공학박사 학위논문

Automated Leukocyte Differential Count System Using Dual-Stage Convolutional Neural Network

이중 합성곱 신경망을 이용한 백혈구 백분율 자동 분석 시스템에 관한 연구

2017 년 8 월

서울대학교 대학원
협동과정 바이오엔지니어링 전공
최 진 우

Ph. D. Dissertation

Automated Leukocyte Differential Count System Using Dual-Stage Convolutional Neural Network

BY

JIN WOO CHOI

AUGUST 2017

INTERDISCIPLINARY PROGRAM IN BIOENGINEERING

THE GRADUATE SCHOOL

SEOUL NATIONAL UNIVERSITY

ABSTRACT

Automated Leukocyte Differential Count System Using Dual-Stage Convolutional Neural Network

Jin Woo Choi
Interdisciplinary Program in Bioengineering
The Graduate School
Seoul National University

Leukocyte or white blood cell differential count is an essential examination modality of hematology laboratory in diagnosis of various blood disorders. However, it requires highly experienced hematologists for correct diagnosis from samples with inter- and intra-sample variations. Due to tedious, time and cost consuming procedure of manual differential count, there has been high demands for development of automated system. In order for it to be applicable in clinical hematology laboratories, an automated system will have to detect and classify leukocytes of different maturation stages, especially in bone marrow aspirate smears. This has been a challenging problem in computer vision, image processing, and machine learning, because of complex nature of bone marrow aspirate

smear. The leukocyte has multiple maturation stages, and these maturation stages have small inter-class differences, so it is difficult to differentiate even with expert knowledge. Moreover, a problem of color, shape, and size variations among samples exists and a problem of touching cell due to high leukocyte density of bone marrow aspirate smear exists.

In this dissertation, an automated leukocyte differential count system for bone marrow aspirate smear was developed to overcome problems of manual differential count and to fulfill clinically unmet needs. The system should perform the differential count with high accuracy and objectivity, and high throughput and efficiency. Moreover, it should overcome challenges of bone marrow aspirate smear. To this end, a large dataset of bone marrow smear was collected for development of a detection and a classification algorithms. Watershed transformation and saliency map were utilized for single-leukocyte detection, and the dual-stage convolutional neural network that learns global and local features of complex leukocyte maturation stages was proposed for classification. Lastly, a probability guidance algorithm was proposed for integration of detection and classification algorithms. The

performance of proposed system was assessed with ten leukocyte maturation stages of myeloid and erythroid series in bone marrow aspirate smears.

Total of 200 large (1388 × 1040) digital images of bone marrow aspirate smears and 2,323 small (96 × 96) single leukocyte digital images were collected. The proposed system showed a state-of-the-art performance. It achieved an average detection accuracy of 96.09% and an average classification accuracy of 97.06%, and it was able to differential count 100 leukocytes in 4 to 5 seconds. This proposes a new paradigm in diagnosis of blood disorder and showed a potential of deep learning, especially the convolutional neural network, in medical image processing. The proposed system is expected to increase the total number of analyzed leukocytes in a sample, which will provide more statistically reliable information of a patient for diagnosis.

Keywords: Leukocyte differential count, Convolutional neural network, Bone marrow aspirate smear, Automated differential count system, Computer-aided diagnosis

Student number: 2013-30303

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AG	Augmented
ALL	Acute lymphoblastic leukemia
AML	Acute myelogenous leukemia
ANN	Artificial neural network
CAD	Computer-aided diagnosis
CBC	Complete blood count
CLL	Chronic lymphoblastic leukemia
CML	Chronic myelogenous leukemia
CNN	Convolutional neural network
GPU	Graphics processing unit
GUI	Graphical user interface
OS	Oversampled
RELU	Rectified linear-unit
ROI	Region-of-interest
SVM	Support vector machine
WBC	White blood cell

CHAPTER 1

INTRODUCTION

1.1. Introduction to Hematology

Leukocytes or white blood cells are key components of human immune system. The main role of leukocytes is protection of body against infectious diseases and foreign invaders. There exists five types of mature leukocytes in peripheral blood, including monocyte, lymphocyte, basophil, eosinophil, and neutrophil. These five types of leukocytes are derived from the progenitor cells, the pluripotent hematopoietic stem cells, in bone marrow niche through a differentiation process called hematopoiesis [1, 2]. Hematopoiesis is a highly regulated system that involves interactions between the genetic program and the bone marrow micro-architecture [2]. It consists of complex multiple differentiation stages, as shown in Figure 1.1 [3]. Therefore, there is a high chance of blood disorder occurrence. The most of blood disorders are observed in myeloid series that starts from myeloblast to band neutrophil, and erythroid series that starts from pronormoblast to erythrocyte, because these contain more maturation stages than other series.

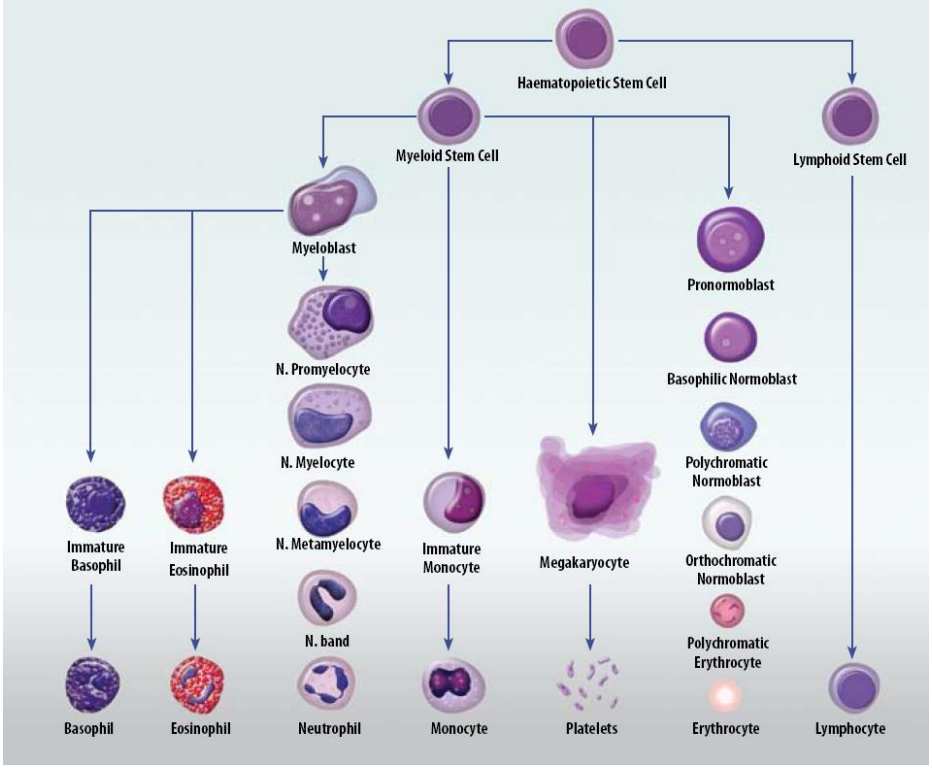

Figure 1.1 Diagram of hematopoiesis in humans.

Hematologic diseases include myeloproliferative disorder, lymphoproliferative disorder, leukemia, anemia, leukopenia, myelodysplastic syndromes, and lymphoma, and many others [4-8].

Among many hematologic disorders, blood cancers, including leukemia, lymphoma, myeloma, and myelodysplastic syndrome, are the most critical diseases that expected to account for 10.2% of all new cancer cases in the United States in 2016 [9]. An

exact cause of the diseases is still unknown, but the incidence rate of these blood cancer has continuously increased as people are more frequently exposed to harmful chemicals and radiation [10].

Lymphoma is blood disorder that results abnormal reproduction of lymphocytes in the lymphatic system. It can be divided into two large categories: Hodgkin lymphoma and non-Hodgkin lymphoma. Usually, non-Hodgkin lymphoma is more commonly occurred than Hodgkin lymphoma [7]. Multiple myeloma is a type of cancer that affects plasma cells [8].

Leukemia, the cancer of the leukocytes, is one of the severe blood disorder that results in high numbers of abnormal immature white blood cells in the bone marrow. There are four types of leukemia: acute myelogenous leukemia (AML), acute lymphoblastic leukemia (ALL), chronic myelogenous leukemia (CML), and chronic lymphoblastic leukemia (CLL) [6]. ALL mostly occur in children and is the most commonly diagnosed cancer among children between 0 and 14 years of age [11]. Other types of leukemia are usually common in adults. However, there exist many subtypes of leukemia due to several different

mutations throughout the multiple stages of hematopoiesis which make diagnosis difficult. Overall 5-year survival rate of leukemia in 2016 was 60.6% [9]. In more details, ALL, CLL, and CML has relatively high 5-year survival rate of 68.2%, 83.2%, and 66.9%, respectively. However, AML has 26.9% of 5-year survival rate, which is as low as that of stomach cancer and brain cancer [9]. According to a recent report from National Cancer Center of Republic of Korea, a relative 5-year survival rate of AML was 9.4% for 65-79 years old and 0% for age above 80 (Figure 1.2) [10]. Therefore, the leukemia still remains as the life threatening disease.

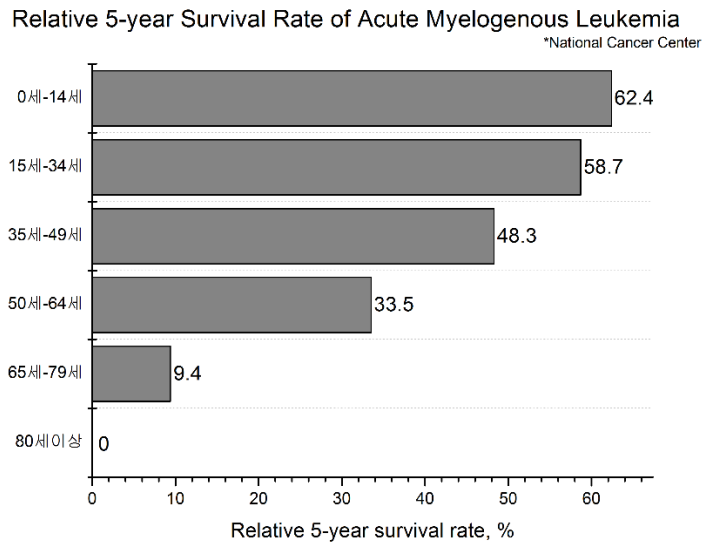

Figure 1.2 Graph of relative 5-year survival rate of acute myelogenous leukemia in Republic of Korea.

