.

Sarwar and et al. [109], introduced a network that incrementally grows by adding convolutional kernels and fully connected layers to the later layers of the network in order to adapt to new classes. An advantage of this method is that the network adapts to new classes without the need to retrain on the previously seen instances, which might not be available at the time of retraining. In [30], Ferrari and collaborators proposed a method that trains deep neural networks incrementally, using new data and a small sample from the previously seen classes. This method is based on the use of a combination of cross-entropy and distillation loss functions. The works in [53, 121] proposed training schemes that

considered exposing the network gradually to the training data in a way that imitates the learning patterns of human beings. The work of Istrate and collaborators [53] performed incremental training by first establishing an original network that was later divided into sub-networks. However, this method slightly improved the accuracy of the model.

Our proposed enhanced incremental training strategy utilises the concept of incremental training as a backbone while adhering additional rules and procedures. In the proposed method, we aim to train the network gradually to identify classes that increase the wrong predictions. Moreover, in order to handle the side effects of classes’ visual resemblance, classes are categorised prior to training, and are undersampled into one class during training in case of performance degradation. Undersampling similar classes into one class can ease the network task, as all instances that look similar will be classified the same. After the completion of the proposed enhanced incremental training algorithm, categories that are recognized to include confusable classes will be classified in a second phase. The proposed algorithm is presented in detail in Algorithm 3 and Figure 9.

Table 6: WBC categories

Category ID	Category Name	Category Members
Cat_0	Separable WBCs Category	Band cell, Basket cell, Basophil, Eosinophil
Cat_1	Monocyte Category	Monocyte, Atypical Monocyte
Cat_2	Lymphocyte Category	Atypical Lymphocyte, Reactive Lymphocyte, Lymphocyte
Cat_3	Neutrophils Category	Hyper-Segmented Neutrophils, Hypo-Segmented Neutrophils, Segmented Neutrophils

The main highlights from Algorithm 3 and Figure 9 are:

1. **Categorization and initialization:** In line 1 all classes are categorised based on their visual similarity. In this study, the WBC classes are clustered in 4 categories. Table 6 lists each category and its member classes. The Lymphocyte, Monocyte, and Neutrophils category members were grouped based on their high visual similarity, i.e.,

Algorithm 3 Enhanced Incremental Training

Parameters:

TS : training set

W_i : the weights resulted from training the network at the i th stage

$CategoryTS$: the training set at the beginning of the stage

$Current_shape$: the number of the network output nodes

▷ **Categoryization and initialization**

1: $Categories \leftarrow \{Cat_0, Cat_1, Cat_2, Cat_3\}$

2: $W_{-1} \leftarrow ImageNetWeights$

3: $TS \leftarrow \emptyset$

4: $i \leftarrow 0$

▷ **Category looping**

5: **for** $Category$ in $Categories$ **do**

6: $CategoryTS \leftarrow TS$

7: $CategoryOutputShape \leftarrow Current_Shape$

8: $CategoryWeight \leftarrow W_{i-1}$

9: $TS \leftarrow TS \cup Category$

10: $Nodes \leftarrow Size(Category)$

11: $Network.Reshape(CurrentShape + Nodes)$

12: $W_i \leftarrow Network.Train(TS, W_{i-1})$

▷ **Stage evaluation**

13: **if** $Evaluate(Network) \geq Threshold$ **then**

14: ▷ **Training revert**

15: $SubSampledCat \leftarrow SubSample(Category)$

16: $TS \leftarrow CategoryTS \cup SubSampledCategory$

17: $Network.OutputShape \leftarrow CategoryShape + 1$

18: $W_{i-1} \leftarrow CategoryWeight$

19: $W_i \leftarrow Network.Train(TS, W_{i-1})$

20: **end if**

21: $i \leftarrow i + 1$

22: **end for**

members of the same group are highly probable to be confused to each other. The Separable WBCs Category comprises the remaining classes that exhibit insignificant morphological and visual similarity and are thought to be easily distinguished from each other. This categorisation was planned based on the advice of a medical expert. The categorization was planned based on human's perspective on what classes are confusable, but since neural networks might have a different point of view on what is confusable, the following steps in the algorithm will let the network test the categorization and decide which categories are confusable from the neural network perspective.

2. **Category looping:** In line 5 the algorithm loops over each category, and at the beginning of every training stage, a new category is added to the training set. Next, in lines 6 to 8, the values of all training parameters at the beginning of each stage are kept, to be used if needed later on in running the "training revert" procedure (more details in point 4). Finally, in lines 9 to 12, a new category is added to the training set while keeping the previously added categories, this act aids in avoiding the "catastrophic forgetting" problem, which refers to the destruction of features learned from previously added data when the neural network was only trained on data of the newly added category [139]. Next, the network output size is adjusted to classify the newly added category, and finally the network weights are assigned to the weights resulted from the previous stage. Utilizing the weights resulted from the previous stages aids in transferring the network knowledge about the previously added categories to the current stage, this way the network will not need to retrain from scratch, but it only fine-tunes the previous knowledge to accommodate the newly added category.
3. **Stage evaluation:** In line 13, the model is tested on the validation set and if the network metrics drop, then we conclude, based on our hypothesis, that the category

contains one or more confusable class(es) that caused a higher number of wrong predictions. The threshold of this drop decides whether the category contains confusable classes or not can be numerically or even logically set depending on the problem domain. The criterion that was empirically chosen for this study is 5% drop in the validation accuracy. Hence, the training revert procedure was executed every time the validation accuracy dropped more than 5% compared with the accuracy obtained in the previous stage. It is also worth mentioning that the validation set of WBCs was chosen to contain equal proportions of data from all classes, which facilitated the usage of the accuracy metric. The evaluation of the first model is an exceptional case since there is no previous stage to compare it to, hence, the first model's evaluation accuracy is compared to a baseline value. In this study the baseline value is the accuracy obtained from the one-phase WBC classification as our goal is to improve the results obtained from that approach.

4. **Training revert:** Training the network incrementally facilitates the implementation of gradual testing, which in turn aids in the recognition of the categories that are responsible for the drop in the network performance. Lines 14 to 18 represent the proposed "training revert" procedure, which is only executed in case of encountering a category that contains confusable classes. In this procedure the network reverts to its state before training on the newly added category, i.e., the network reverts to the weights, training set, and output shape values before adding the category that caused the performance drop. Next, in order to handle the complexity associated with the confusable classes, the newly added category is undersampled and reduced to one class. The point of undersampling the category is to retain the training set balance. The undersampled category is then added as one class to the training set. The purpose of adding the category that contains confusable classes as one class is to alleviate the complexity of classifying the confusable classes. As the network will classify all

the instances that exhibit a high level of visual resemblance as one class, rather than different classes.

After the completion of the proposed enhanced incremental training algorithm, each category that includes confusable classes will be classified in a separate neural network in a second phase as shown in Figure 9. Training separate networks to classify categories that contain confusable classes will give the network a chance to better learn and specialize in the features of the confusable classes, and classify them more accurately. In the second phase, each neural network is trained on the corresponding category classes before subsampling.

For the neural networks of the second phase, multiple deep networks were experimented but the VGG16 was considered because it achieved the best results. The VGG16 classifier was connected to a shallow neural network that consists of a fully connected layer, a dropout layer and finally a softmax layer. The cross entropy loss was used for this task and the batch size, image resolution, and dropout rate hyperparameters were tuned for each of the second phase networks separately. Multiple combinations of augmentation techniques are also experimented for each of the networks to improve the results. The early stopping technique is utilized to stop the training and avoid overfitting.

Batch size tuning

A factor in improving the classification results in the context of peripheral blood cell classification is tuning the batch size hyperparameter [135]. A NN trains with weights W can be seen as an optimisation problem, where we aim to adjust W to reach the minimum possible value for the loss function F . This optimisation problem can be expressed as:

$$\arg \min_W F \tag{11}$$

In this work the mini-batch Gradient Descent algorithm is utilized. During training, at the j -th iteration, W is updated using the formula :

$$W_{j+1} = W_j - \alpha/mb\Delta W_j \quad (12)$$

where α is the learning rate and mb is the batch size [11]. It can be seen from equation 12 that the batch size can directly affect the value of the updated weights, and subsequently affect the value of the loss F that we aim to minimize. In order to tune the neural networks used for WBC classification, the deep networks trained using the enhanced incremental training method will be run on 4 commonly used batch sizes, to examine the effect of tuning this hyperparameter on the classifiers' predictions. Four widely used batch size values were chosen for the experiments: 16, 32, 64, and 128 which was the highest possible value given the available resources.

3.3.5 Considering Determinism

Neural networks use randomness by design. Many forms of randomness are utilized during training, such as, random weights initialization, random mini-batching, random augmentation. This randomness implies that different results will be obtained when training the exact same neural network on the exact same training data multiple times. Introducing determinism to deep networks means controlling the random processes to generate the same random numbers every time in order to guarantee reproducibility.

Since the proposed "enhanced incremental training" method includes training neural networks gradually, while assessing the difference between the evaluation metrics at different stages, considering training the neural networks while taking determinism into account is crucial. This is simply because if the network was non-deterministic, then a different neural network with different parameters will be produced at every stage.

In order to enforce determinism on the training process of all the classifiers considered in this work, technical solutions proposed by Keras [65] and Nvidia [39] were followed. In addition, all NNs were initialized with weights pretrained on the ImageNet dataset, which eliminated the randomness caused by the weight initialization factor. To the best of our knowledge, this is the first work to consider determinism with incremental training.

3.3.6 Evaluation

In order to evaluate the platelet and WBC classification approaches, cross-validation is used. As mentioned earlier, the datasets were partitioned into training, validation, and testing sets. Since determinism is considered in this work, the experiments will not be repeated, as the results will not change.

The final models that result from both platelet classification approaches, and both WBC classification approaches will be tested on the test set and evaluated using the precision, recall of each class. Moreover, each model will be evaluated in terms of the macro-average precision, the macro-average recall, and the accuracy [146].

3.3.7 Experimenting Enhanced Incremental Training with Non-medical Datasets

In order to test how well the proposed enhanced incremental training method interacts with other datasets, two non-medical public datasets available on Kaggle are utilized: humming birds dataset [60] and animals dataset [59].

The humming birds dataset consists of four classes: broad tailed females, broad tailed males, rufous females and an extra class that covers other objects such as bird feeders. In order to implement the proposed method, the dataset was divided into two categories: the females category which consists of both female breeds, and another category that consists of the remaining classes. This categorisation was made based on the visual resemblance

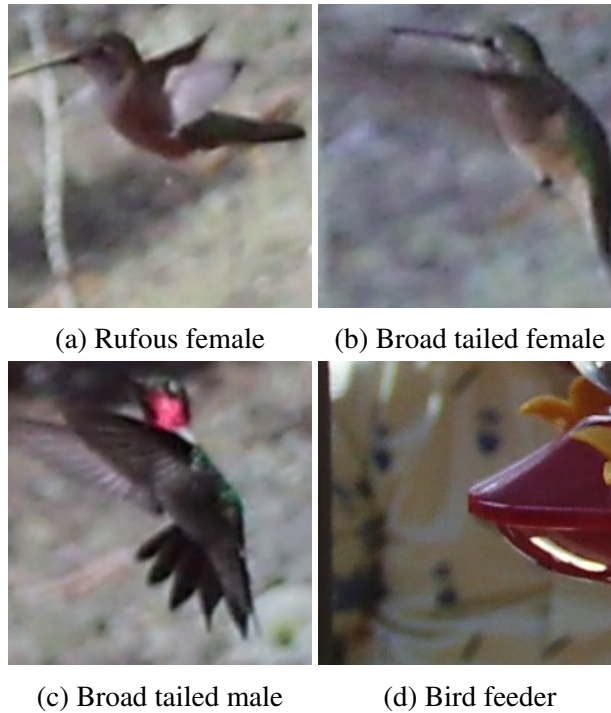


Figure 11: Samples from the humming bird dataset

between the broad tailed females. Figure 11 shows some samples from the humming birds dataset.

On the other hand, only a subset of the animals dataset is considered for this experiment. The seven considered classes are: orangutan, chimpanzee, foxes, wolves, snakes, zebra and turtles. In order to implement incremental training the classes are categorized into: the monkeys category which consists of the orangutan and chimpanzee classes, the canines category which consists of the wolf and fox classes, and finally, all the remaining classes were grouped into one category. The canines and monkeys categories were formed based on their visual resemblance, whereas the classes of the last category exhibit insignificant visual similarities. Figure 12 shows some samples from the monkeys and canines.

For each dataset the VGG16, ResNet50, Xception, Densenet121, MobileNetv2 neural networks are trained and tested and only the best performing network is then experimented using enhanced incremental training.

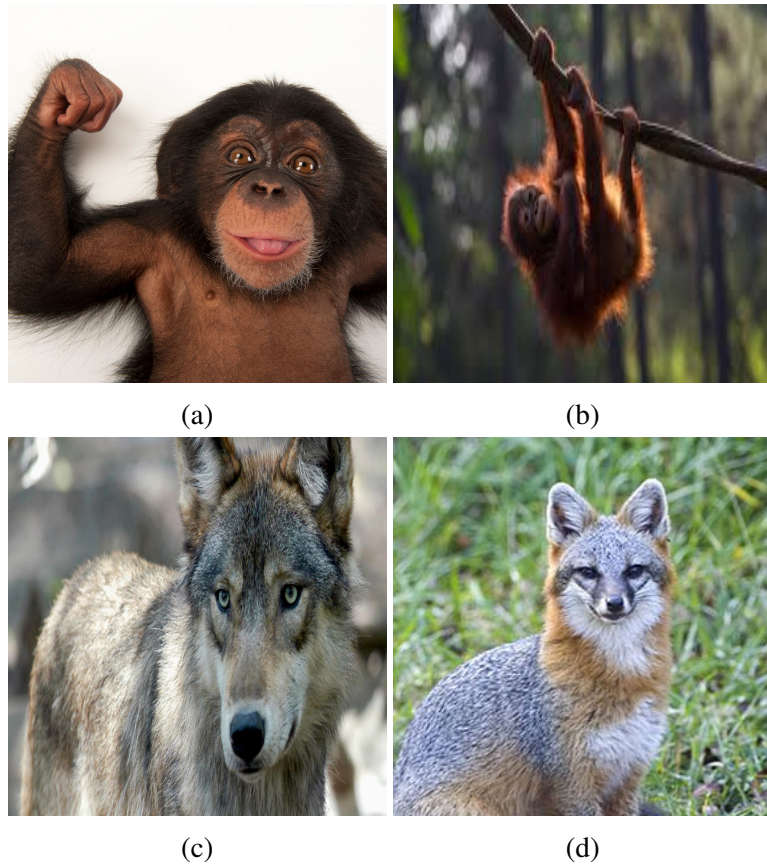


Figure 12: Visual resemblance in the monkeys and canines categories

3.4 Detection of Reticulocytes

In this study we aim to train a deep network to classify reticulocyte cells in human blood smears. The main challenge in this context is data scarcity. Hence, we propose a novel approach to overcome this challenge by using animal reticulocyte cell features as a solution to compensate for the deficiency of human data.