Since there exist numerous blood related diseases, diagnosis is important yet still remains to be difficult task. For diagnosis of hematologic diseases and follow-up of chemotherapy, multiple blood tests are conducted usually following three stages: complete blood count, peripheral blood smear analysis, and bone marrow analysis.

Briefly, complete blood count (CBC) is a common blood test that is initially conducted on patients with suspected blood problem [12]. It enumerates the normal blood cells, including red blood cells, leukocytes, and platelets, using automated hematology analyzers [13]. With advances in technologies, such as electrical impedance, radiofrequency conductivity, optical light scatter, cytochemistry, and fluorescent labeling, many manufacturers developed instruments with combinations of these methodology principles [13, 14]. An evolution of automated hematology analyzers have changed the process of modern hematology laboratories by providing accurate, precise, low-cost analysis with high throughput. Most of these automated hematology analyzers report five major leukocyte differential counts and alarm abnormal specimens for further analysis [13]. When a presence of immature granulocyte or the

abnormal leukocyte distribution is flagged from CBC, peripheral blood smear analysis is conducted for detailed analysis of blood cells [12].

Peripheral blood smear analysis is a fundamental modality in hematology laboratories because of the major implications it has for the diagnosis [15]. Peripheral blood smear analysis is based on blood cell differential count and cytology, which is morphological analysis of red blood cells, five major types of leukocytes, and platelets. Cytology is an important diagnostic criterion where unique features of cells, such as size, shape, and granularity, are assessed [2, 12, 16]. The examination is conducted manually under light microscope by expert hematologists [15]. Therefore, the process is not only time consuming and labor intensive, but also lacks reproducibility and objectivity [15]. Moreover, it is highly exposed to many errors such as intra- and inter-specimen variations, and it highly depends on experience of practitioner [12, 15]. Another main criticisms of the manual leukocyte differential count is its poor efficiency [15]. However, the manual examination is still considered as the gold standard. Therefore, guaranteeing high levels of consistency and maintaining proper quality control of

experts in hematology laboratories are important, yet difficult [12]. These problems have prompted research and development for the automated leukocyte morphology analyzers, and many manufactures have introduced automated morphologic analysis systems for peripheral blood smears, such as CellaVision DM96 [12, 15]. However, these are not widely accepted in clinical hematology laboratories and are not applicable for bone marrow aspirate smears [4].

If an abnormality in blood cells is observed from the peripheral blood smear analysis, a bone marrow analysis is conducted for further analysis. Since the peripheral blood smear analysis assesses the end point of hematopoiesis, the five major types of leukocytes, it is impossible to understand pathologic features without a knowledge of the maturation steps [4, 12]. Therefore, the bone marrow, the source of cells in the peripheral blood, has to be analyzed when abnormal morphology or observance of immature leukocytes in the peripheral blood specimen.

Bone marrow analysis is an essential and critical diagnostic modality in hematology laboratories for diagnosis of patients with not only hematologic disorders, but also with non-

hematologic disorders with unexplained hematologic conditions [2]. It is also conducted for follow-up of hematologic diseases and treatments, such as chemotherapy [4]. Fundamental mechanism of the procedure is similar to that of peripheral blood smear analysis, so the cytology of cells in bone marrow aspirates are assessed. However, the complex nature of bone marrow fluid and the many differences of its composition in comparison to peripheral blood increases difficulties of bone marrow analysis [4]. First, bone marrow aspirates contain many different types of cells in different phases of maturation, whereas only five major leukocytes are considered in peripheral blood smears. Bone marrow aspirates include blasts, promyelocytes, myelocytes, metamyelocytes, band neutrophil, segmented neutrophils, promonocytes, monocytes, eosinophils, basophils, mast cells, nucleated erythroid precursors, lymphocytes, and plasma cells [2, 4]. Since the hematopoiesis is continuous process, it is difficult to establish a discriminative standard. Moreover, bone marrow analysis needs to provide more complicated cytological information such as nuclear maturation and morphology, presence of nucleoli, cytoplasmic granularity, and inclusions. Also, it requires more enumeration,

200–500 consecutive cells, than the peripheral blood smear analysis [2]. Lastly, the problems raised for the peripheral blood smears are more severe in bone marrow aspirate smears. The bone marrow analysis is more variable, and more dependent on hematologists' experiences and knowledges, which make it more time consuming and labor intensive procedure. Therefore, only few expert hematologists are available and bone marrow analysis has been a bottleneck in diagnosis and treatment follow up of blood disorders. However, cytomorphological examination of bone marrow aspirate smears remains a well–established basic method to assess the state of hematological cell lineages, to diagnose hematologic disorders, and to evaluate treatment–related changes of the hematopoietic system. This irreplaceable visual identification and counting of normal and abnormal cell populations remains as a gold standard tool [12]. Technology has revolutionized procedures of clinical laboratories, so many diagnostic tests are currently conducted with highly automated systems. Almost all CBCs are carried out using automated hematology analyzers, and many instruments for automated peripheral blood analysis have been

introduced. However, bone marrow aspirate is less recognized and no significant changes have occurred [4, 12].

1.2. Introduction to Convolutional Neural Network

In the past few year, deep learning has dramatically improved as the state-of-the-art in visual object recognition, speech recognition, and object detection problems of many domains and applications. Among different algorithms of deep learning, the convolutional neural network (CNN) is a well-known deep learning algorithm which was inspired by the natural visual perception mechanism of the living creatures, and it has been most extensively studied for computer visions [11, 17].

The basis of CNN was introduced by LeCun et al. in 1990, which was the first paper on CNN trained by backpropagation [11, 18]. This predecessor of current CNN was structured with a multi-layer artificial neural network called LeNet-5. LeNet-5 was able to learn representations of handwritten digit images with the backpropagation algorithm and could recognize visual patterns directly from raw pixels of the original image with minimal or without pre-processing [18, 19]. However, CNN researches encountered a challenge at the time due to the lack of large training data and high computing power, which hindered training a large high-capacity CNN without overfitting. Later in

2006, CNN has been highlighted once again with introduction of big data, graphics processing unit (GPU), and algorithmic technique to training deep CNN [11, 17]. Moreover, novel CNN architectures were introduced such as AlexNet [20], ZFNet [21], VGGNet [22], GoogleNet [23], and ResNet [24], following a general trend of deeper networks [17]. These modified architectures share similar basic components: convolutional layer, pooling layer, fully connected layer, and activation function [11, 17]. The evolution of the architectures enabled approximation of the target function with increased nonlinearity and get better and more feature representations of images. In result, it showed significant improvements in various problems.

Focusing on biomedical images, CNN demonstrated astounding performance in mitosis detection from breast cancer pathology images [25, 26], in diagnosis of breast lesions in ultrasound images [27], in lung cancer detection and classification from CT scans [27, 28], in detection and classification of nuclei from colon cancer histology images [29], in detection of diabetic retinopathy from retinal fundus images [30], and many others [31].

Deep learning was able to overcome limitations of the conventional machine learning, which hindered from processing natural raw image data. Two major limitations of the conventional machine learning exist in segmentation process and feature extraction process. The conventional machine learning requires a segmentation of region-of-interest (ROI) from a raw image and a feature extraction from the ROI. Therefore, the performance of the conventional machine learning highly depends on segmentation, which is difficult to achieve high or nearly perfect performance. Moreover, the feature extractors are usually hand crafted, which is difficult to create and requires many insights and domain expertise. Therefore, only limited number of features are used in conventional machine learning. Unlike the conventional machine learning methods, deep learning discovers intricate structure in large raw data sets by using the backpropagation algorithm to change internal parameters that are used to compute the representation in each layer [18]. So, deep learning can be operated with natural raw images without error prone segmentation and hand crafted feature extraction processes.

In this thesis, CNN was used to classify multiples maturation stages of leukocytes with complex and indiscrete differential standards. Further details of CNN are described in Chapter 3.

1.3. Thesis Objectives

Currently, a commercial automated bone marrow analysis system does not exist and related researches show insufficient performance to be used in clinical hematology laboratories, which led to the main objectives of this thesis: development of an automated leukocyte differential count system for bone marrow aspirate smears. The proposed system would relieve from time and labor intensive procedure for hematologists, and would enable an inspection on increased number of leukocytes in a patient, which would provide a more reliable statistical information for diagnosis. Other potential advantages are increased throughput and efficiency, reduced costs, increased reproducibility, objectivity of measurements and less variability due to the individual variation in interpretation and the quality of the smear [4].

To this end, the system that satisfies following requirements was developed. The system should detect and classify leukocytes with high accuracy, sensitivity, and specificity. It also needed to overcome the complexity of bone marrow aspirate smear, which contains many touching leukocytes due

to high density of leukocyte, for the detection. Moreover, the system should be able to resolve uncertainty in large various and continuous maturation stages of leukocytes for the classification with high objectivity. The system also required a graphical user interface with fully automated operation for the convenience of users, and high throughput and efficiency.

Prior to the development of the proposed system, a large bone marrow smear dataset was collected and labeled by clinicians. For a classification algorithm, 2,323 small (96 × 96) single leukocyte image patches of ten leukocyte maturation stages were collected, and for a detection algorithm and the integrated system, 200 large (1388 × 1040) digital images of bone marrow aspirate smears were collected (Chapter 2).

Using the collected data, the classification algorithm, that classifies an original raw image of a single leukocyte into subclasses of leukocyte maturation stages, was developed by implementing the CNN. First, the dataset was oversampled and augmented to overcome data imbalance problem and to examine rotational and locational invariance of the trained network. Moreover, the dual-stage CNN was proposed to improve the

accuracy of classification by applying fine adjustments among classes that were highly misclassified (Chapter 3).

For the fully automated differential count system, the detection algorithm, that extracts locations of single leukocytes from a bone marrow aspirate smear image, was developed in addition to the classification algorithm. The proposed detection algorithm utilized saliency mapping and watershed transformation. The development of automated differential count system was completed by integrating these two algorithms. Moreover, a graphical user interface was designed for user convenience and a probability guidance algorithm was proposed to improve the overall performance of the system (Chapter 4).

CHAPTER 2

LEUKOCYTE DATA COLLECTION

Prior to the development of algorithms, a large dataset of bone marrow aspirate smears was collected. The dataset is an important ingredient in developing algorithms, especially for the classification algorithm which utilizes deep learning algorithm. It is widely known that performance of deep learning highly depends on the amount of data. However, any open accessible benchmark or public dataset is not available. As mentioned in the section 1.1, currently almost all differential count processes are manually conducted around the world, so the most of the bone marrow aspirate samples are not digitalized and saved. Therefore, a large dataset was collected for this study.

2.1. Sample Preparation and Acquisition

All bone marrow aspirate smear samples were prepared at the Department of Laboratory Medicine, Seoul National University Hospital. Bone marrow aspirate smears were prepared immediately following aspiration to reduce storage artefact. Smears are made with a glass spreader with beveled edges so that the width of the spreader is narrower than the width of the specimen slide. All slides were followed by staining using the Wright–Giemsa protocol. It yields purple nuclei, light blue

cytoplasm, white granules for leukocytes, and light pink color for red blood cells, as shown in Figure 2.1.

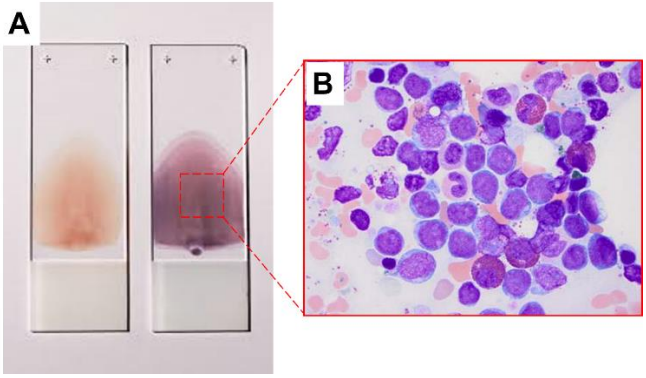

Figure 2.1 (A) Picture of bone marrow aspirate smear slide. (B) Example of leukocytes stained with Wright–Giemsa protocol.

Acquiring an image with high resolution, clarity, and accuracy and fair brightness is a complicated task due to the complex setting of acquisition system. Different quality of acquired images may be obtained from different place, by different people, and in different settings and environments. This affects the performance of image analysis programs that are highly depend on the quality of images. Possible solution to the issue can be steadily obtaining high quality of images with the same acquisition settings, but this is not feasible when the program is used in different facilities for different purposes. Therefore, the

program should be robust enough to cover this variability of image quality. In order to test the robustness of proposed system, there has been an effort to collect a data set with various conditions such as different cell density, color spectrum, and quality of image through acquiring images by experts and non-experts. For this thesis, all the images were acquired at the Department of Laboratory Medicine, Seoul National University Hospital. The acquisition setting is shown in Figure 2.2.

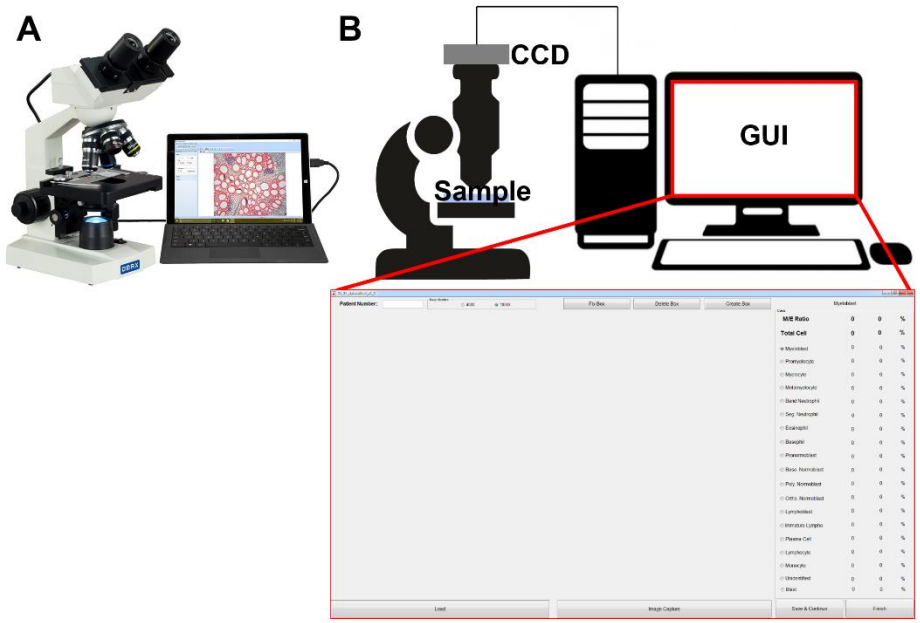

Figure 2.2 (A) Picture and (B) schematic of acquisition setup.

The acquisition setting contains an Olympus CCD camera that was mounted on a light microscope, Olympus CX31, with a magnification of ×1000. The camera with 80 megapixel yielded an image size of 1388 × 1040. For each slide, five to ten acquisition locations were randomly selected, and a total of 200 digitalized bone marrow aspirate smear images were obtained from 30 slides of 10 human subjects. The preparation of images were conducted anonymously, and qualities of all images such as details of features and contrast of the individual cells were confirmed by hematologists.

2.2. Dataset Collection and Preparation

The ground truth locations and classes were labeled from the acquired bone marrow aspirate smear images by hematologists. For the purpose, Graphical user interface (GUI) that saves coordinate indices and maturation stage labels of leukocytes were designed as shown in Figure 2.3. GUI can be easily operated by loading a bone marrow aspirate image and selecting leukocytes on a touchscreen monitor. GUI was implemented with MATLAB (Mathworks, USA).

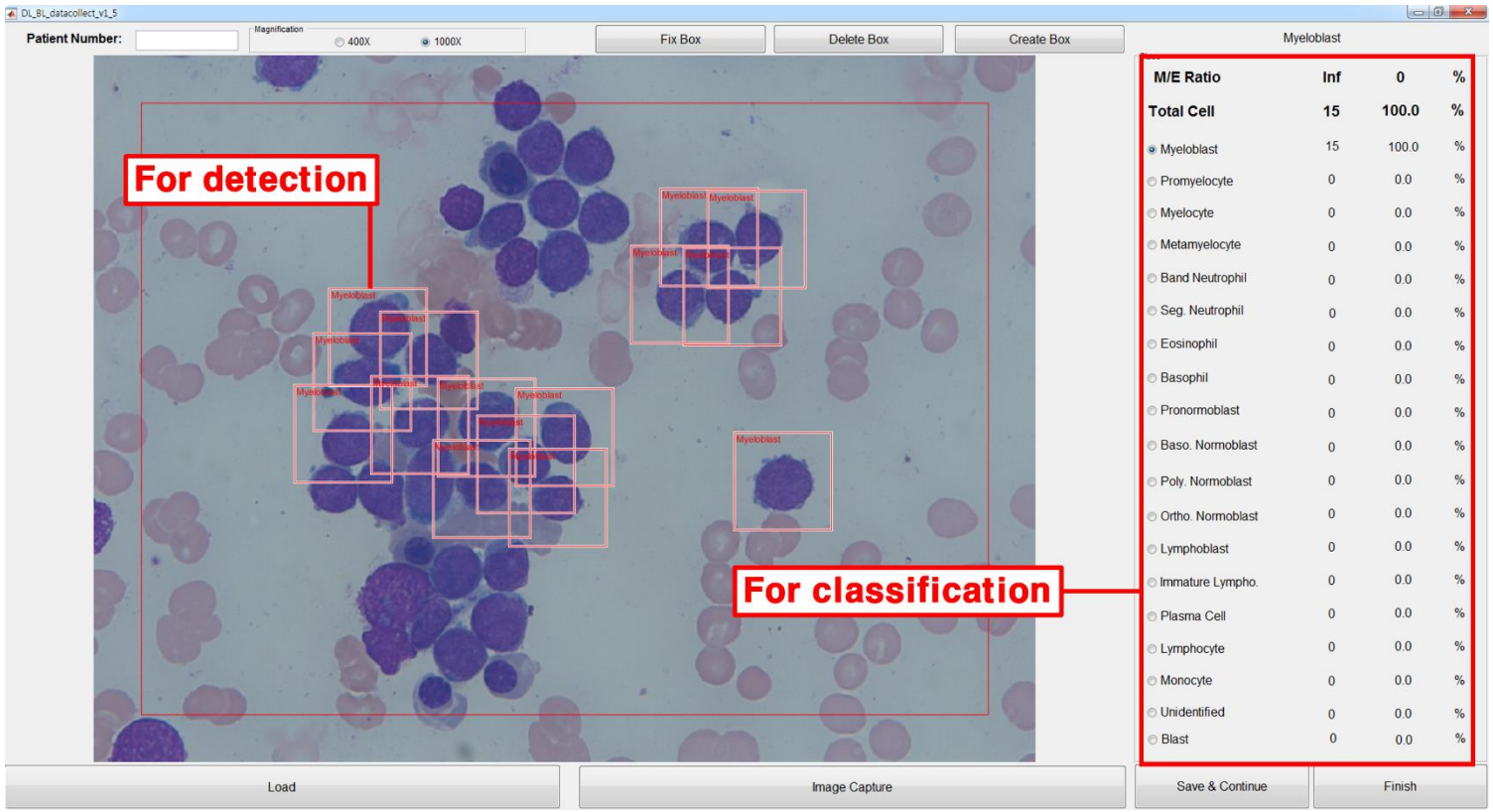

Figure 2.3 Screenshot of the MATLAB GUI for data collection.