Transfer Learning (TL), is reusing the knowledge obtained from a source neural network in training target network(s) in order to improve their performance [147]. This technique is widely used in the literature and it has proven its effectiveness in different application domains. Moreover, utilizing TL was found to be better than initializing weights randomly [143].

In [33], a comparison between utilizing medical image datasets and general large-scale datasets as source for pretrained weights is conducted. The author reviewed many works that utilize both types of source datasets to train target networks that perform medical image analysis. No explicit relation or benefit was found between the type of the source (medical, or large-scale) and the testing results.

The authors of [63] conducted a set of experiments to investigate how ImageNet architectures and pretrained weights relate to performance on downstream medical tasks. In this work, 16 widely used convolutional neural networks were initialized to weights pretrained on the ImageNet dataset to train the target network to classify chest X-rays. The authors observed that ImageNet pretraining resulted in a statistically significant boost in performance.

3.4.1 Datasets

The datasets utilised in this work for the purpose of reticulocyte classification are:

CENPARMI human reticulocyte dataset

This study is a part of the Centre for Pattern Recognition and Machine Intelligence (CENPARMI)'s [3] research projects in the field of artificial intelligence. This dataset was collected for this study and consists of 2461 instances that belong to three main classes: RBCs, reticulocytes, and unknown background objects. Figure 16 (b) depicts the number of instances of each class in the dataset.

All smears were inspected under an oil immersion lens of a Nikon ECLIPSE E100 microscope [81] with a 1000X magnification. The microscope was connected to a 14 Mega Pixel digital microscope camera that captured the images of the blood smears. Next, an experienced hematologist annotated the target cells. This dataset is available for research purposes. Figure 13, Figure 14, and Figure 15 illustrate samples from the reticulocytes,

RBCs and background objects instances respectively.

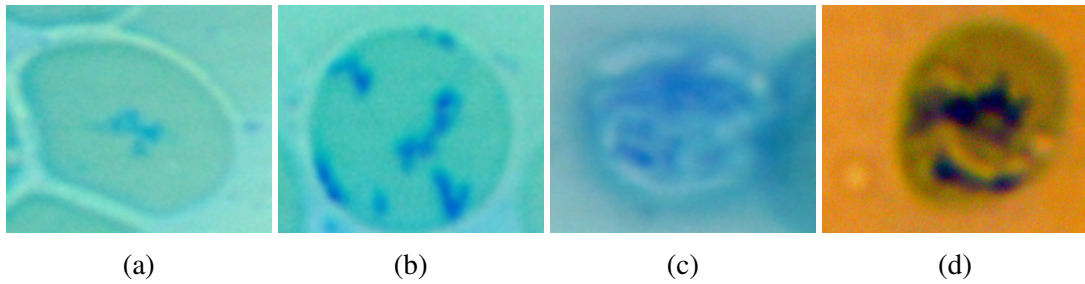


Figure 13: Human reticulocyte samples from the CENPARMI human reticulocyte dataset

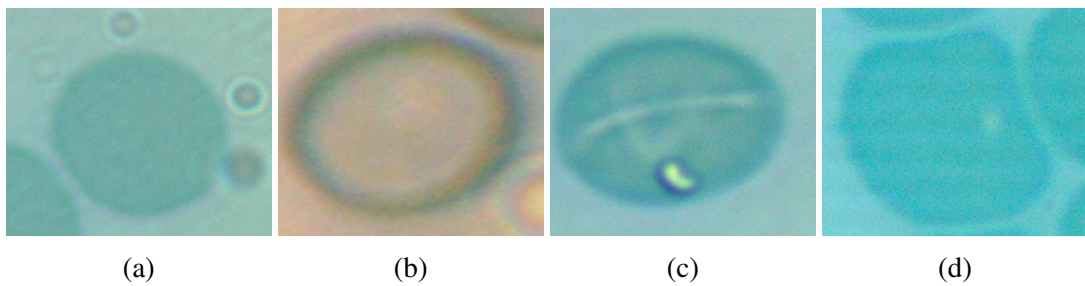


Figure 14: Human RBC samples from the CENPARMI human reticulocyte dataset

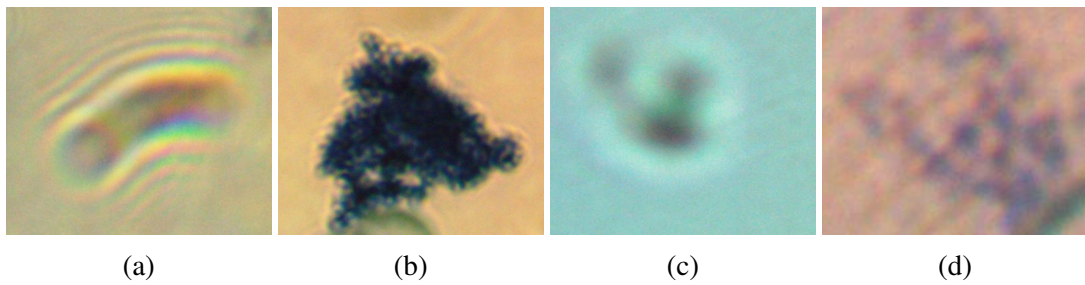


Figure 15: Samples of the background objects class from the CENPARMI human reticulocyte dataset

The feline reticulocytes dataset

This public dataset was collected by Vinicki et al. [128, 48] from two non-anemic male cats. This dataset consists of whole-smear instances that were taken with a Bresser microscope camera and a Samsung Galaxy S6 smartphone camera. The blood cells in this

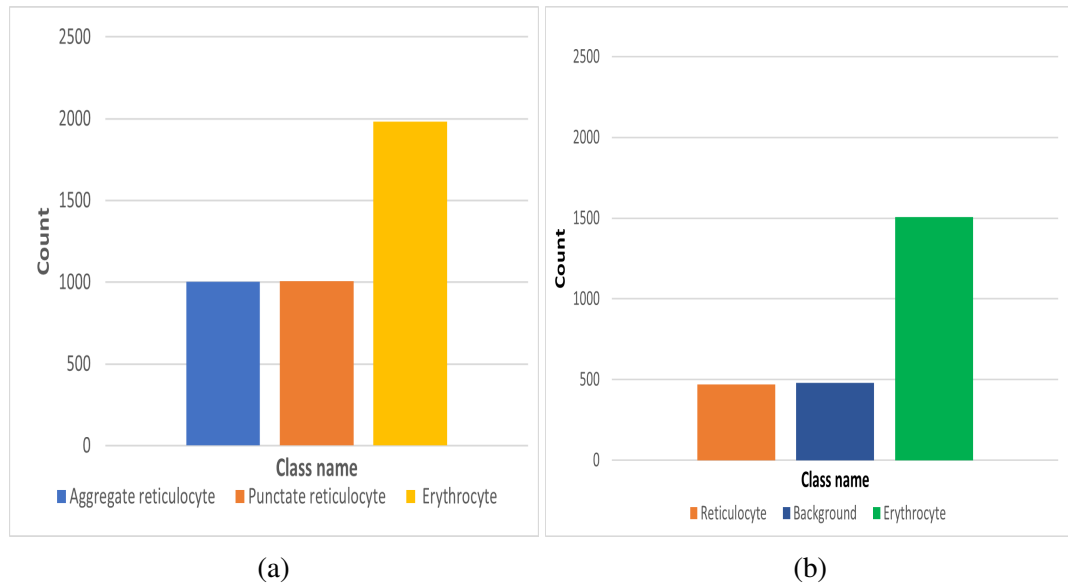


Figure 16: Instance counts of the feline and human reticulocyte datasets
a) The number of instances of each class in the feline reticulocyte dataset, and b) The number of instances of each class in the human reticulocyte dataset

dataset were annotated into three categories: aggregate reticulocytes, punctate reticulocytes and erythrocytes (RBCs). The open-sourced dataset consists of 1086 smear images and their corresponding Pascal VOC XML annotation files. Figure 16 (a) depicts the number of instances of each class, it also shows that the number RBCs is double the number of instances of the other two classes. Different cameras and settings were considered when the dataset was collected, therefore, it can be noted that reticulocyte cells significantly vary in size. Figure 17 depicts 2 examples of aggregate reticulocytes that vary in size in different instances. Finally, some cells are not annotated in the dataset. Figure 18 illustrates 2 examples where some cells were not considered in the annotation files, in the figure the cells outlined by green bounding boxes are not annotated in the original dataset.

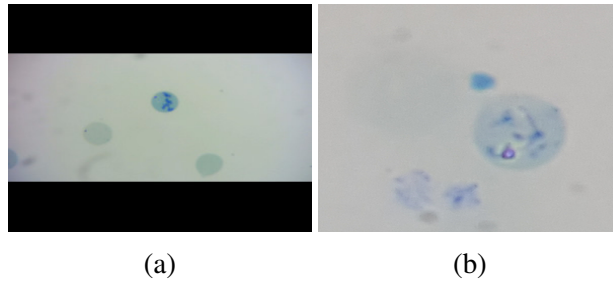


Figure 17: Different whole-slide instances where aggregate reticulocytes vary in size

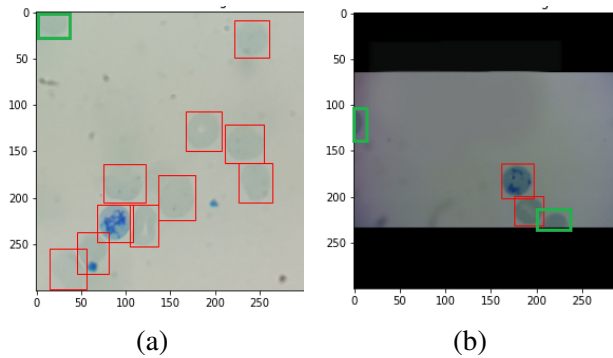


Figure 18: Different whole-slide instances where some cells are not annotated

The ImageNet dataset

This dataset [35, 52] consists of 14,197,122 images from different general categories, such as, animals, plants and instruments. This dataset has been widely used in the literature due to the high diversity of classes and the rich features it provides to neural networks.

3.4.2 Can Animal Cells be a Possible Solution for Medical Data Scarcity?

Data is the backbone of neural networks. In order to train a human reticulocyte classifier, a dataset that consists of a sufficient number of instances is needed. Due to the scarcity of public human reticulocyte datasets, we managed to collect a relatively small dataset, and hence we opted to utilize methods that handle small datasets rather than synthesizing reticulocyte data. However, more data is still needed in order to have a well-trained robust network.

In recent years, researchers have proposed some viable solutions to help neural networks learn and perform well with relatively small datasets. Some techniques that can help neural networks generalise on small datasets are: Data augmentation and transfer Learning. Transfer learning is transferring the knowledge from a source neural network to target network(s) in order to improve their performance [147].

Table 7: Transfer learning formulation

Source	Target
$DS = \{(x_S^i, y_S^i)\}_{i=1}^{ms}$	$DT = \{(x_T^j, y_T^j)\}_{j=1}^{mt}$
$\mathcal{X}_S, \mathcal{Y}_S$	$\mathcal{X}_T, \mathcal{Y}_T$
$\mathcal{F}_S \rightarrow P(Y x_S^i)$	$\mathcal{F}_T \rightarrow P(Y x_T^j)$

Table 7 shows transfer learning setup [144], in each transfer learning problem the source dataset DS comprises ms labeled instances, and a neural network F_S is optimized and trained to output the probabilities of each output class Y , given an input instance x_S^i , where $Y \in \mathcal{Y}_S$ and $x_S^i \in \mathcal{X}_S$. Moreover, \mathcal{Y}_S is the set of all possible output labels (classes) and \mathcal{X}_S is the input space. On the other hand, the target also consists of the same components as the source. By applying transfer learning we aim to tune F_S into F_T in order to accomplish the target task. In general, transfer learning can be applied either by retraining the base network on the target dataset to adapt its weights to perform the target task, or by using F_S without any change to extract features in order to perform predictions for the target task. The earlier technique is called fine-tuning, and the latter is called freezing.

In this work, three experimental setups will be implemented for the purpose of training a reticulocyte classifier, Figure 19 illustrates those setups. In all the experiments that employ TL, the target dataset DT , is the human reticulocyte dataset. Moreover, the set of possible output target classes Y_T consists of 3 classes; human reticulocytes, human RBCs, and the unknown background objects class. In those experiments, we aim to optimize the

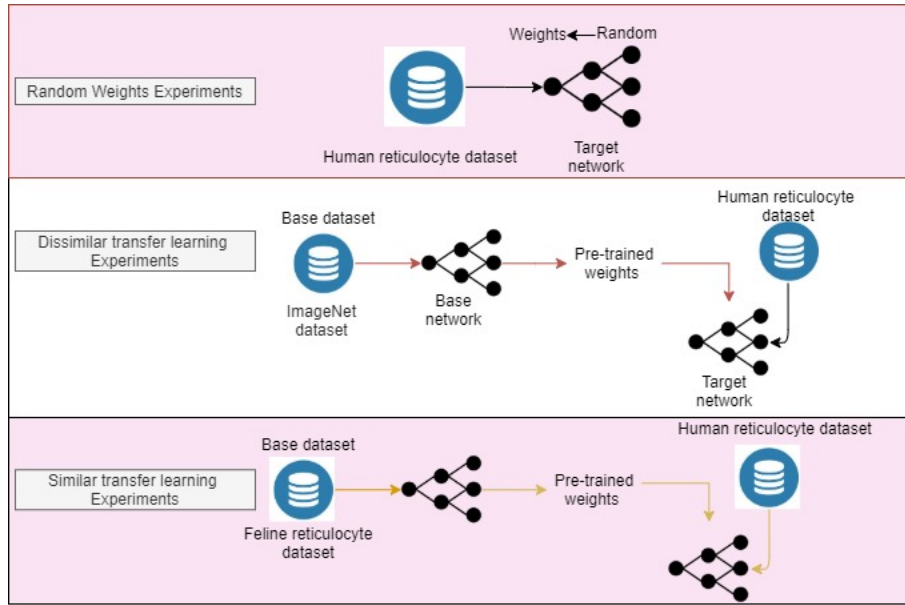


Figure 19: The proposed experiments

neural network F_T to maximize the conditional probability of the correct class. The next subsections list more details about each of the experimental setups described in Figure 19.

3.4.3 Transfer Learning Between Similar Datasets: Feline to Human Reticulocyte Transfer Learning Experiments

Human reticulocytes are round in shape and differ from RBCs in terms of size and morphology. Human RBCs are 6-8 μM whereas reticulocytes are 5-10 μM [25], moreover, RBC cells have a clear appearance whereas reticulocytes are distinguished by the reticular network of ribosomal RNA. On the other hand, felines have two types of reticulocytes; aggregate and punctate and both types differ in size and morphology. Feline punctate reticulocytes and mature RBCs are approximately 5.5-6.3 μM [46], whereas aggregate reticulocytes are larger. Moreover, punctate cells exhibit less reticular RNA remnants compared to aggregate cells.