The coordinate indices of a selected single leukocyte was saved and the area was cropped into 96 × 96−pixel single−cell patch images. The corresponding leukocyte maturation stage label among four consecutive stages of the erythroid series—pronormoblast (C1), basophilic normoblast (C2), polychromatic normoblast (C3), and orthochromatic normoblast (C4)—and six consecutive stages of the myeloid series—myeloblast (C5), promyelocyte (C6), myelocyte (C7), metamyelocyte (C8), band neutrophil (C9), and segmented neutrophil (C10), was selected and confirmed as having been placed in the correct classes by two expert hematologists. Figure 2.4 shows examples of collected single leukocyte images.

Total 200 of bone marrow aspirate smear images (1388 × 1040) were collected with coordinate indices, and these were used for developing the detection algorithm. Total 2,323 of single leukocyte images (96 × 96) were collected with corresponding labels among 10 maturation stages, and these were used for developing the classification algorithm.

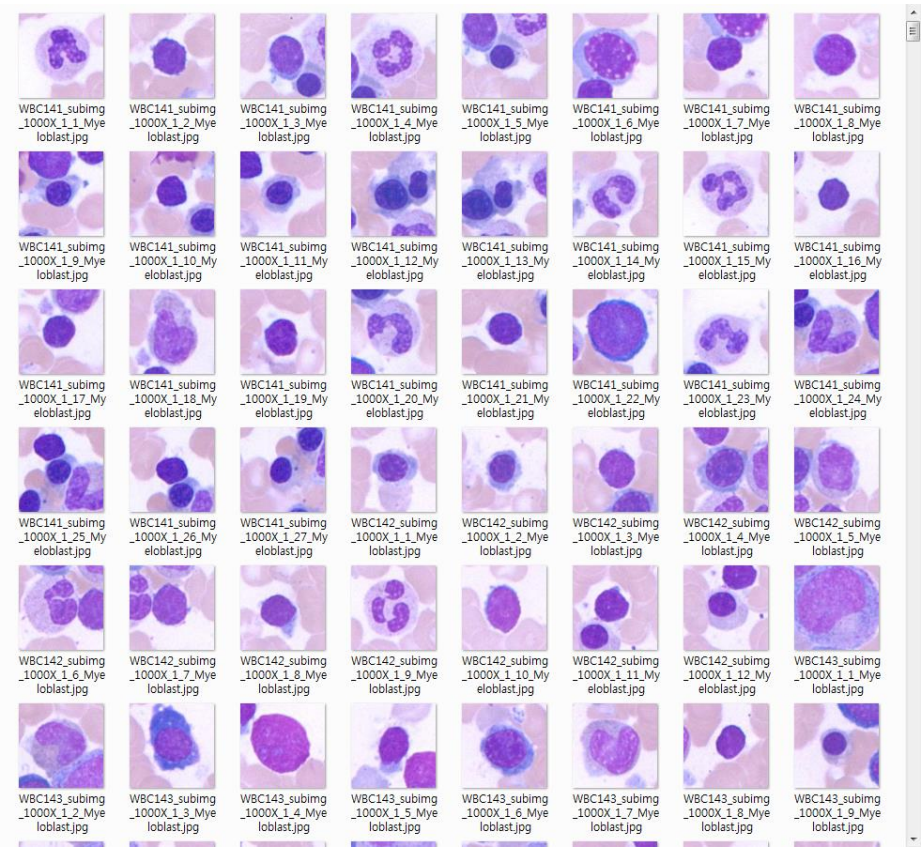

Figure 2.4 Examples of collected single leukocytes sub-images.

CHAPTER 3

LEUKOCYTE CLASSIFICATION

3.1. Introduction

The differential count of white blood cells (WBCs) is an essential examination in clinical hematology that is conducted on peripheral blood and bone marrow smears. Information obtained from these assessments is used for such purposes as the diagnosis of leukemia, lymphoma, myeloma, myeloproliferative neoplasm, and anemia, and for follow-up care after chemotherapy [32]. This important examination is still manually performed by trained hematologists. They assess the characteristics of cells, such as size, shape, and granularity, using a light microscope. Therefore, the process is not only tedious and labor intensive, but also vulnerable to many sources of error. Intra- and inter-cell variations exist because the morphological characteristics of cells differ within a patient and among patients. Image properties, such as color and contrast, also vary among samples due to the methods used for staining as well as the quality of image acquisition. These make it difficult to obtain an accurate count of WBCs. Since the results are qualitative and highly dependent on the hematologist's skill and experience, variations within the results obtained by a

hematologist, as well as those among measurements by several experts, are unavoidable [33, 34]. In order to solve these problems, a quantified automated analysis system is highly demanded [34–36].

A number of studies have been conducted on automated WBC differentiation in a peripheral blood smear, and commercial computer-aided diagnosis (CAD) systems are available for this purpose [34, 37]. However, an automated WBC differential count of bone marrow smears is problematic and has not been sufficiently researched. Classification of WBCs in bone marrow smears is complex and challenging. In peripheral blood smears, five fully maturated WBCs—basophil, eosinophil, segmented neutrophil, monocyte, and lymphocyte—are usually observed and analyzed. These WBC types have distinct characteristics, so they are relatively easier to discriminate. However, bone marrow smears are typically used to consider the maturation stages of the WBCs. These stages involve more cell types, such as myeloblast, promyelocyte, myelocyte, metamyelocyte, band neutrophil, segmented neutrophil, pronormoblast, basophilic normoblasts, polychromatic normoblast, orthochromatic normoblast, lymphoblast, lymphocyte, monocyte, basophil,

eosinophil, and plasma cell. In the diagnosis of hematologic diseases, knowing the ratio of these immature and mature cell types is necessary [37, 38]. Not only do more types of cells need to be discriminated, these stages of maturation are also challenging in the context of defining discrete standards for each cell type, because small inter-class differences exist among continuous stages [39]. Moreover, the cell density of WBCs in the bone marrow smears is higher than that in peripheral blood smears. Due to the high density of bone marrow smears, many WBCs touch one another, which makes it difficult to segment single cells. This is critical in developing an automated WBC differential counter using image processing and traditional machine learning methods, since single-cell segmentation is required for feature extraction and classification [34, 40]. Despite the importance of bone marrow analysis and the high demand for a quantified automated bone marrow smear analyzer, these difficulties hinder progress, and have only been addressed in a few studies.

Attempts have been made in research to classify WBCs in bone marrow smears using image processing and machine learning algorithms. These traditional methods follow the sequence of

segmentation, feature extraction, and classification. Many such studies only focused on single-cell segmentation from a bone marrow smear image. However, relatively few studies have been devoted to the classification of the WBCs in bone marrow images. In Theera-Umpon et al. [37], WBCs in bone marrow smear images were classified using the morphological granulomere of the nucleus for six myeloid series. The study used Bayes and artificial neural network (ANN) classifiers on four extracted features from the nucleus. This approach achieved an accuracy of 63.3% and 65.7% using Bayes and ANN, respectively. It attempts to overcome the problem of touching cells in high-density bone marrow smears by segmenting only the nuclei. This reduced the error rate in cell segmentation, but led to many clinically significant features of the cytoplasm being discarded, hence yielding insufficient classification accuracy. Moreover, the work of Reta et al. [40] showed that features extracted from the nucleus and the cytoplasm are powerful. Several other studies used different methods, such as the multilayer perceptron, the support vector machine (SVM), and feed-forward neural networks for classification [41-46]. Osowski et al. [39] increased the

classification accuracy to 83.2% on 11 classes of white blood cells using an SVM and a genetic algorithm. More recently, Staroszczyk et al. [47] showed that using an ensemble of classifiers is more effective to this end. A model combining different feature selection methods and an SVM for an ensemble was proposed, and improved the accuracy to 85.7%. Attempts have also been made to use a CAD system for classifying WBCs of different maturation stages [48–51]. Lee et al. [48] and Briggs et al. [50] used CellaVision DM 96 (CellaVision AB, Sweden) to classify six maturation stages of the myeloid series: blasts, promyelocytes, myelocytes, metamyelocytes, band neutrophils, and segmented neutrophils. In these studies, the CAD system yielded a correlation of 0.86 and 0.74, respectively, with results obtained by expert hematologists. The CAD system was much quicker than the experts, but its performance was unsatisfactory. These studies have been unable to solve the problem of touching cells in high-density bone marrow smears, and have not achieved a desirable accuracy for WBC classification in multiple stages of maturation. These limitations of traditional image processing and machine learning methods are difficult to overcome because

of their low learning capacities and their use of handcrafted features.

Deep learning has been spotlighted in machine learning research due to advancements in parallel computation using graphics processing units (GPUs), large datasets, and algorithms, the essential ingredients of deep learning. In contrast to traditional image processing and machine learning, deep learning algorithms incorporate feature extraction and classification. Therefore, it can be applied to raw data with minimal pre-processing, such as mean and standard deviation normalization of datasets; deep learning algorithms can also learn more features than handcrafted methods. Deep learning has shown outstanding performance in the classification and recognition of images and signals [20, 52-54]. Its application to medical images has only recently been actively studied. However, the number is rapidly increasing and it is demonstrating astounding performance in various applications. Kainz et al. studied bone marrow cell classification using deep learning, which is the only previous research that applied deep CNN in white blood cell classification [55]. They proposed rotation-invariant WBC classification on a raw image using a

recurrent neural network. Approximately 157 images were augmented to 944 images in five classes (four WBC types and one background) for training and testing. The method achieved an accuracy of 96.4% but had some limitations. The background class, which achieved 100% accuracy, should be excluded from WBC classification accuracy for fair comparison, since background images are distinct from WBC images. Moreover, this study tried to discriminate WBCs in the maturation stages, but the number of cell types was not sufficiently large for clinical practice. Further, it required a long processing time and a large amount of computation as the images had to be rotated at every degree. Time and computational effort could have been saved by reducing the frequency of rotation, but this would have come at the cost of classification accuracy.

In this study, we propose an automatic bone marrow WBC differential counter using a convolutional neural network (CNN). The CNN is a deep learning algorithm that has shown strong performance in image recognition and classification [20, 56]. With sparse interactions, parameter sharing, and equivalent representation, the CNN can learn multi-level features from minimally processed raw data and detect complex interactions

among the features. Therefore, we aim to exploit the CNN to classify WBCs of the myeloid and erythroid series. The proposed method does not require single-cell segmentation or hand-crafted feature extraction. WBC images were collected and labeled, and trained and tested to classify 10 WBC cell types in different stages of maturation. The proposed method also solves the problem of imbalanced data through oversampling and augmentation and improves classification performance through the dual-stage network of a global and a local model.

3.2. Methods

3.2.1. Data Description

The dataset was composed of 2,323 single leukocyte images of 10 WBC classes in the stages of maturation, including four consecutive stages of the erythroid series—pronormoblast (C1), basophilic normoblast (C2), polychromatic normoblast (C3), and orthochromatic normoblast (C4)—and six consecutive stages of the myeloid series—myeloblast (C5), promyelocyte (C6), myelocyte (C7), metamyelocyte (C8), band neutrophil (C9), and segmented neutrophil (C10). The examples of WBCs in the stages of maturation are shown in Figure 3.1A. The cells in the same series had small inter-class differences due to the continuous maturation process.

A number of images for each class is shown in Figure 3.1B. The collected dataset had an imbalanced distribution among classes. This problem was unavoidable while collecting the dataset, since the natural distribution of white blood cells is imbalanced, as shown in Figure 3.1C.

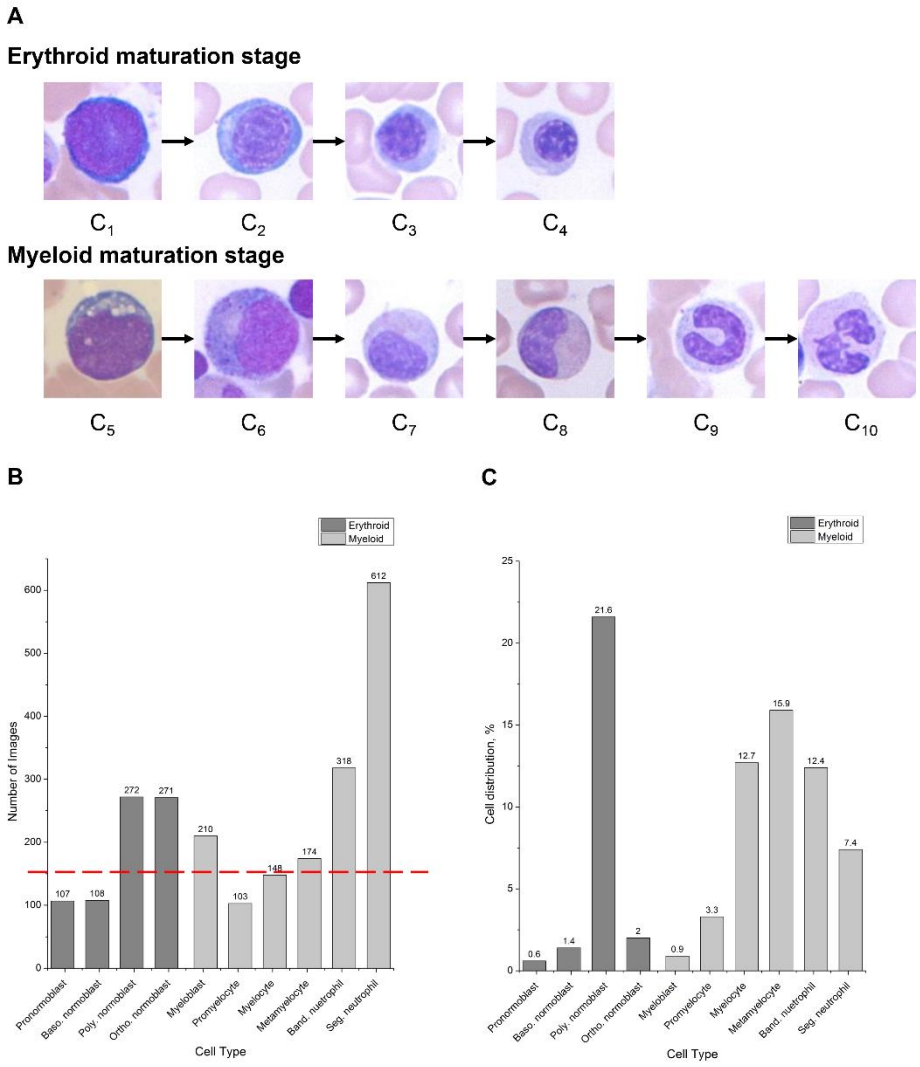

Figure 3.1 Description of collected data. (A) Examples of white blood cells in erythroid series (C1–4) and myeloid series (C5–10). (B) Distribution of collected data. (C) Cellular component distribution in bone marrow.

3.2.2. Data Oversampling and Augmentation

The imbalanced dataset can be a problem in training the network, as only a few features are learned from classes with relatively small number of data [57]. In an attempt to resolve the problem of imbalanced data due to the heterogeneous distribution of white blood cells, oversampling was conducted for classes with relatively small numbers of data items. Usually an oversampling indicates a duplication of same data, but the oversampling of the study was conducted in slightly different manner. During data preparation, images of the same cells at slightly different centers for classes with few data were cropped to increase the number of dataset. This increased the number of data and provided more diverse data. Examples of oversampled images are shown in Figure 3.2A.

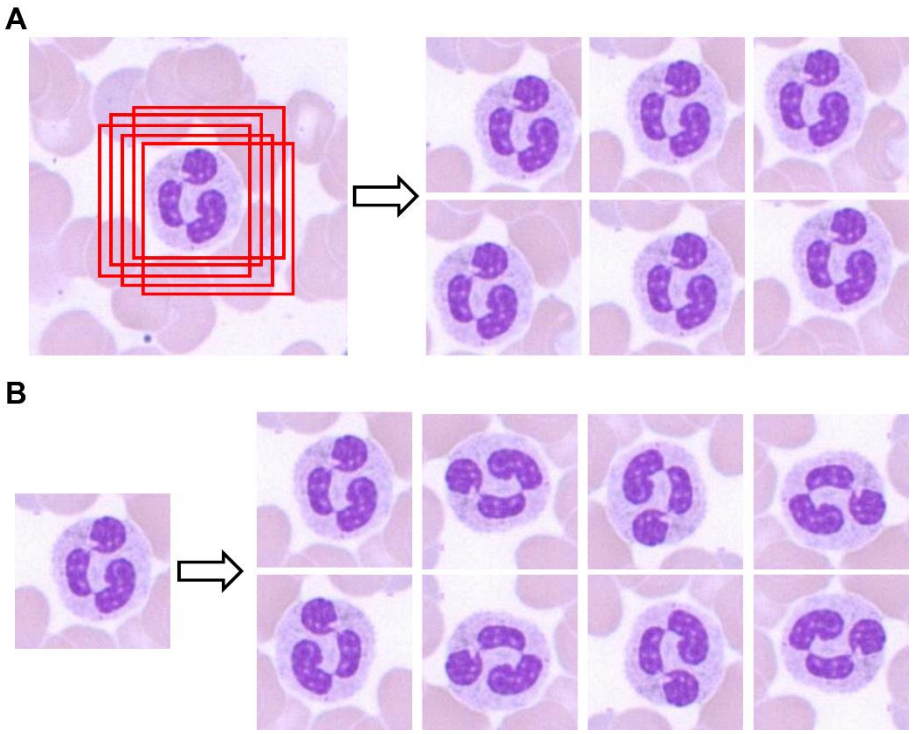

Figure 3.2 Examples of data preparation. (A) Oversampling and (B) Augmentation.

In order to solve the problem of the rotation variation in white blood cell classification, the dataset was augmented by a factor of eight. Image patches were transformed through combinations of four angles of rotation (0°, 90°, 180°, 270°) and two flips (horizontal and vertical). We ensured that no duplication existed between the original and augmented images. Augmentation solves the problem of rotation variation and

increases the number of data to help network training. Examples of augmented images are shown in Figure 3.2B.

3.2.3. Convolutional Neural Network Architecture and Dual-stage Convolutional Neural Network

The architecture of the CNN was inspired by VGGnet developed by Simonyan et al. [22]. The network used in this study was composed of 16 layers, 13 convolutional layers with max-pooling, and three fully connected layers followed by a softmax classifier layer. The dimensions of the input were set to 96×96×3 and 3×3 filters were used for the convolutional layers. The max-pooling layers were operated in 2×2 regions with a stride of 2. In order to prevent overfitting of the trained network, a dropout [58] was placed between all pairs of convolutional layers. The dropout randomly deactivated some weights in the convolutional layer; the dropout ratio was set in the range 0.3 to 0.5. Batch normalization was implemented after the convolutional layers with a batch size of 40 to prevent overfitting. Moreover, a rectified linear-unit (ReLU) activation function [20] was used following each instance of batch

normalization for effective learning and fast convergence. The network architecture is shown in Figure 3.3A.