The purpose of this set of experiments is to assess the feasibility of utilizing animal cell features as a solution to the scarcity of human data. In addition, the high visual similarity

of the feline and human reticulocytes could be very beneficial, as utilizing transfer learning between similar datasets provides the target network with pretrained domain specialized features.

Before conducting this experiment some operations were performed, in order to further bridge the gap between the source and the target dataset:

1. The cells labeled as aggregate or punctate in the feline dataset were merged into one class named "reticulocyte". We opted to perform this operation because both feline reticulocyte types map to human reticulocytes, hence merging them in one class could help the source network focus more on learning general features of reticulocytes rather than on learning the features of punctate and aggregate reticulocytes separately.
2. Due to the structure of the instances in the human dataset, the cells in the feline dataset were cropped in order to have one ROI (cell) showing in each instance.

According to the formulation in Table 7, in this experiment, $\mathcal{X}_S = \mathcal{X}_T$, since both input spaces are images. Moreover, after merging the feline reticulocyte classes the output category spaces will be very similar but yet not homogeneous since the source dataset lacks the "unknown background objects" class; $\mathcal{Y}_S \subset \mathcal{Y}_T$. Finally, the target neural network will be experimented by freezing the layers of F_S .

3.4.4 Transfer Learning Between Dissimilar Datasets: ImageNet to Human Reticulocyte Transfer Learning Experiments

In this experiment the ImageNet features will be transferred to classify human reticulocytes. The set of all possible output classes in the source, \mathcal{Y}_S , consists of hundreds of general classes. Hence, in this context, only the input spaces are equal, whereas $\mathcal{Y}_S \neq \mathcal{Y}_T$ and

$DS \neq DT$. The neural networks in this experiment are experimented by freezing the layers of F_S .

3.4.5 No Transfer Learning Experiments

In this experiment the neural network will be trained from scratch, i.e., without any prior knowledge, on the target reticulocyte dataset.

3.4.6 Dataset Size and Balance Effect

The 3 experimental setups described in subsections 3.4.3, 3.4.4 and 3.4.5 utilize the CENPARMI reticulocyte dataset to train the target neural networks. In this work, the dataset size and balance aspects will be experimented too; all the experiments in this work will be conducted twice:

1. Using all the instances of the CENPARMI dataset, which is highly imbalanced.
2. Using a balanced version of the dataset that consists of equal proportions of all 3 classes. In order to obtain the balanced version of the dataset, the RBCs class, which is the dominant class in the dataset, was subsampled to 500 instances.

In the full dataset, the number of instances is higher and the dataset is imbalanced due to the high number of RBCs compared to the other 2 classes, whereas, in the balanced version of the dataset the total number of instances is 1000 less, but it is balanced. By comparing the results of utilizing each version of the dataset we can assess the importance of more training instances over the balance of the dataset.

3.4.7 Technical Setup

The neural network utilized in this work is the VGG16 network, other deep networks such as the DenseNet, ResNet50 and VGG19 were experimented but the results were very poor

Table 8: The properties of the fully connected neural network utilized in the target networks

Framework	Keras 2.4.3
Loss function	Cross Entropy
Learning rate	.01
Weight decay	1e-6
Momentum	.9
Optimization algorithm	mini-batch Gradient Descent
Activation function	ReLU, Softmax
Fully connected layer size	1000
Batch size	32

hence we opted to only utilize the VGG16. In order to train the human reticulocyte classifier, the desired pretrained weights were first loaded. Next, the human reticulocyte dataset instances were resized to 100 x 100. Next, the resulting images were augmented by applying rotations, and horizontal and vertical flips. The augmentation techniques and the resizing factor were empirically set. Finally, the images were standardized using the Keras "preprocess input" function. Next, an extra fully connected layer and a softmax layer were attached to the network. The fully connected network utilized in the experiments is summarized in Table 8. Finally, the early stopping technique was used to stop the training and avoid overfitting.

The pretrained weights of the ImageNet were loaded from the Keras applications library. On the other hand, in order to get the pretrained weights of the feline cells, the VGG16 network was trained from scratch on the feline dataset, and the early stopping technique was used to stop the training. The network stopped training after 100 epochs with a training loss value of .057, accuracy of 97.8 %, macro F-score of 97.5 %, and macro precision of 97.6 %.

3.4.8 Evaluation Metrics

The stratified k-fold cross validation method will be used to train and evaluate the neural networks in this work. Stratified k-fold cross validation is an extension of the k-fold validation method where the ratio between the dataset classes is the same in each fold as it is in the full dataset. The number of folds in all experiments is 5 [17, 31].

Each experiment will be evaluated in terms of the average precision, average accuracy, and average F-score. The metrics are all averaged because they result from training the network 5 times. Moreover, the models will be evaluated and compared in terms of the following boost measures:

1. Pretraining boost: We study the effect of pretraining on the overall performance by defining the pretraining boost measure [63] as the difference between the macro F-score of the human reticulocyte classifier model initialized with pretrained weights (on the ImageNet dataset or the feline reticulocytes dataset) and the macro F-score of the classifier without pretraining, as shown in Equation 13.

$$\text{Pretraining boost} = Fscore_{DS} - Fscore_{pretraining} \quad (13)$$

2. Datasets similarity boost: Since the feasibility of utilizing animal cells is the focus of this study, a boost measure that studies the effect of datasets similarity on the results is defined as the difference between the macro F-score of the human reticulocyte classifier initialized with weights pretrained on the feline reticulocytes and the macro F-score of its counterpart pretrained on the ImageNet dataset, as shown in Equation 14. It is worth mentioning that the sign of the result of this equation is crucial, as it indicates how it benefits the feline cells weights compared to the ImageNet pretrained weights.

$$\text{Similarity boost} = Fscore_{Feline} - Fscore_{ImageNet} \quad (14)$$

3. Dataset size boost: Two variations of the CENPARMI dataset are utilized in this work as explained in Section 3.4.6. In this evaluation metric we aim to assess which dataset is more beneficial to the models' performance. This metric is defined as the difference between the macro F-score of each of the three human reticulocyte classifiers trained on the entire imbalanced dataset and the macro F-score of its counterpart trained on the balanced dataset, as shown in Equation 15.

$$\textit{Similarity boost} = Fscore_{imbalancedDS} - Fscore_{balancedDS} \quad (15)$$

Chapter 4

Results

4.1 Results of Synthetic Blood Smears Generation Using Locality Sensitive Hashing

In this section, the results of the proposed synthetic dataset generation method explained in Section 3.2 and the medical assessment method mentioned in Section 3.2.3, are presented.

4.1.1 Synthetic Dataset Generation Results

To conduct this set of experiments, Algorithm 1 was implemented and executed. During the execution, it was noted that the initial set of cells were often pasted without any overlapping and without the need to retrieve neighbors due to space availability. Hence, to further decrease the cost of execution, the initial 60 cells were pasted without neighbor retrieval. The threshold 60 was chosen empirically.

To further tune our algorithm, the assignment of the threshold parameter T from equation 7 was adjusted; T was assigned to a random value from an acceptable range that reflects the real distribution pattern, instead of setting it to a constant value throughout the execution. With this improvement, the synthetic blood smear instances reflected more realistic

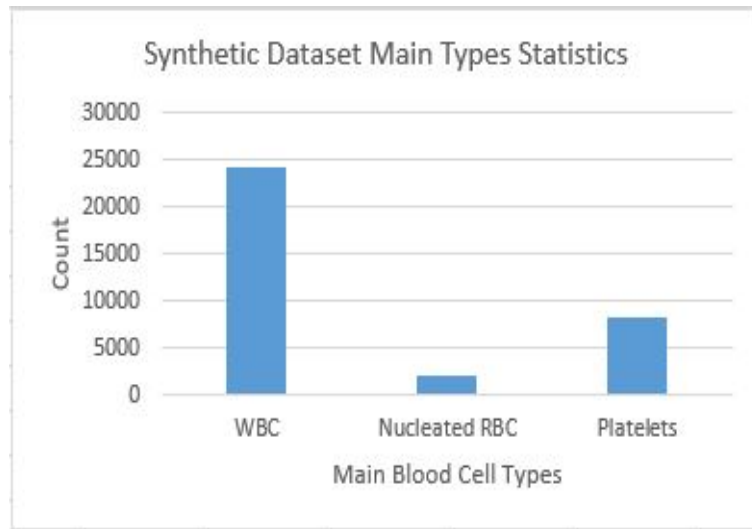


Figure 20: Synthetic dataset main types statistics

scenarios, For example, some RBCs were stacked in many smears forming patterns similar to the well-known Rouleaux formation. The resulting dataset consists of 2500 blood smears, the dataset is balanced in terms of subtypes and each subtype has around 2000 total instances. Figures 20 and 21 depict more statistical details of the synthetic dataset. Figure 21 demonstrates the dataset balance.

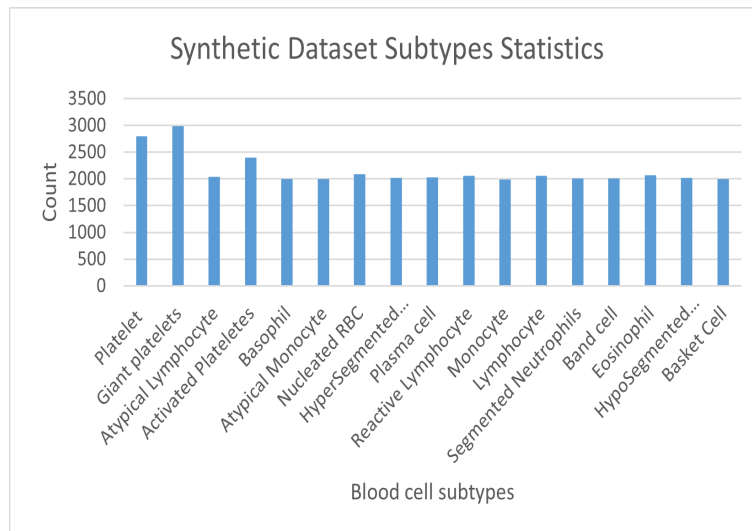


Figure 21: Synthetic dataset Subtypes Statistics

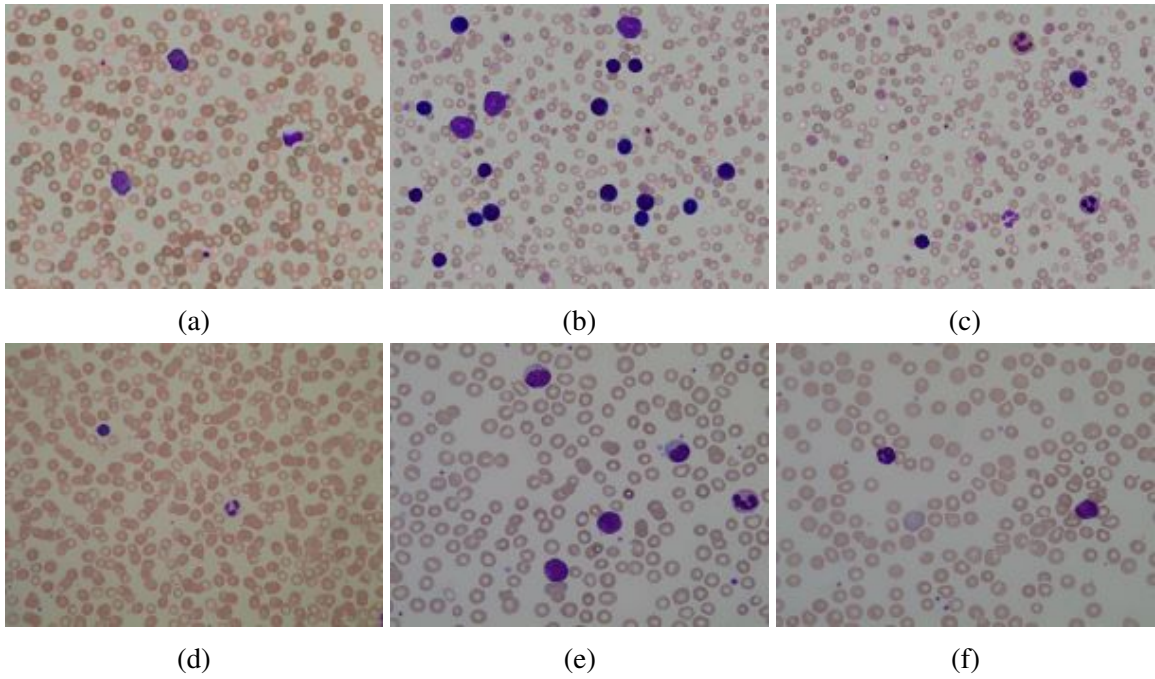


Figure 22: Instances from real and synthetic blood smears

Figure 22 illustrates some samples from the synthetic dataset; smears (a) to (c) belong to the synthetic dataset, while the rest belong to ALL-IDB1 dataset.

4.1.2 Medical Assessment Results

Figure 23 demonstrates the results of the questionnaire, where the responses to each question are represented by a column in the chart. In each column, similar answers were grouped by color, and the blue color is used for answers that are identical to the ground truth. Figure 23 shows clearly that all responses were identical and met the expected answers in all questions except for questions 6 and 10. The sixth question is one of the questions that were used to verify factor 3. It showed a synthetic blood smear and asked the respondent to classify the platelets that appeared in the smear. The platelet is sub-classified as a normal platelet in the second set of annotations of the proposed dataset; however, one of the hematologists classified it as a giant platelet. It is worth mentioning that it is common that medical experts have different opinions about medical data annotation [97].

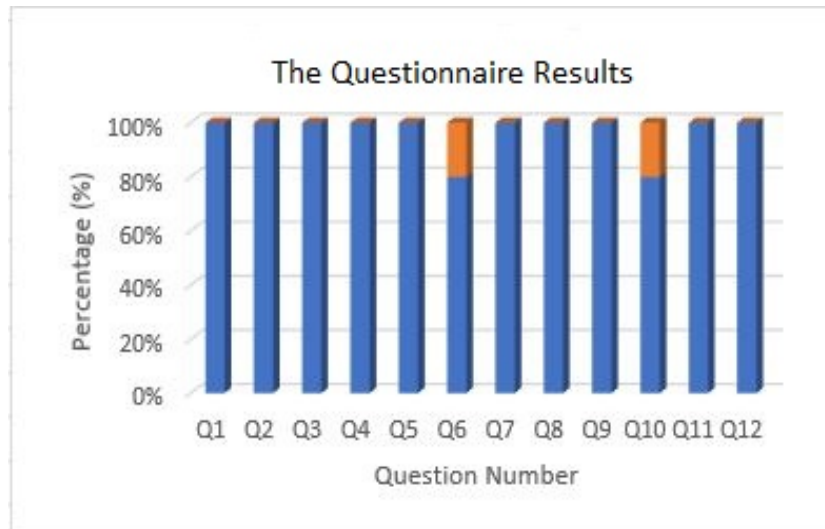


Figure 23: The results of the questionnaire

On the other hand, the tenth question tested the ratio of the blood cell sizes in the smear. One of the hematologists expressed that some cells appeared a little larger than it should be, where the other 4 hematologists believed that the ratios were appropriate. Factor 5 verifies the correctness of the proposed algorithm and due to its importance, the haematologists were asked to assess all the blood smears that appeared in the questionnaire in terms of the quality of cell distribution and occlusion. All respondents approved that all the slides shown in the questionnaire were thin and all cells were distributed in a natural pattern with a ratio of occlusion that meets the standard in real blood smears.

The proposed dataset proved its usefulness in the context of PBS analysis, it also provides the following benefits:

1. The synthetic dataset is not subject to any privacy constraints since it does not belong to real people.
2. Rare subtypes, like plasma cells, are sufficiently present in the dataset.
3. The dataset was automatically annotated for blood cell subtypes and a wide range of morphological abnormalities, without any extra efforts of medical experts.

4. The dataset is balanced in terms of the 15 subtypes covered in this study.

4.2 Results of Phase 1: Classification of Main Blood Cell Types using YOLO Deep Network

The results of the first phase of classification mentioned in Section 3.3.1 are presented in this section. A Tiny YOLOv3 network [100] was first trained on the synthetic dataset, but due to its relatively small size (13 convolutional layers) and the high complexity of the context, the network achieved a low Mean Average Precision (MAP) score of approximately 40%. To improve the results, a larger network, namely YOLOv2 [99], that consists of 23 convolutional layers, was trained on the dataset. This network achieved a much better MAP score of 97.59%.

To further improve this result, the network input resolution was randomly resized every 10 batches during training, this regime which was proposed in [99] works like data augmentation and helps the network learn to better generalise. By exposing the network to randomness, the MAP score was improved to 98.72%. Both Nucleated RBCs and WBCs classes were classified with 100% Average Precision (AP), while Platelets scored 96.4%. In order to better understand the wrong platelet predictions made by the network, we further discussed the platelet results with a hematologist who suggested that the majority of the errors resulted from confusing the network between platelet cells and stain residues that some times appear on smears. Hence, in order to handle this issue, platelet cells are further examined against fixed ranges in the second classification phase which is explained in more detail in Section 3.3.3. The utilized neural network ran with a momentum value of 0.9, learning rate of 0.001, and network input resolution of 800 as an initial resolution. Table 9 summarizes the blood cell classification experiments and results.

Table 9: Blood cell classification results using YOLO

Deep Network	Mean average precision
Tiny YOLOv3	40%
YOLOv2 without random resizing	97.59%
YOLOv2 with random resizing	98.72%

4.3 Phase 2: Results of Blood Cell Subtypes Classification

4.3.1 Platelets classification results

Two approaches are proposed in this work for platelets classification; one-phase classification, and two-phase classification. The results of both approaches, are presented in the next two sub sections.

One-phase Classification

In this approach, a single neural network was trained to classify all 3 platelet types. A VGG19 network was trained first, but it only achieved 78% test accuracy. In the second experiment, a ResNet50 was trained to classify all 3 classes. The trained ResNet50 achieved a test accuracy score of 82.67%.

Two-phase Classification

In this approach platelets are classified in two phases. In the first phase, a VGG19 network was trained and tested on both classes. This network achieved an accuracy score of 98.4%. Classifying both confusable classes (normal and giant platelets) as one class reduced the level of confusion, and the rate of misclassifications which resulted in a better accuracy.

In the second experiment, a ResNet50 was also trained and tested on both classes. This network test result surpassed all the previous networks trained in this study for platelet classification. An accuracy test score of 100% was achieved by this network, hence, we

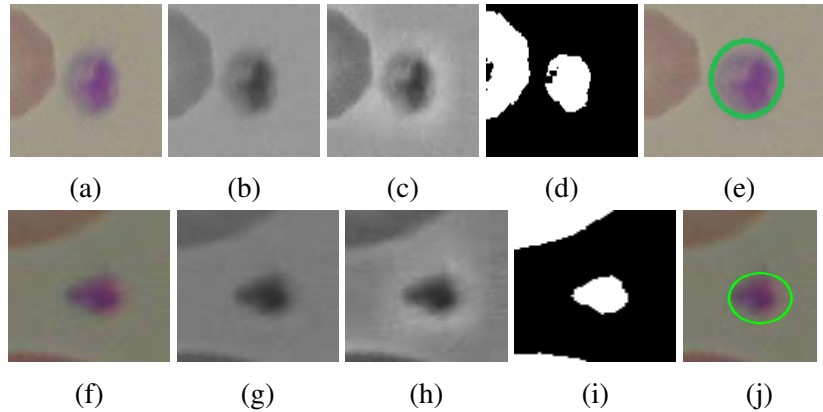


Figure 24: Sample results from the second phase of classification.

(a) to (e): test instance 1 classification process, (f) to (j): test instance 2 classification process

(a),(f): the original image; (b),(g): the extracted green channel; (c),(h): the resulting image after applying CLAHE; (d),(i): the resulting image after applying Otsu's thresholding; (e),(j) the platelet is outlined after applying Hough Circle Transform

Table 10: Confusion matrix resulted from the one-phase classification approach

		Predicted			
	Activated	Giant	Normal	Recall	
Activated	50%	50%	0%	50%	
Giant	0%	76%	24%	76%	
Normal	1%	16%	83%	83%	
Precision	98%	54%	78%		
Accuracy	82.67% (CI 95% 79.11, 86.23)				
Macro-average recall	69.6%				
Macro-average precision	76.6%				

chose to only consider it for the second phase. In the second phase, the instances classified as "other platelets" went through a set of image processing techniques and filters, as illustrated in Figure 8.

Figure 24 illustrates the classification process of two test samples in the second phase, each row in the figure represents one test sample. Only a few instances were misclassified in the second phase, leading to an overall accuracy score of 98.6%.

Moreover, in order to have a better understanding of the obtained results, Table 10 shows the confusion matrix of the one-phase ResNet50 model, and Table 11 shows the

Table 11: Confusion matrix resulted from the two-phase classification approach

		Predicted		
	Activated	Giant	Normal	Recall
Activated	100%	0%	0%	100%
Giant	0%	94%	6%	94%
Normal	0%	1%	99%	99%
Precision	100%	99%	94%	
Accuracy	98.6% (CI 95% 97.5, 99.7)			
Macro-average recall	97.6%			
Macro-average precision	97.6%			

confusion matrix of the two-phase ResNet50. As can be seen in Table 10, one reason for the low accuracy in the one-phase approach was the high visual similarity between giant and normal platelets. For instance, 24% of the giant platelets were confused with platelets, and 16% of the platelets were confused with giant platelets, which negatively affected the overall classification accuracy. Hence, combining both classes in the "other platelets" class aided in improving the classification results. It is also worth mentioning that the activated platelets classification results were also another reason for the low accuracy of the one phase approach but investigating the reason is beyond the scope of this study, as this class is not considered as confusable, hence it is not related to the study hypothesis. On the contrary, in the two-phase approach only 6% of the giant platelets were confused with platelets, and 1% of the platelets were confused with the giant platelets.

4.3.2 WBC classification Results

Two approaches are proposed in this work for WBC classification: 1) one phase WBC classification, and 2) deep neural networks pipeline based on enhanced incremental training.

One phase WBC classification

In this approach all WBC instances from all 12 classes are classified by a single VGG19 neural network.

Table 12: One phase WBC Classification Results

Batch Size	Accuracy (%)
16	33.7
32	39.1
64	44
128	61.5

The results obtained are listed in Table 12. The poorest result was obtained with the smallest batch size. On the other hand, the results kept improving as the batch size value increased. However, the best achieved result was 61.5%.

Deep neural networks pipeline based on enhanced incremental training

In this approach, WBCs are classified into two phases. During the implementation of the proposed method, Cat_0 was added to the training set first, as according to the medical expert, the odds of having a confusable class in this category is very low. Hence, training this category first helps to have a solid basis of trained weights before training the network on the rest of the categories which contain classes with high visual resemblance.

In the proposed method, the network is tested on the validation set during training and only the final model that results from training the network on all categories is tested on the test set. Four neural networks were trained using enhanced incremental training, each classifier only differed in the batch size value.

Table 13 lists the classes that were included in each stage. Figure 26 depicts all training stages for each of the four neural networks, along with the validation accuracy results obtained from each stage. It is clear from the figure that all classifiers followed the same pattern during the training process. Moreover, Figure 25 depicts the workflow of executing the proposed training scheme. Some highlights from Table 13, Figures 25 and 26 regarding all four classifiers are:

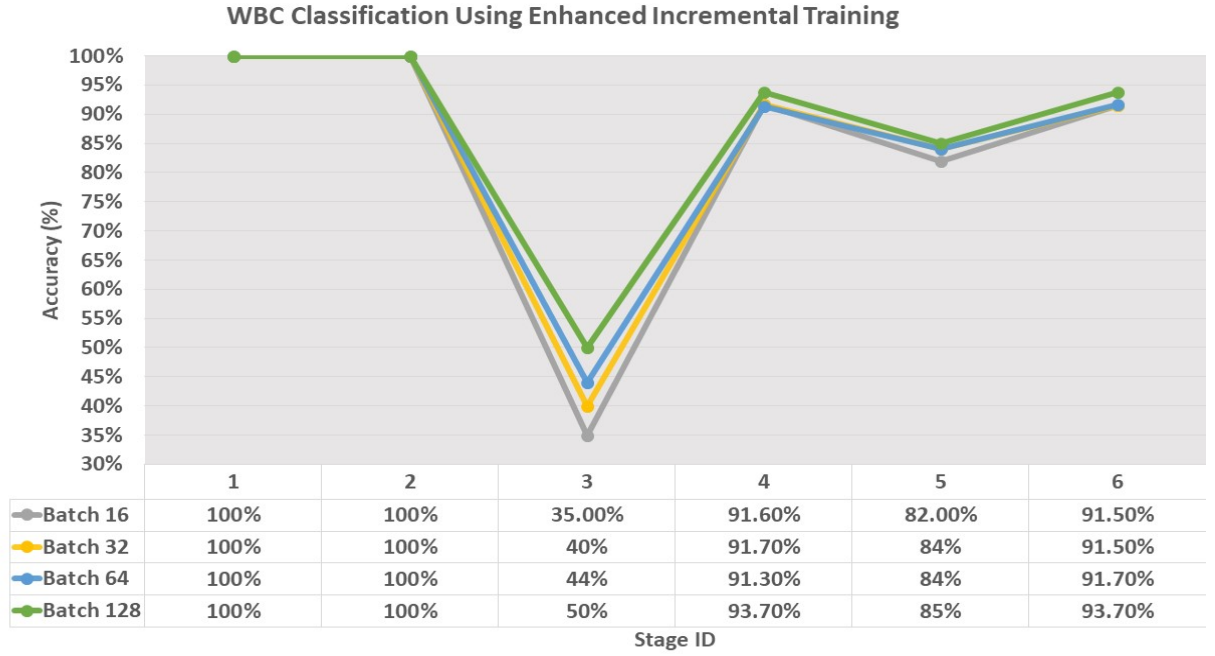


Figure 25: The workflow of training the four neural networks using enhanced incremental training and the accuracy results obtained from each stage using the validation set

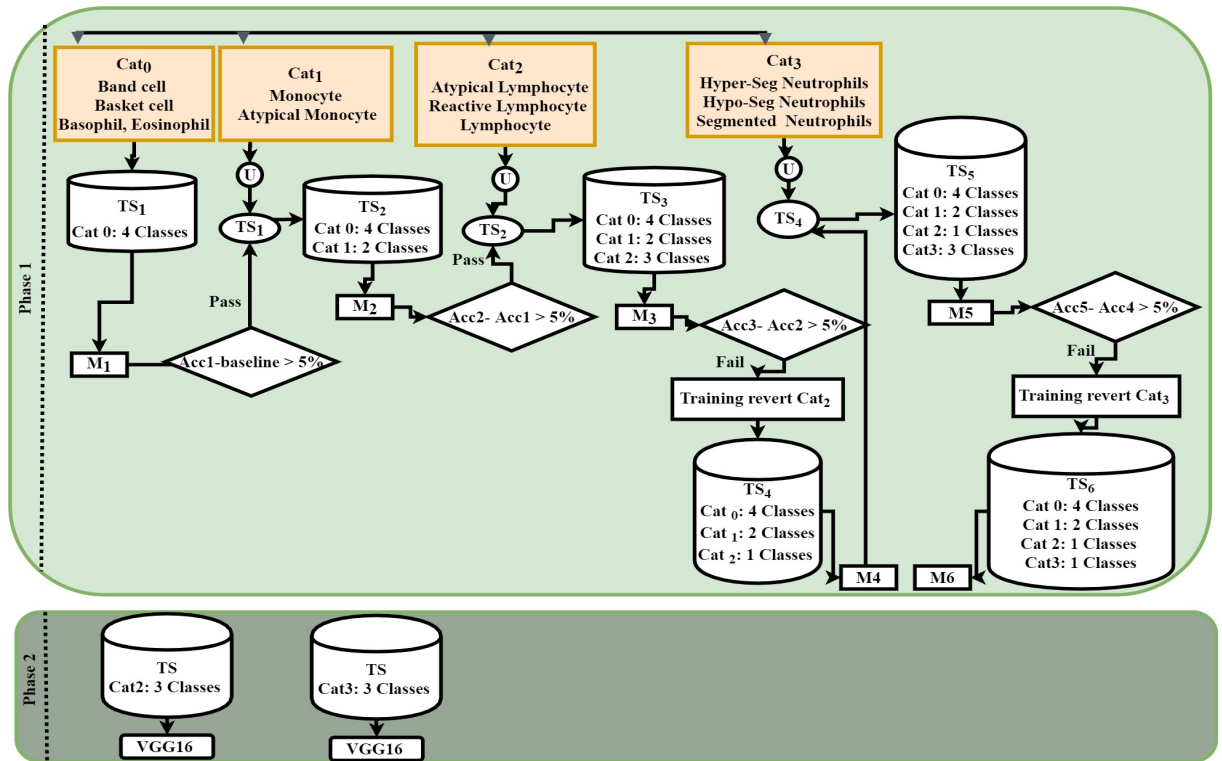


Figure 26: The workflow of training using the proposed training scheme

Table 13: Classes included in each stage

Category	Class Name	Stage ID					
		1	2	3	4	5	6
Cat_0	Band cell	✓	✓	✓	✓	✓	✓
Cat_0	Basket cell	✓	✓	✓	✓	✓	✓
Cat_0	Basophil	✓	✓	✓	✓	✓	✓
Cat_0	Eosinophil	✓	✓	✓	✓	✓	✓
Cat_1	Monocyte		✓	✓	✓	✓	✓
Cat_1	Atypical Monocyte		✓	✓	✓	✓	✓
Cat_2	Reactive Lymphocyte			✓	✓	✓	✓
Cat_2	Atypical Lymphocyte			✓			
Cat_2	Lymphocyte			✓			
Cat_3	Hyper-Segmented Neutrophils					✓	✓
Cat_3	Hypo-Segmented Neutrophils					✓	
Cat_3	Segmented Neutrophils					✓	

- The validation accuracy was 100% after adding Cat_0 and Cat_1 . Despite our expectations that the accuracy will drop after adding Cat_1 due to the visual similarity of the Monocyte and atypical Monocyte cells, the model actually kept a high accuracy score.
- In Stage 3, classes of category Cat_2 were added and the accuracy score dropped by more than 5%. Hence, Cat_2 was identified as a category that comprises confusable classes. The confusion matrix at this point was analyzed, and it was observed that the drop in the accuracy was mainly due to the high level of confusion between the instances of the three lymphocyte category classes, which supports the assumption that visual resemblance can be a key factor in negatively affecting the classification results.
- In Stage 4, the "training revert" procedure was executed; the weights and the training set were reverted to their values just after training category Cat_1 , i.e., Stage 2. All Cat_2 classes were then randomly undersampled and reduced to one class. The network was trained again with the reduced category and the accuracy score increased.