We propose a dual-stage CNN with a global and a local model. The last fully connected layer of the local model was changed from ten to two. A global model of 10 classes and a local model of two classes were trained and combined, as shown in Figure 3.3B, for testing. This customized architecture for the WBC differential count was designed to fine-tune cases where the global model incorrectly classifies two consecutive maturation stages. In this study, we trained the local model for the band neutrophils and segmented neutrophils classes, which are difficult to classify due to their complex nuclear shapes and granularities. It is widely acknowledged that band cells cannot be reliably distinguished from segmented neutrophils on blood film by human observers [15, 59]

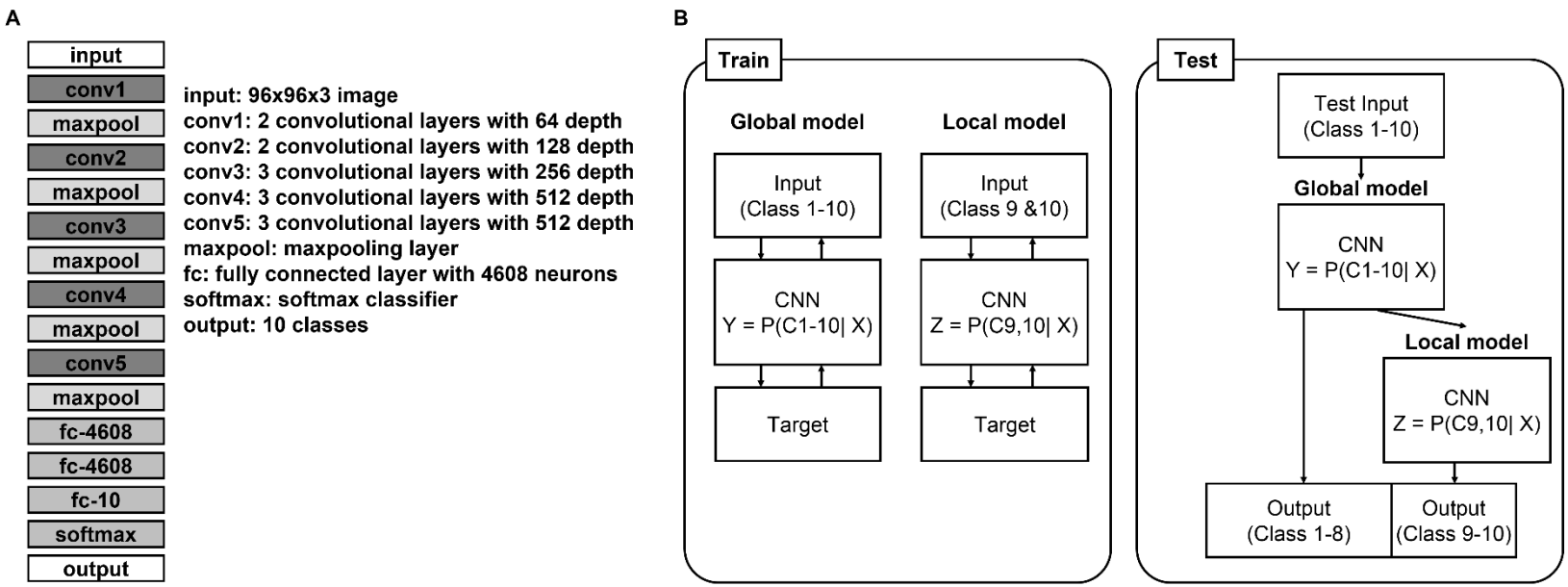

Figure 3.3 Description of networks. (A) Illustration of the convolutional neural network. (B) Description of the proposed dual-stage convolutional neural network.

3.2.4. Convolutional Neural Network Training

The weights of each layer were initialized using the MSRA method [60]. It randomly assigns weights from a zero-mean Gaussian distribution with a standard deviation of $\sqrt{2/N_{in}}$, where Nin is the number of inputs to a neuron. Following initialization, the network was trained and optimized through stochastic gradient descent to minimize the cross-entropy loss function:

$$\text{Loss}: \mathcal{L} = \frac{1}{N}\sum_i \mathrm{D}(S_i, L_i) \qquad (3.1)$$

where N is a total number of images, and i is the i-th single image of a trained dataset. S is an output vector of the softmax classifier that assigns probabilities of classes between 0 and 1, and L is a label vector of the image, where 1 is assigned for a correct assignment to a class and 0 for incorrect assignments. The cross-entropy of each image was computed as follows:

$$\text{Cross entropy}: \mathrm{D}(\mathrm{S}, \mathrm{L}) = -\sum_c L_c \log S_c$$

$$\text{Softmax}: \mathrm{S}(y_c) = \frac{e^{y_c}}{\sum_j e^{y_j}} \qquad (3.2)$$

where c denotes a class among many, y denotes a vector of the class output scores of a network, and j also denotes classes.

The prepared image dataset was divided into training-validation and testing datasets with a ratio of 5:1. The training-validation dataset was further divided into a ratio of 4:1 ratio for five-fold cross-validation, and was used to train and optimize the hyper-parameters: learning ratio, momentum, learning rate decay, and weight decay. The images of each dataset were transformed from the RGB color channel to YUV. The training dataset was normalized using the mean and standard deviation of each channel. The mean and standard deviation of the training dataset were recorded to normalize the validation and test datasets. The networks were trained from scratch without any pre-training for 150 epochs.

To assess the effects of the number of data, the augmentation, and oversampling, the network was trained using six datasets; the original dataset, the augmented dataset, the dataset oversampled to 300 images per class, oversampled to 600 images per class, the dataset augmented and oversampled to 300 images per class, and that augmented and oversampled to

600 images per class. These datasets contained 2,323, 18,584, 3,000, 6,000, 24,000, and 48000 images, respectively. For these cases, only the training dataset was augmented and/or oversampled, and the test dataset was used only for performance evaluation. To analyze the location invariance and the rotation invariance of the trained networks, randomly selected images from the band neutrophil and segmented neutrophil test datasets were oversampled and augmented.

In order to further analyze the effects of augmentation and oversampling, additional following datasets were examined: original (150 samples per class), oversample ×2, augmentation ×2 (300 samples per class), oversample ×4, augmentation ×4, combination over. ×2 + aug. ×2 (600 samples per class), combination over. ×2 + aug. ×4, combination over. ×4 + aug. ×2 (1200 samples per class), combination over. ×2 + aug. ×8, combination over. ×4 + aug. ×4, combination over. ×8 + aug. ×2 (2400 samples per class), combination over. ×4 + aug. ×8, combination over. ×8 + aug. ×4 (4800 samples per class), and combination over. ×8 + aug. ×8 (9600 samples per class).

Moreover, the dual-stage CNN was constructed using the datasets augmented and oversampled to 600 images per class, which yielded the best performance, as the global model and the local model of classes 9 and 10, which corresponds to band neutrophil and segmented neutrophil. These two maturation stages were selected for training the local model, because it is widely acknowledged that band cells cannot be reliably distinguished from segmented neutrophils on blood films by human observers [15, 59].

3.2.5. Implementation

The proposed method was implemented using the Torch7 framework and the CUDA toolkit with the cuDNN library on Linux OS. All experiments were performed with a CPU i7-6700 (3.40 GHz), RAM 16 GB, and GPU NVIDIA GTX 980 (4 GB).

3.2.6. Evaluation Metrics

The trained network was tested on the test dataset and classification performance was assessed quantitatively through

the following metrics: mean accuracy, precision, recall, and F1 score,

$$\text{Accuracy} = \frac{T_P+T_N}{T_P+F_P+T_N+F_N},$$

$$\text{Precision or Positive predictive value} = \frac{T_P}{T_P+F_P},$$

$$\text{Recall or Sensitivity} = \frac{T_P}{T_P+F_N},$$

$$\text{F1 score} = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (3.3)$$

where T_P is the number of true positive classifications, T_N the number of true negatives, F_P is the number false positive classifications, and F_N the number of false negatives. A confusion matrix of the classes was also created to analyze class-wise performance.

3.3. Results and Discussion

Experiments were conducted on the six datasets, and the classification performance of the proposed method was compared for the six datasets according to the above evaluation metrics. The classification performance of the network trained on original data (OG network), augmented data (AG network), oversampled data with 300 images per class (OS 300 network), oversampled data with 600 images per class (OS 600 network), augmented and oversampled data with 300 images per class (AG+OS 300 network), and augmented and oversampled data with 600 images per class (AG+OS 600 network) is summarized in Table 3.1.

Table 3.1 Classification performance of the network trained on different datasets.

Dataset	Accuracy	Precision	Recall	F1 score
Original	57.80	83.36	48.69	61.47
Augmentation	71.90	65.17	65.00	65.08
Oversampling 300	65.62	68.59	65.61	67.07
Oversampling 600	90.57	91.04	90.57	90.80
Augmentation + Oversampling 300	85.05	85.02	85.05	85.04
Augmentation + Oversampling 600	95.68	95.49	95.68	95.58

The accuracy of the OG network yielded the worst accuracy at 57.8%, but the AG network and the OS 300 network recorded slightly higher accuracy values of 71.9% and 65.62%,

respectively. The OS 600 network and the AG+OS 300 network significantly improved classification performance with accuracies of 90.57% and 85.05%, respectively. The AG+OS 600 network demonstrated the best performance with a 95.68% accuracy, which was as good as the state-of-art method of Kainz et al. It is widely accepted that deep learning algorithms train better networks with larger amounts of data, and the result showed a similar trend in general, except for the OS 600 network. This network, which was trained on 6,000 images, showed better performance to the AG+OS 300 and the AG networks, which were trained on 24,000 and 18,584 images, respectively. This indicates that the diversity of data, in addition to the total number of data, is also important in training the CNN for WBC differential count.

To confirm the statement, further analysis on effects of oversampling and augmentation was conducted. The results are summarized in Table 3.2 and plotted as shown in Figure 3.4.

Table 3.2 Effects of oversampling and augmentation on the classification performance.