Table 14: Test results of the WBC classifiers trained using the proposed method

Batch Size	Accuracy (%)
16	91
32	91.3
64	92.2
128	95

- The accuracy dropped again by more than 5% after adding the last category, i.e. Cat_3 . Hence, the "training revert" procedure was executed again, and the entire category was reduced to one class and added to the training set. The accuracy score increased and the final validation accuracy was higher than 90% for all batch sizes.
- All classifiers that were trained on different batch sizes showed the same pattern during the training stages, hence the observations above apply to all four classifiers.
- Comparing the accuracies of the four classifiers that were trained on different batch sizes, it is noted that increasing the batch size increased the accuracy.

All the results shown in Figure 25 are the validation set results, the final model resulted from Stage 6 is the only one that was tested on the test set. The results of the test set were very similar to the validation results. The highest test accuracy result obtained was 95% achieved by setting the batch size value to 128. Table 14 lists all test accuracy results.

In the second phase to classify Cat_2 and Cat_3 , two VGG16 networks were utilized; the first one was trained to classify the Lymphocyte category into, Lymphocyte, reactive Lymphocyte, and atypical Lymphocyte. The second network was trained to classify the Neutrophils category into hypo-segmented Neutrophils, hyper-segmented Neutrophils, and segmented Neutrophils.

The best Lymphocyte classification results were obtained with a batch size value of 128. Moreover, the literature reveals that the size of the input image of a neural network aids in improving its performance [74, 12]. Hence, 10 different resolutions ranging from

200 to 600 were experimented: {200, 250, 300, ..., 600}, where 200 is the original image size emitted from the main classifier in phase 1, and 600 is the highest value possible with the available hardware. The lowest accuracy was 92%, scored with the lowest resolution. Moreover, the accuracy kept improving as the input resolution value increased. Hence, all instances were upscaled to 600 x 600.

On the other hand, to classify the Neutrophils category, another VGG16 classifier was tuned to upscale all instances to 400 x 400, this input resolution was tuned using the same procedure followed in the Lymphocyte category classifier. The lowest accuracy score was 95%, achieved with the lowest resolution, and the accuracy score was 100% for all resolution values higher than 350. It is worth mentioning that only the best classifier from phase one was tested with the classifiers from phase two, and both networks in phase two achieved 100% accuracy.

To sum up, during the test phase every test instance will be first fed to the main VGG19 neural network trained in the first phase, the instances that are classified as either Lymphocyte category or Neutrophil category are further emitted to the networks in the second phase for more detailed classification.

Moreover, in order to have a better understanding of the obtained WBC classification results, the confusion matrices of the utilized methods are shown in Figures 27 and 28. Figure 27 shows the 12x12 normalized confusion matrix of the top model trained using the one phase approach on the test set, and Figure 28 shows the 8x8 normalized confusion matrix of the top model trained using the proposed approach when applied to the test set. The latter is smaller in size because two out of the four categories were subsampled during training.

The following insights can be observed about each of the four categories from the confusion matrices:

1. Classes of Cat_0 :

	Predicted												Recall
	A.Lymphocyte	A.Monocyte	Band	Basket	Basophil	Eosinophil	Hyper S Neutrophil	Hypo S Neutrophil	Lymphocyte	Monocyte	R. Lymphocyte	S.Neutrophil	
A. Lymphocyte	57%	0	0	0	0	0	0	14%	15%	14%	0	0	57%
A.Monocyte	0	100%	0	0	0	0	0	0	0	0	0	0	100%
Band Cell	0	0	100	0	0	0	0	0	0	0	0	0	100%
Basket Cell	2%	0	0	94%	0	0	0	0	4%	0	0	0	94%
Basophil	0	0	0	0	100%	0	0	0	0	0	0	0	100%
Eosinophil	0	0	0	0	0	100%	0	0	0	0	0	0	100%
Hyper S Neutrophils	0	0	0	0	0	0	86%	0	0	0	0	14%	86%
Hypo S Neutrophils	0	0	0	0	0	0	0	100%	0	0	0	0	100%
Lymphocyte	37%	0	0	7%	0	0	0	1%	53%	1%	1%	0	53%
Monocyte	0	0	0	0	0	0	12.5%	0	0	75%	0	12.5%	75%
R. Lymphocyte	17%	0	0	0	0	0	0	0	66%	0	17%	0	17%
S.Neutrophils	0	0	0	9%	0	0	34%	0	0	0	0	57%	57%
Precision	50%	100%	100%	85%	100%	100%	65%	87%	38%	83%	94%	68%	
Accuracy	61.5%	Macro-average recall					78.25%	Macro-average precision				76.6%	

Figure 27: The normalized confusion matrix of the top model trained using the one phase WBC classification method

	Predicted									Recall
	A.Monocyte	Band Cell	Basket Cell	Basophil	Eosinophil	Lymphocyte Cat	Monocyte	Neutrophils Cat		
A.Monocyte	100%	0	0	0	0	0	0	0	0	100%
Band Cell	0	100%	0	0	0	0	0	0	0	100%
Basket Cell	0	0	95%	0	0	5%	0	0	0	95%
Basophil	0	0	0	100%	0	0	0	0	0	100%
Eosinophil	0	0	0	0	100%	0	0	0	0	100%
Lymphocyte Cat	0	0	4%	0	0	96%	0	0	0	96%
Monocyte	0	0	0	0	0	12%	88%	0	0	88%
Neutrophils Cat	0	4%	5%	0	0	0	0	91%	0	91%
Precision	100%	71.43%	84%	100%	100%	98.77%	100%	100%		
Accuracy	95%	Macro-average recall			96.25%	Macro-average precision			94.27%	

Figure 28: The normalized confusion matrix of the top model trained using the proposed method

- (a) The Recall values for the Band cells, Basophils, Eosinophils and Basket cells were approximately the same in both experiments.
- (b) The precision values for the Basophils, Eosinophils and Basket cells were approximately the same in both experiments. It is also noted that the precision score of the Band cells decreased in the model trained with the proposed method.

2. Classes of Cat_1 :

- (a) The recall score of the atypical Monocyte class remained 100% in both experiments, whereas, the proposed method improved the recall of the Monocyte class from 75% to 88%.
- (b) The precision score of the atypical Monocyte class remained 100% in both experiments. On the other hand, the proposed method improved the Monocyte class precision from 83% to 100%.

3. Classes of Cat_2 : There is no exact class-to-class comparison for this category because the classes of this category were blended into one class during the execution of the proposed method. The classes of this category were highly confused with each other in the one-phase approach, for example:

- (a) 37% of the Lymphocytes were incorrectly classified as atypical ones.
- (b) 15% of the atypical Lymphocytes were misclassified as Lymphocytes.
- (c) 17% of the reactive Lymphocytes were incorrectly classified as atypical Lymphocytes. Moreover, 66% of the reactive Lymphocytes were misclassified as Lymphocytes.

The recall and precision ratios of the Lymphocyte category in the proposed method were 96%, and 98.7% respectively.

4. Classes of Cat_3 : The following observations can be made from the one phase approach:
 - (a) 14% of the hyper-segmented Neutrophils were incorrectly classified as segmented Neutrophils.
 - (b) 34% of the segmented Neutrophils were misclassified as hyper-segmented Neutrophils.
5. The accuracy, macro average recall, and the macro average precision of the proposed method are 95%, 96.25%, and 94.27% respectively. which is higher than the results obtained from the one phase approach.

After the proposed system finishes classifying the cells in a blood smear, cells are grouped based on their types and then counted. Overall, it can be seen that the proposed method has improved the quality of the classification results. A drawback of utilizing the proposed method is that it needed 3 neural networks to complete the task, compared to one neural network in the one phase approach. But it is worth mentioning that automatic PBS analysis is not a real time application, and like other medical applications, the correctness of the results is a high priority.

The proposed method can be utilized in other application domains where confusable classes are thought to be the reason behind low classification results. Moreover, the proposed method was utilized on top of pretrained models available in an open source framework i.e. Keras. This indicates that the method can be easily reproduced and deployed by other researchers.

Table 15: Comparison of the results of WBC and platelet classification using our proposed system with existing works in the literature in terms of Accuracy

Method	Accuracy (%)
Our proposed work (WBCs & platelets)	96.7 (CI 95% 96.69,96.709)
Wang et al. [135] (SSD 300x 300)	75.1 (CI 95% 72.1,78.1)
Qin et al. [91] (shallow residual network)	60 (CI 95% 57.3,62.7)

4.3.3 Comparison with existing works

A comparison of our proposed method with the works of Wang et al. [135] and the work of Qin et al. [91] is shown in Table 15. In order to do the comparison, the overall accuracy score of the top platelet and WBC classification models was considered. Next, the entire training set (i.e. platelets and WBCs) was fed to the networks reproduced from both [135, 91]. The results in the table show the superiority of our proposed model. It is also noted that the results of the models from the literature were relatively close to the accuracy scores obtained from implementing the one-phase platelet classification and one phase WBC classification.

The method proposed in [135] scored an overall accuracy of 90.09% in the original paper, but it only achieved an accuracy score of 75.1% when trained and tested on the datasets of this study. One possible reason for this is that the method in the original paper was trained on a dataset that was only annotated for 11 types of platelets and WBCs, and this combination did not comprise many confusable classes which made the classification task easier. Moreover, the method reproduced from [91] scored the lowest accuracy, possibly because of its shallow structure which did not comply to the complexity level of the problem .

4.3.4 Results of Experimenting Enhanced Incremental Training with Non-medical Datasets

Two sets of experiments were conducted using each dataset. The first set of experiments were conducted using a set of widely used neural networks without utilizing enhanced incremental training. Next, in the second set of experiments the proposed method was utilized. Finally the results of both experiment sets are compared in order to evaluate the effectiveness of the proposed method with non-medical datasets.

Humming Birds Classification Results

In the first set of experiments, the entire dataset was fed to five widely used neural networks namely: VGG16, ResNet50, Xception, Densenet121, and MobileNet. The best performing network was the Xception network with a macro f-score of 87%, a macro average precision score of 89%, and a macro average recall of 88%. Those results were obtained with a batch size of 32. In order to further tune the batch size, the Xception network was trained again with a batch size of 8, 4, and 2. It is worth mentioning that 32 was the highest possible batch size given the available resources. Better results were obtained when the batch sizes was set to 8, where the test macro f-score, macro average precision and macro average recall values were 94.9%, 95%, and 95%, respectively. Hence, only the Xception network with a batch size of 8 and an input network resolution of 300, was considered for the enhanced incremental training experiment.

In the second set of experiments, the Xception network was first trained on the broad tailed males and the other objects classes. The validation macro f-score, macro average precision and macro average recall values of the resulting model were all 100%. Next, in the second stage, the females category, which consists of the broad tailed female and Rufous female classes, was added. The validation macro f-score, macro average precision and macro average recall values of the resulting model were all 95%. Since the evaluation

metrics dropped, the females category was considered confusing and the "training revert" procedure was executed. During the execution of the training revert procedure, the network was trained again on a training set that comprises of the first category and an undersampled females category that comprises of one class. After the execution of the training revert, the validation evaluation metrics improved again and the macro f-measure was 100%.

Finally, an extra neural network is needed to classify the female classes. Hence, the following networks were experimented: VGG16, ResNet50, Xception, Densenet121, and MobileNet. The Xception net was the best performing network with a batch size of 8 and an input resolution value of 300. Multiple augmentation techniques were also utilized, such as, horizontal flipping, rotation, and width shifting. This network was able to classify the female classes with a test macro f-score of 100%. Comparing the results of the experiment sets shows that the classification results improved when the proposed method was applied.

Animals Classification Results

In the first set of experiments, the same group of deep classifiers that were employed in the humming birds experiments were trained on the entire dataset. Some classifiers exhibited poor performance, for example the VGG16 and the ResNet50 both scored an f-score of 2%. On the other hand, the Xception network was the best performing network with an f-score of 96.7%. Moreover, the macro average precision and the macro average recall were 97% and 96%, respectively.

In the second set of experiments, the category that comprises classes with the least visual resemblance were first fed to the neural network. Hence, the first training set consisted of three classes: snake, turtle, and zebra. The validation macro average f-score was 100%. Next, the monkeys category was added to the training set in the second stage. At this point, the f-score decreased to 98.6%. Next, the training revert procedure was executed and the neural network was trained again on the first category and the undersampled monkey

category. Executing the training revert procedure improved the validation results, and the macro average f-score was 100%. In the final stage, the fox category was added to the training set, however, despite our expectations, the validation results did not change and the f-score of the resulting model was 100%.

For the purpose of monkeys classification, another Xception network was trained to only classify images into chimpanzee or orangutan. The network achieved a macro f-score of 98.6%. Moreover, the macro average precision and the macro average recall were both 99%.

4.4 Results of Reticulocytes Detection

In this section, the results of the approaches listed in Section 3.4 will be presented. Table 16 lists the main results of all experiments in terms of the average accuracy, macro f-score, and macro precision. The results in this table are obtained from averaging each metric among the 5-folds of each experiment.

Table 16: Results of 5-fold human reticulocyte classifier training

	Average accuracy (%)	Average macro precision (%)	Average macro F-score (%)
CENPARMI human reticulocyte dataset			
Pretrained on ImageNet	98.5	98.9	98.1
Pretrained on Feline data	98.9	98.9	98.6
No transfer learning	95.8	95.7	95
Balanced CENPARMI human reticulocyte dataset			
Pretrained ImageNet	96	96.4	96
Pretrained on Feline data	96.6	96.8	96.6
No transfer learning	89.1	90.9	88.4

Some highlights from the Table 16 are:

1. The best performing model is the one trained on the entire CENPARMI dataset, with weights initialized to the weights pretrained on the Feline dataset. This model

achieved 98.9%, 98.9%, 98.6% average accuracy, average macro precision, and average macro F-score respectively.

2. The worst performing model is the one trained on the balanced CENPARMI dataset, without TL. This model achieved 89.1%, 90.9%, 88.4% average accuracy, average macro precision, and average macro F-score respectively.
3. Overall, the experiments that were conducted without TL were the worst performing among the other experiments on the same dataset.
4. In general, the human reticulocyte classifiers that were pretrained on the feline dataset performed best among the other experiments trained on the same dataset.

4.4.1 Boost Measures Results

Table 17: Pretraining and similarity boost measure results

	CENPARMI human reticulocyte dataset (%)	Balanced CENPARMI human reticulocyte dataset (%)
ImageNet pretraining boost	3.1	7.6
Feline pretraining boost	3.6	8.2
Similarity boost	.5	.6

Table 17 lists the results of the pretraining boost measure and the dataset similarity boost. Some highlights from the table are:

1. All pretraining boost measure results are positive, which means that employing TL was always beneficial and resulted in better performance.
2. The similarity boost measure is positive using both datasets, this indicates that utilizing animal cells as a base dataset is beneficial, especially that the results of this utilization outperformed the results of the experiments that used Imagenet, which means that 2000 instances of animal cells were able to beat a large-scale dataset that consists of millions of images.

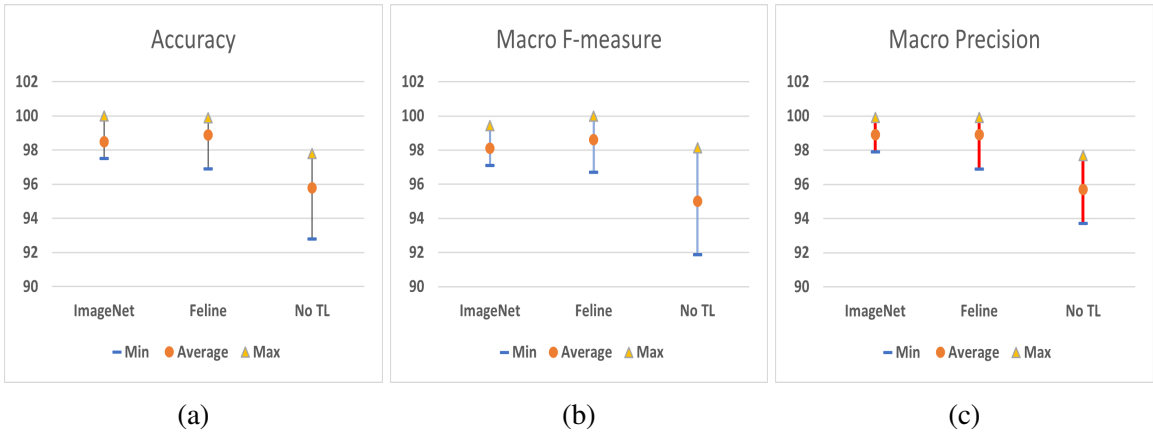


Figure 29: Min Max Average plots of the K-fold experiments using the CENPARMI dataset

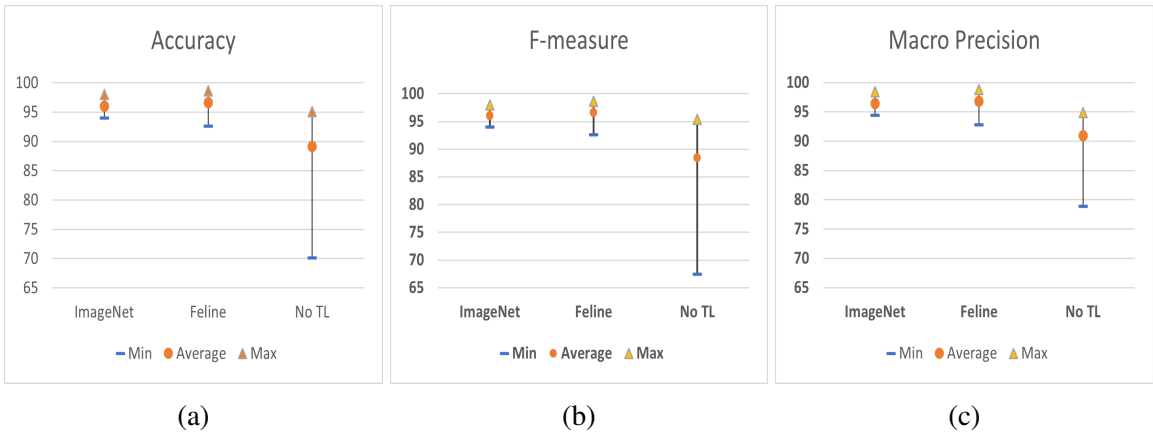


Figure 30: Min Max Average plots of the K-fold experiments using the balanced version of CENPARMI dataset

Next, the effect of utilizing the imbalanced CENPARMI dataset compared to the smaller balanced version of the same dataset, is discussed.

Figure 29 and Figure 30 illustrate the minimum, maximum and average values of the accuracy, f-score and macro precision of each experiment. The minimum in the figures represents the minimum value reported among the 5 folds of the experiment, and the maximum is also the maximum among the 5-folds. From the figures it is clear that the evaluation metrics of the experiments that did not utilize TL spans over the widest range of values; the gap between the minimum and the maximum is the highest.

Moreover, Table 18 lists the average loss and Mean Absolute Deviation (MAD) of each

Table 18: The average loss and mean absolute deviation of the conducted experiments

	Average loss	MAD
CENPARMI human reticulocyte dataset		
Pretrained on ImageNet	0.035	0.019
Pretrained on Feline data	0.040	0.012
No transfer learning	0.26	0.15
Balanced CENPARMI human reticulocyte dataset		
Pretrained on ImageNet	0.10	0.04
Pretrained on Feline data	0.11	0.03
No transfer learning	0.26	0.15

experiment. MAD is the average distance between each data value and the mean, see Equation 16, where \bar{x} is the average of the values and n is the number of data points. The larger the MAD, the greater variability there is in the data. Variability refers to how spread out the values are in a set of data points; high variability means the data is spread out, whereas low variability means the data is clustered together.

$$MAD = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (16)$$

The MAD of the experiments that did not utilize TL were approximately triple the MAD scores of the other experiments that utilized TL, this is because TL added robustness to the training process in a way that made the network perform the same regardless of the random test fold and the other 4 training folds. In addition, the MAD scores of the experiments that used the full imbalanced dataset were all lower than the ones of the balanced subset, which in terms indicates more robustness and less sensitivity to changes made on the training and test folds. Those conclusions are also supported by the minimum and maximum values in Figures 29 and 30, where it can be seen that the ranges of values of the experiments on the balanced dataset are wider than those on the full imbalanced dataset.

Moreover, Table 19 lists the dataset boost measure, from the table it is shown that the bigger dataset had a positive impact on the performance. Hence, given all the highlights and

conclusions about the dataset size, it can be concluded that the imbalance is not necessarily harmful on the networks performance, as more data points could aid in learning more sophisticated features and play a role in developing a function that separates data better.

Table 19: Dataset size boost measure

Experiment	Dataset size boost (%)
Pretrained on ImageNet	2.1
Pretrained on Feline data	2
No transfer learning	6.6

To sum up, the similar and dissimilar transfer learning experiments performed better compared to the model which was trained without TL in terms of the evaluation metric scores and the variability of the scores throughout the 5 folds. In addition, this study presented a novel conclusion regarding the utilization of animal medical data as a viable solution for the scarcity of human’s medical data. In the conducted experiments the human reticulocyte classifiers that were pretrained on feline data outperformed all the other classifiers. This conclusion holds a promising solution for medical data scarcity, which is a main challenge in computer aided diagnosis systems.

Chapter 5

Limitations

Due to the COVID-19 pandemic which constituted unprecedented challenges and pressure to the health care sector, we did not have access to values of human-level performance. Hence, only computer based performance is reported in this thesis.

Chapter 6

Conclusion and Future Work

In this thesis, a novel solution to creating a dataset of synthetic blood smears is proposed. The fact that each thin blood smear contains hundreds of blood cells, implies that the synthesizing procedure can be highly complicated. The novel synthetic dataset consists of 2500 instances, and was automatically annotated for 17 essential blood cell types and abnormal morphologies during the instances generation process. This type of datasets is useful since labeled medical data is scarce due to extra security and privacy constraints on it.

In order to create this dataset, 18 image pools were created in the first phase. In the second phase, RBCs, WBCs and platelets counts were selected from Gaussian distributions. Next, LSH was employed to divide cells' space into N projections and all near objects were hashed into the same bucket. All cells that hashed in the same bucket were tested against each other using Jaccard similarity, and all cells caused collision higher than an acceptable threshold were rejected. Three YOLO neural networks were trained on the first set of annotations to classify instances into WBCs, RBCs or platelets. The model was tested on the ALL-IDB dataset, and an accuracy score of 98.72% was achieved. The dataset was also reviewed by a group of highly experienced hematologists to ensure that it meets the general standards of making thin blood smears.

In order to further classify the instances into the subtypes and morphological abnormalities of platelets and WBCs, a set of experiments was conducted in the second stage of classification. For classifying platelets, a one-phase classification approach is experimented and a hybrid approach of deep learning and image processing techniques is proposed. The proposed approach handled the confusion caused by the visual resemblance between the giant and normal platelets while improving the platelet classification accuracy and macro-average precision from 82.6% to 98.6% and 76.6% to 97.6% respectively.

Moreover, for white blood cell classification, a novel scheme for training deep networks is proposed, namely, Enhanced Incremental Training, that automatically recognises and handles classes that confuse and negatively affect neural network predictions. To handle the confusable classes, a procedure called "training revert" is also proposed. Application of the proposed method has improved the classification accuracy and macro-average precision from 61.5% to 95% and 76.6% to 94.27%. The proposed method can be utilized in other application domains where confusable classes are thought to be the reason behind low classification results. Moreover, the proposed method was utilized on top of pretrained models available in an open source framework such as Keras. This indicates that the method can be easily reproduced and deployed by other researchers. The proposed enhanced incremental learning method was also experimented on two non-medical datasets that comprise confusable classes. Applying the proposed method improved the classification results on both datasets.

Finally, the classification of another type of smears, namely, reticulocyte smears, was discussed. Reticulocyte count blood test can reveal a wide range of illnesses and medical conditions. In this work, multiple experiments that utilized the concept of transfer learning were conducted. Moreover, three measures were defined, namely, the pretraining boost, the dataset similarity, and the dataset size boost measures to compare the effectiveness of the utilized experimental setups.

Evaluating those measures, three main conclusions can be drawn. First, using animals medical data holds a promising solution for human medical data scarcity, as utilizing weights that were pretrained on a medium sized feline reticulocyte dataset outperformed the model that utilized weights that were pretrained on the large scale ImageNet dataset.

Moreover, the experiments have shown that data imbalance is not always harmful to the performance as the bigger imbalanced version of the dataset outperformed the smaller balanced one. Third, utilizing transfer learning has a positive impact on the performance and generalizability of neural networks when compared to networks that initialize weights randomly.

For future work, the enhanced incremental training method will be tested on large-scale medical and non-medical datasets. Moreover, the influence of some factors on the results will be examined, such as category orderings, and the usage of more complicated criteria to evaluate the training progress between stages. Moreover, more animal datasets will be utilized for human medical applications. Moreover, the scope of studying the feasibility of utilizing animal data as a data scarcity solution will be further investigated by experimenting this option on different problem domains where human and animal data have mutual properties.

In addition, two aspects of PBS analysis that will be studied in my future work are: virtual staining and PBS quality assessment. Staining is the first step in preparing blood smears. This step is crucial since it reveals the nuclear and cytoplasmic details of cells, and hence, plays a major role in recognizing different blood cell types and abnormalities. Despite the importance of this step, it is considered costly, in terms of both time and money. Being able to examine blood smears without having to implement the physical staining step can be a huge transforming step in laboratory services. This research direction has not been thoroughly investigated in the literature of computer science.

Furthermore, the quality of digital blood smears can directly affect blood cell classification results [107]. Hence, developing computerized PBS quality assessment systems can be used as an educational tool for beginner lab technician and students. This area of research has not been completely addressed and there is still room for improvement in this context, as it only exists as a supplementary step in a few works in the literature.

Chapter 7

Publications

1. Rabiah Al-qudah, C. Y. Suen (2022), Blood Smear Analysis Using Deep Learning: Current Challenges and Future Directions, Invited book chapter, “The Computational Intelligence and Image Processing in Medical applications book” which will be a part of the “World scientific Publishing Co Series in Computer Vision, edited by C. H. Chen
2. Rabiah Al-qudah, C. Y. Suen (2021), Can Animal Medical Data be a Possible Solution for Human Medical Data Scarcity?, submitted to the “Neural Processing Letters Journal”. Impact factor: 2.9
3. Rabiah Al-qudah, C. Y. Suen (2021), Intensive Survey on Peripheral Blood Smear Analysis Using Deep Learning, N. Nobile, M. Blom and C. Y. Suen (eds.), Advances in Pattern Recognition and Artificial Intelligence, Volume 6, Pages 23-45, DOI: https://doi.org/10.1142/9789811239014_0002
4. Rabiah Al-qudah, C. Y. Suen (2021), Improving Blood Cells Classification in Peripheral Blood Smears Using Enhanced Incremental Deep Learning, Computers in Biology and Medicine, Volume 131, Pages 104265. DOI:<https://doi.org/10.1016/j.compbiomed.2021>. Impact factor: 4.5

5. Rabiah Al-qudah, Suen C.Y. (2020) A Survey on Peripheral Blood Smear Analysis Using Deep Learning. In: Lu Y., Vincent N., Yuen P.C., Zheng WS., Cheriet F., Suen C.Y. (eds) Pattern Recognition and Artificial Intelligence. ICPRAI 2020. Lecture Notes in Computer Science, vol 12068. Springer, Cham. https://doi.org/10.1007/978-3-030-59830-3_63

6. Rabiah Al-qudah, C. Y. Suen (2020), Synthetic Blood Smears Generation Using Locality Sensitive Hashing and Deep Neural Networks, IEEE Access Journal, Volume 8, Pages 102530-102539, DOI: 10.1109/ACCESS.2020.2999349. Impact factor: 3.367

Bibliography

- [1] Amber Erickson Gabbey. Reticulocyte count: Purpose, procedure, and results. <https://www.healthline.com/health/reticulocyte-count>, June 2017.
- [2] Ask Hematologist, Understand Hematology. Miscellaneous red cell abnormalities. <https://askhematologist.com/miscellaneous-red-cell-abnormalities/>, 03 2016. Accessed in March 2021.
- [3] Concordia University. Centre for pattern recognition and machine intelligence. <https://www.concordia.ca/research/cenparmi.html>, May 2021.
- [4] National Institute of Health. National institute of health malaria dataset. <https://lhncbc.nlm.nih.gov/publication/pub9932>, 2018. Accessed in March 2021.
- [5] World Health Organization. Who guidelines on drawing blood: Best practices in phlebotomy. <https://www.ncbi.nlm.nih.gov/books/NBK138650/>, 01 2010.
- [6] N. A. Ahmed, A. Yiğit, Z. Isik, and A. Alpkocak. Identification of leukemia subtypes from microscopic images using convolutional neural network. *Diagnostics*, 9(3):104, 2019.
- [7] G. Abdul Hamid. *Manual of Hematology*, page 53. 01 2012.
- [8] G. Abdul Hamid. *Manual of Hematology*, page 57. 01 2012.
- [9] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.
- [10] A. Adewoyin and B. Nwogoh. Peripheral blood film - a review. *Annals of Ibadan postgraduate medicine*, 12:71–9, 12 2014.
- [11] D. Aditya, N. Maxim, and G. Michael. Adabatch: Adaptive batch sizes for training deep neural networks, 2018.
- [12] R. Al-qudah and C. Y. Suen. Enhancing yolo deep networks for the detection of license plates in complex scenes. In *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, DATA '19*, New York, NY, USA, 2019. Association for Computing Machinery.