Original		
N = 1500 (150 per class)	Original	Accuracy: 79.2%

	Oversampling		Augmentation	
N = 3000 (300 per class)	Over. x2	Accuracy: 89.1%	Aug. x2	Accuracy: 89%
N = 6000 (600 per class)	Over. x4	Accuracy: 92.5%	Aug. x4	Accuracy: 92.1%
N = 12000 (1200 per class)	Over. x8	Accuracy: 94.8%	Aug. x8	Accuracy: 94%

(Continued)

	Combination	
N = 6000 (600 per class)	Comb. Ox2 + Ax2	Accuracy: 90%
N = 12000 (1200 per class)	Comb. Ox2 + Ax4	Accuracy: 92%
	Comb. Ox4 + Ax2	Accuracy: 93.8%
N = 24000 (2400 per class)	Comb. Ox2 + Ax8	Accuracy: 92.9%
	Comb. Ox4 + Ax4	Accuracy: 95.9%
	Comb. Ox8 + Ax2	Accuracy: 95.2%
N = 48000 (4800 per class)	Comb. Ox4 + Ax8	Accuracy: 95.8%
	Comb. Ox8 + Ax4	Accuracy: 96%
N = 96000 (9600 per class)	Comb. Ox8 + Ax8	Accuracy: 96%

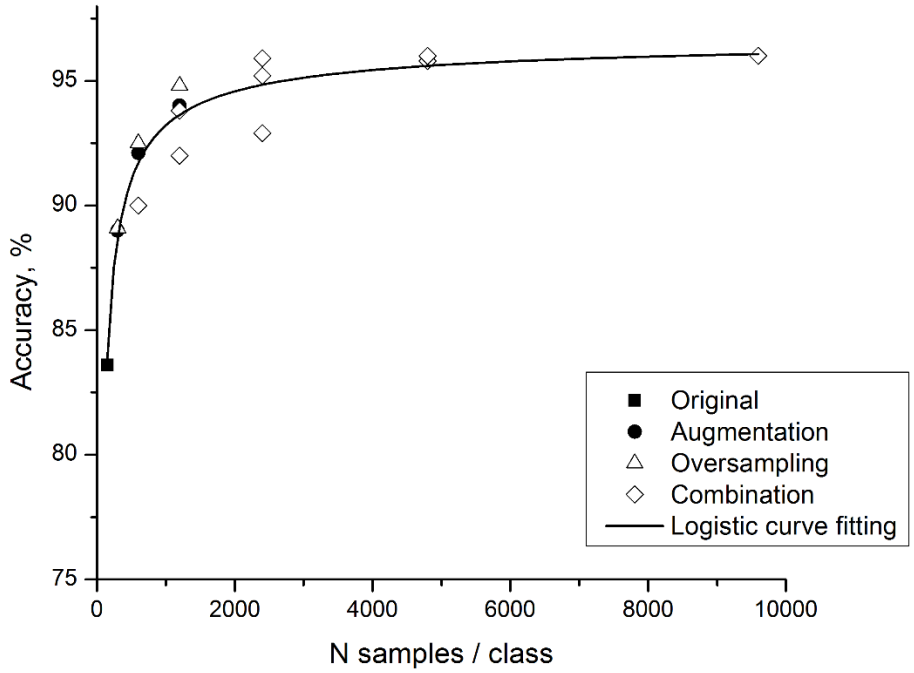

Figure 3.4 Graph of classification accuracy demonstrating the effects of oversampling and augmentation.

This experiment was conducted on the different train and test dataset from the same data pool, so the accuracy was slightly different from those of Table 3.1. However, it showed the general trend of increased performance with increased number of dataset. By directly comparing the effects of oversampling and augmentation on the equal number of data per class, the oversampling showed slightly better performance than the augmentation. This was also observed in the combination

dataset. For example, the combination over. ×8 + aug. ×2 showed higher classification accuracy than that of the combination over. ×2 + aug. ×8, which contained the same amount of data. So it had confirmed that the oversampling is more effective than the augmentation. The classification accuracy reached plateau above 2400 samples per class.

Most cell types have round-shaped cells and nuclei; thus, the rotational augmentation of images was less effective in creating diverse data than oversampling. Hence, while augmentation is widely used to increase the number of data in deep learning in general, this may not be effective or practical for WBC images. Rescaling augmentation cannot be applied to WBC classification because the size of cell is an important characteristic. Moreover, in other medical images, even rotational augmentation is not acceptable, since the orientations of images represent important information. However, augmentation was able to improve performance with the same number of oversampled data. The precision, the recall, and the F1 score also showed similar trends as accuracy, such that the AG+OS 600 recorded the best performance. The problems of an imbalanced dataset and

low performance were solved through data oversampling and augmentation.

Further analysis was conducted on the AG+OS 600 network. To confirm that it had been successfully trained, the accuracy of the validation dataset and training loss were plotted over 150 epochs (Figure 3.5A). Accuracy converged after approximately 100 epochs, when the training loss also converged to 0.004. The final trained network achieved 99.7% accuracy on the training dataset and 96% on the validation dataset. A confusion matrix of the classification results on the test dataset was generated to evaluate class-wise performance (Figure 3.5B). The first four classes represent WBCs of the erythroid maturation series and the last six the WBCs of the myeloid maturation series. As shown in the confusion matrix, misclassifications occurred within the same maturation series, mostly within the consecutive maturation stage. A majority of misclassifications occurred in the band neutrophil and the segmented neutrophil, which achieved 89% and 85% in terms of recall, respectively.

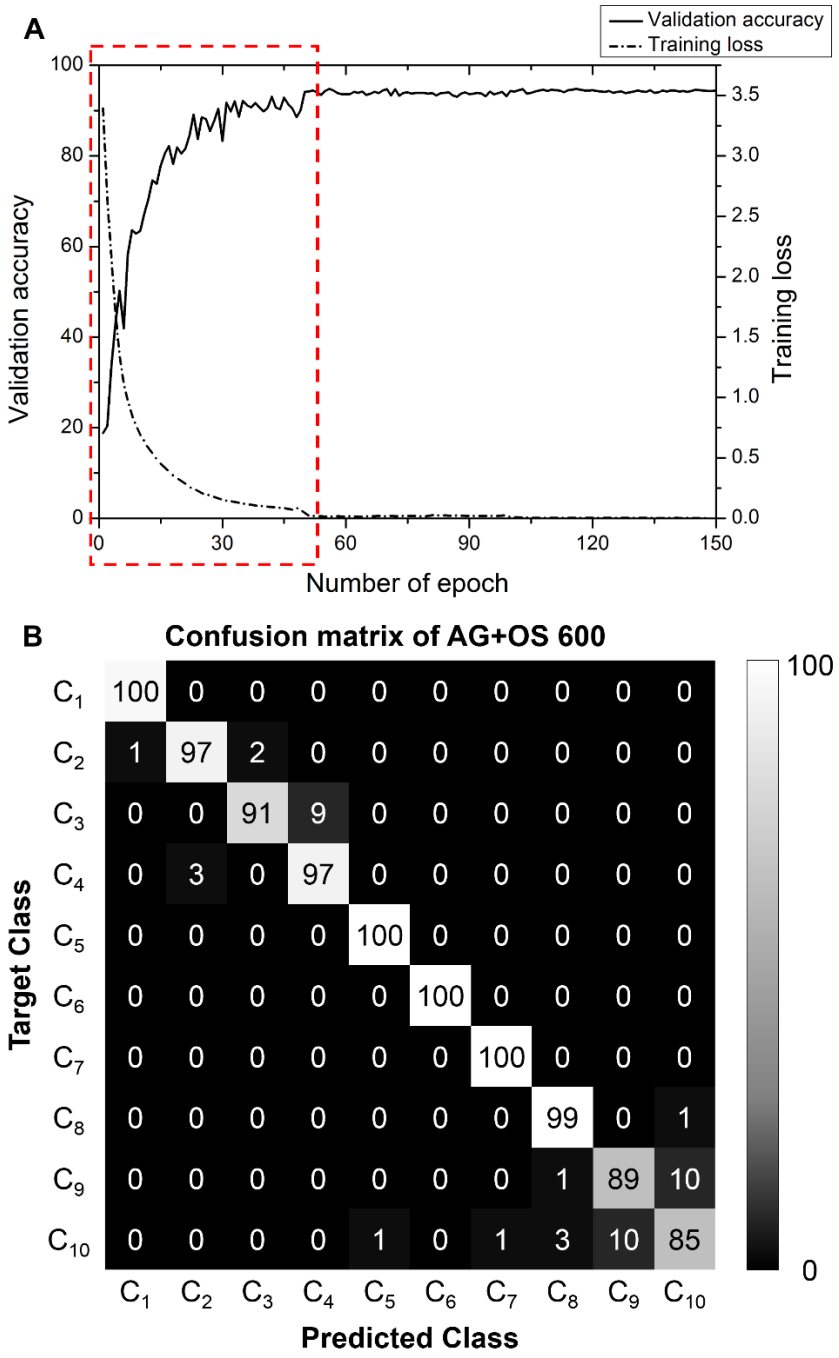

Figure 3.5 Details of training networks. (A) Graph of validation accuracy and training loss during training of network. (B) Confusion matrix of AG+OS 600.

The classification of band neutrophil and segmented neutrophil was explored in greater detail on the AG+OS 600 network. Figure 3.6 shows the correctly classified images of the band neutrophil and the segmented neutrophil. The network was able to classify images with backgrounds, as shown in Figure 3.6A. These images not only contained red blood cells and parts of other WBCs in the background, but also featured cells touching the target cells. This result indicates that the network was background invariant, and can be applied to raw images without WBC segmentation, which is a problem in traditional image processing and machine learning methods. Moreover, a randomly selected image from the band neutrophil and the segmented neutrophil test datasets was oversampled and tested to assess the network's location invariance (Figure 3.6B). The AG+OS 600 network was able to correctly classify the oversampled images. To ensure that this was due to data oversampling, we tested the same oversampled images on other networks. The AG network was not able to classify most of these oversampled images, whereas the OS network correctly classified a few. This location-invariant network can help develop a combined system of detection and classification. If

the detection algorithm slightly misses the center of the cell, the trained model can still deliver good performance. The augmented images were also assessed to show that the network was rotation invariant. A randomly selected image from the band neutrophil and the segmented neutrophil test datasets was augmented and tested (Figure 3.6C). The AG+OS 600 network was able to correctly classify the augmented images. To ensure that this was the consequence of data augmentation, we tested the same augmented images on other networks. The OS network was not able to classify most augmented images, whereas the AG network correctly classified a few. We can thus conclude that data augmentation is effective in training a rotation-invariant network.