- [13] S. Alferez, A. Merino, A. Acevedo Lipes, L. Puigvi, and J. Rodellar. Color clustering segmentation framework for image analysis of malignant lymphoid cells in peripheral blood. *Medical and Biological Engineering and Computing*, 57, 02 2019.
- [14] H. A. Aliyu, R. Sudirman, M. A. Abdul Razak, and M. A. Abd Wahab. Red blood cell classification: Deep learning architecture versus support vector machine. In *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, pages 142–147, Kuching, Malaysia, July 2018.
- [15] N. Alok, K. Krishan, and P. Chauhan. *Deep Learning-Based Image Classifier for Malaria Cell Detection*, chapter 12, pages 187–197. John Wiley and Sons, Ltd, 2021.
- [16] R. Alqudah and C. Suen. *A Survey on Peripheral Blood Smear Analysis Using Deep Learning*, volume 12068, chapter 63, pages 725–738. Springer, Switzerland, 11 2020.
- [17] M. Alsaedi, T. Fevens, A. Krzyzak, and L. Jelen. Cytological malignancy grading systems for fine needle aspiration biopsies of breast cancer. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 705–709, 2017.
- [18] J. Angulo and G. Flandrin. Automated detection of working area of peripheral blood smears using mathematical morphology. *Analytical Cellular Pathology : the journal of the European Society for Analytical Cellular Pathology*, 25:37–49, 02 2003.
- [19] K. Anilkumar, V. Manoj, and T. Sagi. A survey on image segmentation of blood and bone marrow smear images with emphasis to automated detection of leukemia. *Biocybernetics and Biomedical Engineering*, 40(4):1406 – 1420, 2020.
- [20] S. Arslan, E. Ozyurek, and C. Gunduz-Demir. A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images. *Cytometry Part A*, 85(6):480–490, 2014.
- [21] V. Arunagiri and R. B. Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, pages 1–21, 2019.
- [22] V. Athitsos, M. Potamias, P. Papapetrou, and G. Kollios. Nearest neighbor retrieval using distance-based hashing. In *2008 IEEE 24th International Conference on Data Engineering*, pages 327–336, April 2008.
- [23] O. Bailo, D. Ham, and Y. Min Shin. Red blood cell image generation for data augmentation using conditional generative adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Long Beach, CA, USA, June 2019.

- [24] F. Baker and R. Silverton. 28 - introduction to haematology. In *Introduction to Medical Laboratory Technology (Fifth Edition)*, pages 549 – 558. Butterworth-Heinemann, fifth edition edition, 1976.
- [25] S. Banton. Human peripheral reticulocyte isolation and exosome release in vitro. *Georgia Institute of Technology*, 1:9–11, 2017.
- [26] Y. Baydilli and U. Atila. Classification of white blood cells using capsule networks. *Computerized Medical Imaging and Graphics*, 80:101699, 01 2020.
- [27] D. Bibin, M. S. Nair, and P. Punitha. Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access*, 5:9099–9108, 2017.
- [28] British Society for Immunology. Eosinophils. <https://tinyurl.com/2p9aes8y>, 12 2020.
- [29] R. W. Budenz and G. W. Bernard. Osteogenesis and leukopoiesis within diffusion-chamber implants of isolated bone marrow subpopulations. *American Journal of Anatomy*, 159(4):455–474, 1980.
- [30] F. M. Castro, M. J. Marín-Jiménez, N. Guil, C. Schmid, and K. Alahari. End-to-end incremental learning. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision – ECCV 2018*, pages 241–257, Switzerland, 2018. Springer International Publishing.
- [31] F. Chang, H.-C. Chen, and H.-C. Liu. Double k-folds in svm. In *2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pages 384–387, 2015.
- [32] C. Chen, M.-Y. Liu, O. Tuzel, and J. Xiao. R-cnn for small object detection. In S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, editors, *Computer Vision – ACCV 2016*, pages 214–230, Switzerland, 2017. Springer International Publishing.
- [33] V. Cheplygina. Cats or cat scans: Transfer learning from natural or medical image source data sets? *Current Opinion in Biomedical Engineering*, 9:21–27, 2019. Futures of BME: Digital Health and BME, Biomedical Imaging: Cardiovascular Imaging.
- [34] Coinmac, CC BY-SA 3.0. Stained and unstained blood smears. <https://commons.wikimedia.org/w/index.php?curid=20155466>, 2020.
- [35] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [36] C. Di Ruberto, A. Loddo, and G. Puglisi. Blob detection and deep learning for leukemic blood image analysis. *Applied Sciences*, 10:1176, 02 2020.

- [37] Z. Ding, N. F. Lepora, and E. Johns. Sim-to-real transfer for optical tactile sensing. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1639–1645, 2020.
- [38] DonatingPlasma.Org. What is plasma? <https://www.donatingplasma.org/donation/what-is-plasma>. Accessed in Jan 2021.
- [39] N. Duncan Riach. Determinism in deep learning. <https://developer.nvidia.com/gtc/2019/video/S9911>, 2019. Accessed in April 2021.
- [40] S. M. El Jamal, C. Salib, A. Stock, N. I. Uriarte-Haparnas, B. S. Glicksberg, J. Teruya-Feldstein, F. R. Dembitzer, G. N. Nadkarni, and A. Firpo-Betancourt. Atypical lymphocyte morphology in sars-cov-2 infection. *Pathology - Research and Practice*, 216(9):153063, 2020.
- [41] Farlex Partner Medical Dictionary. Human blood. (n.d.). <https://medical-dictionary.thefreedictionary.com/Human+Blood>, 01 2021.
- [42] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 03 2018.
- [43] P. Fulkerson and M. Rothenberg. Targeting eosinophils in allergy, inflammation and beyond. *Nature reviews. Drug discovery*, 12, 01 2013.
- [44] X. Geng and K. Smith-Miles. *Incremental Learning*, pages 731–735. Springer US, Boston, MA, 2009.
- [45] B. George-Gay and K. Parker. Understanding the complete blood count with differential. *Journal of PeriAnesthesia Nursing*, 18(2):96 – 117, 2003.
- [46] C. B. Grindem. Schalm’s veterinary hematology, 6th edition. editors: Douglas j. weiss, k. jane wardrop. *Veterinary Clinical Pathology*, 40(2):121–133, 2011.
- [47] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama. Gan-based synthetic brain mr image generation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 734–738, 2018.
- [48] R. Harrand. Feline reticulocytes. <https://www.kaggle.com/tentotheminus9/feline-reticulocytes>, March 2018.
- [49] C. Hartanto, S. Kurniawan, D. Arianto, and A. Arymurthy. Dcgan-generated synthetic images effect on white blood cell classification. volume 1077, pages 12–33, 02 2021.

- [50] S. C. Hellewell, A. Conquest, L. Little, S. Vallance, J. Board, R. Bellomo, D. J. Cooper, and M. C. Morganti-Kossmann. Epo treatment does not alter acute serum profiles of gfap and s100b after tbi: A brief report on the australian epo-tbi clinical trial. *Journal of Clinical Neuroscience*, 76:5 – 8, 2020.
- [51] Q. Huang, W. Li, B. Zhang, Q. Li, R. Tao, and N. H. Lovell. Blood cell classification based on hyperspectral imaging with modulated gabor and cnn. *IEEE Journal of Biomedical and Health Informatics*, 24(1):1, 2019.
- [52] M.-Y. Huh, P. Agrawal, and A. A. Efros. What makes imagenet good for transfer learning? *ArXiv*, abs/1608.08614, 2016.
- [53] R. Istrate, A. C. I. Malossi, C. Bekas, and D. Nikolopoulos. Incremental training of deep convolutional neural networks. In *Proceedings of the International Workshop on Automatic Selection, Configuration and Composition of Machine Learning Algorithms*, pages 41–48, 09 2017.
- [54] P. J., den Braber I., K. L. Vrisekoop N., de Boer RJ., B. JA., T. K., and K. L. Abnormalities in leukocyte morphology and number. *Blood*, 116:625–627, 07 2010.
- [55] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, and K. Bousmalis. Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12619–12629, Los Alamitos, CA, USA, jun 2019. IEEE Computer Society.
- [56] Y. Jiang, M. Yang, S. Wang, X. Li, and Y. Sun. Emerging role of deep learningbased artificial intelligence in tumor pathology. *Cancer Communications*, 40:154–166, 2020.
- [57] W. Johnson and J. Lindenstrauss. Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics*, 26:189–206, 01 1984.
- [58] B. Journal. Rouleaux formation. *Blood*, 107(11):4205–4205, 06 2006.
- [59] Kaggle. Animals dataset. <https://www.kaggle.com/iamsouravbanerjee/animal-image-dataset-90-different-animals>, Dec 2021.
- [60] Kaggle. Humming birds dataset. <https://www.kaggle.com/akimball002/hummingbirds-at-my-feeders>, Dec 2021.
- [61] S. C. Kalkan and O. K. Sahingoz. Deep learning based classification of malaria from slide images. In *2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)*, pages 1–4, Istanbul, Turkey, April 2019.
- [62] M. Kaspar, J. Muñoz Osorio, and J. Bock. Sim2real transfer for reinforcement learning without dynamics randomization. pages 4383–4388, 10 2020.

- [63] A. Ke, W. Ellsworth, O. Banerjee, A. Ng, and P. Rajpurkar. Chextransfer: performance and parameter efficiency of imagenet models for chest x-ray interpretation. *Proceedings of the Conference on Health, Inference, and Learning*, 2021.
- [64] Keras. Keras applications. <https://keras.io/api/applications/>. Accessed in Jan 2020.
- [65] Keras. Keras faq. https://keras.io/getting_started/faq/#how-can-i-obtain-reproducible-results-using-keras-during-development. Accessed in Feb 2020.
- [66] M. Khodashenas, H. Ebrahimpour-komleh, and A. M. Nickfarjam. White blood cell detection and counting based on genetic algorithm. In *2019 Advances in Science and Engineering Technology International Conferences (ASET)*, pages 1–4, 2019.
- [67] V. Krunoslav, F. Pierluigi, B. Maja, and T. Romana. Using convolutional neural networks for determining reticulocyte percentage in cats, 2018.
- [68] R. Kumar, S. K. Singh, and A. Khamparia. Malaria detection using custom convolutional neural network model on blood smear slide images. In *Advanced Informatics for Computing Research*, pages 20–28, Singapore, June 2019.
- [69] R. Labati, V. Piuri, and F. Scotti. All-idb: the acute lymphoblastic leukemia image database for image processing. In *18th IEEE International Conference on Image Processing*, pages 2045–2048, Brussels, Belgium, Sept 2011.
- [70] P. Lanzkowsky. Chapter 11 - disorders of white blood cells. In P. Lanzkowsky, editor, *Manual of Pediatric Hematology and Oncology (Fifth Edition)*, pages 272 – 320. Academic Press, San Diego, fifth edition edition, 2011.
- [71] J. Lemley, S. Bazrafkan, and P. Corcoran. Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, 5:5858–5869, 2017.
- [72] G. Liang, H. Hong, W. Xie, and L. Zheng. Combining convolutional neural network with recursive neural network for blood cell image classification. *IEEE Access*, 6:36188–36197, 2018.
- [73] K. Liu, R. Shuai, L. Ma, and ZeXu. Cells image generation method based on vae-gan. *Procedia Computer Science*, 183:589–595, 2021. Proceedings of the 10th International Conference of Information and Communication Technology.
- [74] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *Computer Vision – ECCV 2016*, pages 21–37, Switzerland, 2016. Springer International Publishing.
- [75] Y. Liu and F. Long. Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning. *bioRxiv*, 2019.

- [76] P. Manescu, L. Neary-Zajiczek, M. J. Shaw, M. Elmi, R. Claveau, V. M. Pawar, J. Shawe-Taylor, I. Kokkinos, M. A. Srinivasan, I. Lagunju, O. O. Sodeinde, B. J. Brown, and D. Fernandez-Reyes. Deep learning enhanced extended depth-of-field for thick blood-film malaria high-throughput microscopy. *ArXiv*, abs/1906.07496, 2019.
- [77] A. E. Mast, M. A. Blinder, and D. J. Dietzen. Reticulocyte hemoglobin content. *American Journal of Hematology*, 83(4):307–310, 2008.
- [78] I. Mondragón-García, P. Flores-Guzmán, and H. Mayani. Human cord blood hematopoietic cells acquire neural features when cultured in the presence of neurogenic cytokines. *Blood Cells, Molecules, and Diseases*, 85:102485, 2020.
- [79] D. Mundhra, B. Cheluvaraju, J. Rampure, and T. Rai Dastidar. Analyzing microscopic images of peripheral blood smear using deep learning. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 178–185, Cham, 2017. Springer International Publishing.
- [80] D. Mundhra, B. Cheluvaraju, J. Rampure, and T. Rai Dastidar. Analyzing microscopic images of peripheral blood smear using deep learning. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 178–185, Switzerland, 2017. Springer International Publishing.
- [81] Nikon. Eclipse e100 microscope. <https://www.microscope.healthcare.nikon.com/products/upright-microscopes/eclipse-e100>. Accessed in May 2021.
- [82] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
- [83] W. Pan, Y. Dong, and D. Wu. *Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks*, pages 159–172. IntechOpen, 09 2018.
- [84] A. Patil, M. Patil, and G. Birajdar. White blood cells image classification using deep learning with canonical correlation analysis. *IRBM*, 42(5):378–389, 2021.
- [85] P. A. Pattanaik, M. Mittal, and M. Z. Khan. Unsupervised deep learning cad scheme for the detection of malaria in blood smear microscopic images. *IEEE Access*, 8:94936–94946, 2020.
- [86] S. Peter Klinken. Red blood cells. *The International Journal of Biochemistry and Cell Biology*, 34(12):1513 – 1518, 2002.
- [87] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pages 337–345, 1990.

- [88] A. Prakash, S. Debnath, J.-F. Lafleche, E. Cameracci, G. State, and M. Law. Sim2sg: Sim-to-real scene graph generation for transfer learning, 11 2020.
- [89] J. Prellberg and O. Kramer. Acute lymphoblastic leukemia classification from microscopic images using convolutional neural networks, 2020.
- [90] S. Purwar, R. K. Tripathi, R. Ranjan, and R. Saxena. Detection of microcytic hypochromia using cbc and blood film features extracted from convolution neural network by different classifiers. *Multimedia Tools and Applications*, 79:4573–4595, 2019.
- [91] F. Qin, N. Gao, Y. Peng, Z. Wu, S. Shen, and A. Grudtsin. Fine-grained leukocyte classification with deep residual learning for microscopic images. *Computer Methods and Programs in Biomedicine*, 162(8):243–252, 2018.
- [92] Rabiah Al-Qudah. Synthetic blood smears questionnaire. <https://forms.gle/1b36HbNRwFVhpVQeA>.
- [93] A. Rahman, H. Zunair, M. S. Rahman, J. Q. Yuki, S. Biswas, M. A. Alam, N. B. Alam, and M. R. C. Mahdy. Improving malaria parasite detection from red blood cell using deep convolutional neural networks, 2019.
- [94] A. Rahman, H. Zunair, M. S. Rahman, J. Q. Yuki, S. Biswas, M. A. Alam, N. B. Alam, and M. R. C. Mahdy. Improving malaria parasite detection from red blood cell using deep convolutional neural networks, 2019.
- [95] S. Rajaraman, S. Antani, M. Poostchi, K. Silamut, R. Maude, S. Jaeger, and G. Thoma. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, 6(e4568), 2018.
- [96] S. Rajaraman, S. Jaeger, and S. K. Antani. Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ*, 7, 2019.
- [97] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. P. Langlotz, K. S. Shpanskaya, M. P. Lungren, and A. Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *ArXiv*, abs/1711.05225, 2017.
- [98] M. I. Razzak and S. Naz. Microscopic blood smear segmentation and classification using deep contour aware cnn and extreme machine learning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 801–807, July 2017.
- [99] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. *arXiv preprint arXiv:1612.08242*, 2016.

- [100] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.
- [101] EclinPath, Cornell University College of Veterinary Medicine. Normal leukocytes. <https://eclinpath.com/hematology/morphologic-features/white-blood-cells/normal-leukocytes/>, 2020. Accessed in Feb 2021.
- [102] H. Rezatofghi and H. Soltanian-Zadeh. Automatic recognition of five types of white blood cells in peripheral blood. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 35:333–43, 02 2011.
- [103] B. F. Rodak, G. A. Fritsma, and K. Doig. *Hematology: Clinical Principles and Applications*. Elsevier Health Sciences, 2007.
- [104] D. Roy, P. Panda, and K. Roy. Tree-cnn: A deep convolutional neural network for lifelong learning. *ArXiv*, abs/1802.05800, 2018.
- [105] G. D. Sad, L. D. Terissi, and J. C. Gómez. Class confusability reduction in audio-visual speech recognition using random forests. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 584–592, Switzerland, 2018. Springer International Publishing.
- [106] S. Saleem, J. Amin, M. Sharif, M. Almas Anjum, M. Iqbal, and S. Wang. A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models. *Complex and Intelligent Systems*, -():1–16, 2021.
- [107] U. Sara, M. Akter, and M. Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 07:8–18, 01 2019.
- [108] M. Saraswat and K. Arya. Automated microscopic image analysis for leukocytes identification: A survey. *Micron*, 65:20 – 33, 2014.
- [109] S. Sarwar, A. Ankit, and K. Roy. Incremental learning in deep convolutional neural networks using partial network sharing. *IEEE Access*, PP, 12 2017.
- [110] S. Shafique and S. Tehsin. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technology in Cancer Research and Treatment*, 17:1–7, 2018.
- [111] B. Shah, N. Burg, and M. H. Pillinger. Chapter 11 - neutrophils. In G. S. Firestein, R. C. Budd, S. E. Gabriel, I. B. McInnes, and J. R. O’Dell, editors, *Kelley and Firestein’s Textbook of Rheumatology (tenth edition)*, pages 169 – 188.e3. Elsevier, tenth edition edition, 2017.
- [112] A. Shakarami, M. B. Menhaj, A. Mahdavi-Hormat, and H. Tarrah. A fast and yet efficient yolov3 for blood cell detection. *Biomedical Signal Processing and Control*, 66:102495, 2021.

- [113] P. Shamsolmoali, M. Zareapoor, E. Granger, H. Zhou, R. Wang, M. E. Celebi, and J. Yang. Image synthesis with adversarial networks: A comprehensive survey and case studies. *Information Fusion*, 72:126–146, 2021.
- [114] M. Shanker, M. Hu, and M. Hung. Effect of data standardization on neural network training. *Omega*, 24(4):385 – 397, 1996.
- [115] Shenggan Github account. Bccd dataset. https://github.com/Shenggan/BCCD_Dataset.
- [116] N. Singla and V. Srivastava. Deep learning enabled multi-wavelength spatial coherence microscope for the classification of malaria-infected stages with limited labelled data size. *Optics & Laser Technology*, 130:106335, 2020.
- [117] A. Stefanovska. Physics of the human cardiovascular system. *Contemporary Physics*, 40(1):31–55, 1999.
- [118] F. K. T. J. S. M. M. S. M. N. R. T. Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. *Diagnostics*, 10:329, 05 2020.
- [119] T. T., T. K., P. J., B. J., and K. L. What’s your age again? determination of human neutrophil half-lives revisited. *Journal of leukocyte biology*, 94:595–601, 04 2013.
- [120] A. Talaat, P. Kollmannsberger, and A. Ewees. Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Scientific Reports*, 10, 02 2020.
- [121] Y. Tao, Y. Tu, and M. Shyu. Efficient incremental training for deep convolutional neural networks. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 286–291, 2019.
- [122] T. Thanh, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, and K.-W. Kang. Blood cell images segmentation using deep learning semantic segmentation. In *2018 IEEE International Conference on Electronics and Communication Engineering (ICECE)*, pages 13–16, Xian, China, Dec 2018.
- [123] T. Thanh, C. Vununu, S. Atoev, S.-H. Lee, and K.-R. Kwon. Leukemia blood cell image classification using convolutional neural network. *International journal of computer theory and engineering*, 10(2):54–58, 2018.
- [124] C. Touzeau, C. Pellat-Deceunynck, T. Gastinne, F. Accard, G. Jego, H. Avet-Loiseau, N. Robillard, J. Harousseau, R. Bataille, and P. Moreau. Reactive plasmacytoses can mimic plasma cell leukemia: Therapeutical implications. *Leukemia and lymphoma*, 48:207–8, 02 2007.
- [125] L. Tremolizzo, G. Sala, and C. Ferrarese. *Platelet Activation*, pages 1034–1035. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

- [126] University of Leeds, the Histology Guide. White blood cells. https://www.histology.leeds.ac.uk/blood/blood_wbc.php. Accessed in June 2020.
- [127] US Drug and Food Administration. Recommendations for investigational covid-19 convalescent plasma. <https://tinyurl.com/4y3v8cjm>, 08 2020.
- [128] K. Vinicki, P. Ferrari, M. Belic, and R. Turk. Using convolutional neural networks for determining reticulocyte percentage in cats. *ArXiv*, abs/1803.04873, 2018.
- [129] L. H. Vogado, R. M. Veras, F. H. Araujo, R. R. Silva, and K. R. Aires. Leukemia diagnosis in blood slides using transfer learning in cnns and svm for classification. *Eng. Appl. Artif. Intell.*, 72(C):415–422, 2018.
- [130] L. H. S. Vogado, R. D. M. S. Veras, A. R. Andrade, F. H. D. D. Araujo, R. R. V. e. Silva, and K. R. T. Aires. Diagnosing leukemia in blood smear images using an ensemble of classifiers and pre-trained convolutional neural networks. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 367–373, Campo Grande, Brazil, Oct 2017.
- [131] H. J. Walker HK, Hall WD. *Clinical Methods: The History, Physical, and Laboratory Examinations*. Elsevier, 1990.
- [132] G. Wang, T. Zhao, Z. Fang, H. Lian, X. Wang, Z. Li, W. Wu, B. Li, and Q. Zhang. Experimental evaluation of deep learning method in reticulocyte enumeration in peripheral blood. *International Journal of Laboratory Hematology*, 43(4):597–601, 2021.
- [133] J. L. Wang, A. Y. Li, M. Huang, A. K. Ibrahim, H. Zhuang, and A. M. Ali. Classification of white blood cells with patternnet-fused ensemble of convolutional neural networks (pecnn). In *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 325–330, Louisville, KY, USA, Dec 2018.
- [134] Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, and S. Yang. Deep learning approach to peripheral leukocyte recognition. *PloS one*, 14(6), 2019.
- [135] Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, and S. Yang. Deep learning approach to peripheral leukocyte recognition. *PloS one*, 14(6), 2019.
- [136] X. Wang, T. Xu, J. Zhang, S. Chen, and Y. Zhang. So-yolo based wbc detection with fourier ptychographic microscopy. *IEEE Access*, 6:51566–51576, 2018.
- [137] Y. Wang and Y. Cao. Human peripheral blood leukocyte classification method based on convolutional neural network and data augmentation. *Medical Physics*, 47, 11 2019.
- [138] Wikipedia. A cbc specimen in front of a printout displaying cbc and differential results. https://en.wikipedia.org/wiki/Complete_blood_count, 2020.

- [139] T. Xiao, J. Zhang, K. Yang, Y. Peng, and Z. Zhang. Error-driven incremental learning in deep convolutional neural network for large-scale image classification. In *ACM Multimedia*, pages 177–186, November 2014.
- [140] M. Xu, D. Papageorgiou, S. Abidi, M. Dao, H. Zhao, and G. Karniadakis. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLOS Computational Biology*, 13:e1005746, 10 2017.
- [141] Xueyin Lin and W. Wee. Shape detection using range data. In *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, pages 34–39, 1985.
- [142] F. Yang, M. Poostchi, H. Yu, Z. Zhou, K. Silamut, J. Yu, R. Maude, S. Jaeger, and S. Antani. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1, 2019.
- [143] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3320–3328. Curran Associates, Inc., 2014.
- [144] K. You, Z. Kou, M. Long, and J. Wang. Co-tuning for transfer learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17236–17246. Curran Associates, Inc., 2020.
- [145] M. Zareapoor, P. Shamsolmoali, D. Kumar Jain, H. Wang, and J. Yang. Kernelized support vector machine with deep learning: An efficient approach for extreme multiclass dataset. *Pattern Recognition Letters*, 115:4 – 13, 2018. Multimodal Fusion for Pattern Recognition.
- [146] Y. Zhang, C. Jia, and C. K. Kwok. Predicting the interaction biomolecule types for lncRNA: an ensemble deep learning approach. *Briefings in Bioinformatics*, 10 2020.
- [147] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, page 1–34, 2020.
- [148] G. Zini. Abnormalities in leukocyte morphology and number. *Blood and Bone Marrow Pathology*, pages 247–261, 01 2011.
- [149] G. Zini. Abnormalities in leukocyte morphology and number. *Blood and Bone Marrow Pathology*, page 257, 01 2011.

Appendix A

WBC types

A.1 Neutrophils

Neutrophils are multilobed-nucleus cells. The cytoplasm of this type of cells contains numerous purple granules called azurophilic that contain microbicidal agents. Moreover, Neutrophils make up approximately 50% of the WBC population and are responsible for the phagocytosis and degradation of foreign organisms, through activating proteases and other antibiotic molecules, and generating toxic oxygen radicals [111]. Neutrophils are produced daily and normally have a circulating lifespan of 5 to 135 hours in humans [54, 119].

Alterations in Neutrophils cytoplasm or nuclei lobes is considered as a morphological abnormality. Nucleus of normal Neutrophils have an average of three lobes and always less than five lobes. hypersegmentation is an abnormality that can be diagnosed in lab when more than 3 cells have 5 lobes or a single cell with six lobes is present in a sample of 100 cells [148]. On the other hand, a defect in the Neutrophil lobulation is called Hyposegmenation resulting in the presence of dumbbell-shaped neutrophils with bilobed nuclei.

Finally, a third example of Neutrophil abnormalities is band cells. This abnormality is characterized by having a curved but not lobular nuclei. Band cells are immature Neutrophils, and an increase in its count typically means that the bone marrow has been signaled to increase the production of WBCs due to infection or inflammation.

A.2 Eosinophils

Eosinophils are major effector cells. Eosinophils make up to 1% of the white blood cells in normal individuals [43, 28].

Human eosinophils have bi-lobed nucleus, pink staining with eosin and characteristic cytoplasmic granules. They are also active participants in many immune responses.

A.3 Basophil

Basophils are the only WBCs that contain histamine. Moreover, they are the least frequent WBCs; as they make up less than 1% of the WBC population in humans [149]. Basophils play a major role in preventing blood clotting, because they contain heparin, which is a natural blood-thinning substance. Moreover, when the immune system is exposed to an allergen this type of cells release histamine and mediate in handling allergic reactions.

A.4 Monocytes

Monocytes are the largest cells of the blood measuring between 16 and 22 μm in diameter. Monocytes typically have one nuclei that is usually centrally placed within the cell and often kidney shaped [24]. Cells of this type fight certain infections and regulate immunity against foreign substances, moreover, they act as scavenger cell.

A.5 Lymphocytes

Lymphocytes make up about 25% of the total WBC count. Two forms of Lymphocytes are: B cells, and T cells. Each B cell is set to make one specific antibody, that can only match one specific antigen. On the other hand, T cells process foreign substances for removal. Reactive Lymphocyte, atypical Lymphocyte, Plasma cells, and Basket cells are all considered as morphological abnormalities of Lymphocytes. Reactive Lymphocytes and plasma cells are considered as large lymphocytes. On the other hand, Basket cells are remnants of cells that lack any identifiable structure. Such cells are associated with disorders such as chronic lymphocytic leukemia (CLL). Finally, atypical Lymphocytes [40] are a heterogeneous group of lymphocytes with morphological characteristics that are not considered among the normal ones. Each of the mentioned abnormalities, when accompanied with clinical findings, can aid in the diagnosis of many illnesses.