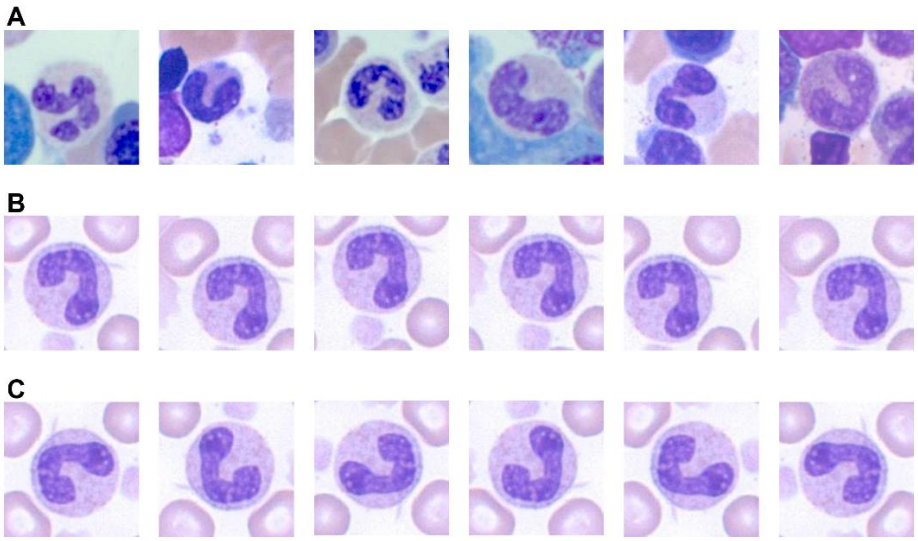

Figure 3.6 Examples of correctly classified cells by the AG+OS 600 network. (A) WBCs with backgrounds showing background invariance of the network. (B) Oversampled WBCs showing location invariance of the network. (C) Augmented WBCs showing rotation invariance of the network.

Incorrectly classified images of the band neutrophil and the segmented neutrophils were also analyzed. Some cases are shown in Figure 3.7 with probability distribution results from the softmax classifier. The deep learning algorithm predicted the classes according to the probability of the softmax classifier. The correct class in the left column of Figure 3.7 is band neutrophil; yet, the network predicted these images as belonging to the segmented neutrophil. The correct class on the

right column of Figure 3.7 is the segmented neutrophil; yet, the network predicted the images as belonging to the band neutrophil. Although, the data collection was carefully conducted over almost a year by two experts who were trained over 7 years, the process of data collection is tedious and error-prone, so these misclassified cells were confirmed by a third expert. There were controversies on few cases, for example, the third expert had different opinion on the last two rows of both band neutrophil and segmented neutrophil in Figure 3.7C-D and G-H. The person commented that these cells are on a board between band neutrophil and segmented neutrophil. Specifically, the cell of Figure 3.7C should be labeled as segmented neutrophil and the cell of Figure 3.7G should be labeled as band neutrophil. The expert was uncertain about Figure 3.7D and Figure 3.7H. However, the maturation of white blood cell is a continuous process, so it is difficult to provide discrete discriminative standards and there may exist different opinions. This demonstrates the issue of differential count, which highly depends on the experts' opinions and experiences.

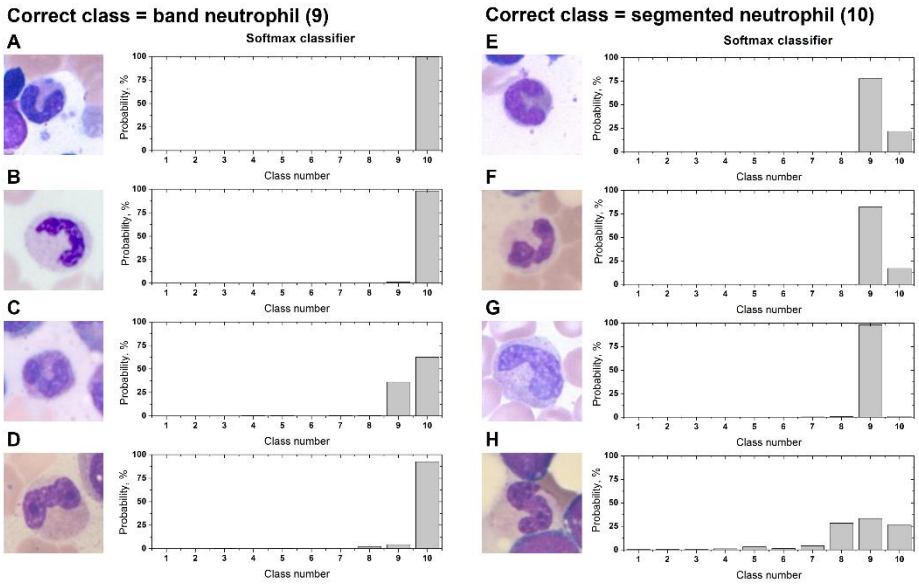

Figure 3.7 Examples of incorrectly classified leukocytes (band neutrophil and segmented neutrophil) by the AG+OS 600 network and their confidence values. (A)－(D) Cells whose ground truth is band neutrophil. (E)－(H) Cells whose ground truth is segmented Neutrophil.

The confusion matrix of the dual－stage CNN classification for the 10 classes considered on the test dataset is shown in Figure 3.8. The dual－stage CNN was able to correct the misclassifications in the band neutrophil and the segmented neutrophil of the AG+OS 600 network as shown in the red dotted boxes of Figure 3.8. The dual－stage CNN achieved a 97.06% accuracy, a 97.13% precision, a 97.06% recall, and a

97.1% F1 score. It thus outperformed any previously reported research for cases involving a large number of WBC types, which had hitherto yielded accuracy values below 90%. It even surpassed the 96.4% accuracy of the state-of-the-art method proposed by Kainz et al, which only classified four WBC types. Moreover, the dual-stage CNN classified images with backgrounds, augmented images, and oversampled images in Figure 3.6 as well, which indicates that it is rotation and location invariant, and can be applied to raw images. Since the classification evaluation metrics are not familiar to all readers, average specificity and sensitivity (recall) of each class were calculated and provided for convenience to understand the results. The dual-stage CNN achieved 99.73% of average specificity and 97.06% of sensitivity.

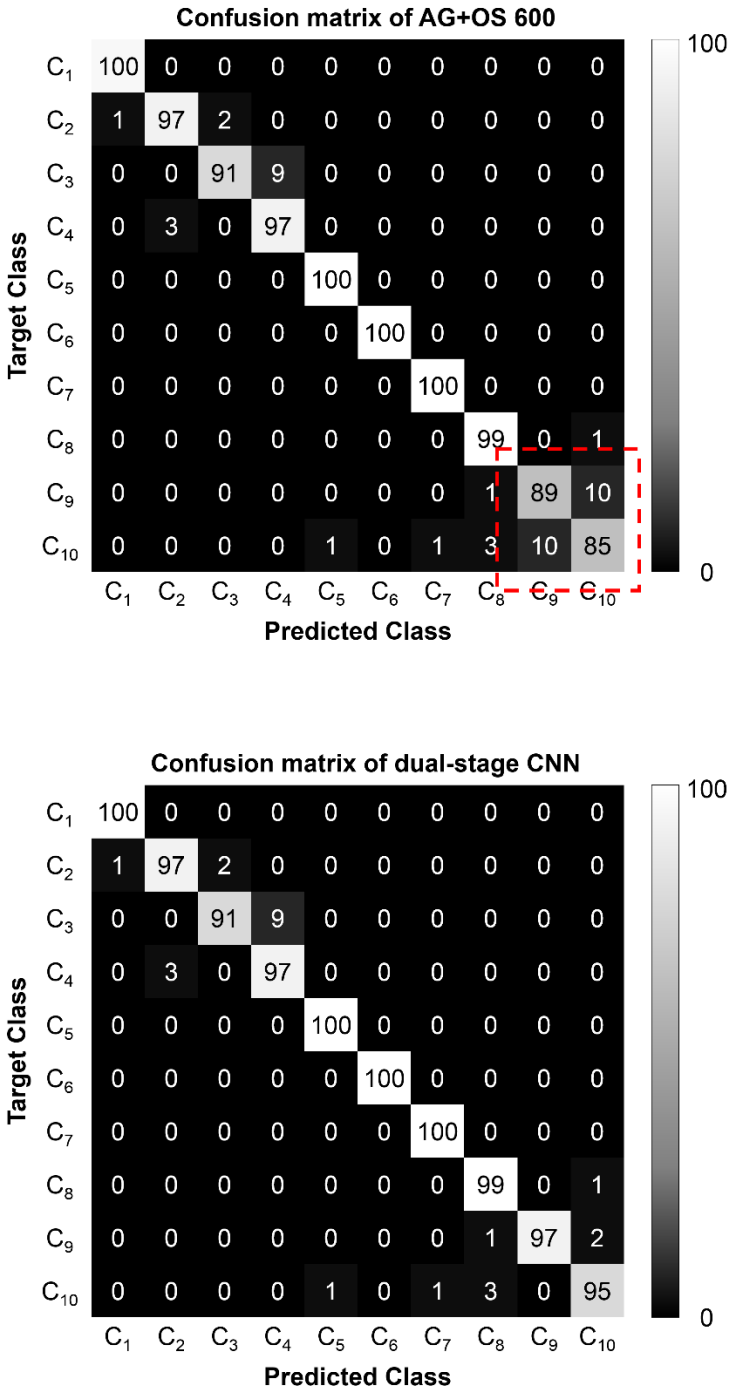

Figure 3.8 Comparison of confusion matrices of AG+OS 600 and dual-stage CNN.

The proposed work focused on WBC classification and yielded impressive performance. However, there is some room for improvement from the perspective of the entire system. The automated WBC differential counting system should include a detection process for high-content screening. Such screening has been used computer-aided diagnosis with a high-resolution microscopic scanner [35, 61]. This can help reduce time, cost, and labor, and can enhance the accuracy of diagnosis by assessing a larger number of cells in a bone marrow smear slide. Moreover, the proposed method can discriminate 10 WBCs of the maturation stages, but should include a complete list of WBC types, such as plasma cells, lymphocytes, and others. Moreover, disease models, such as stages of leukemia, should be included. A complete list of cell types and disease modeling are important, but a dataset that can be used for this purpose does not yet exist. Therefore, a dataset of complete WBC types should first be created. Training of CNN highly depends not only on the size of data, but also on the reliability of ground truth labels. However, the process of collecting data is very tedious task and is error-prone. Moreover, it is difficult to separate continuous maturation stages with discrete

discriminative standards as mentioned previously. So, the dataset should be carefully prepared with these considerations and a protocol that can collect reliable data without any controversies should be developed.

3.4. Conclusion

Chapter 3 demonstrated the promise of the proposed dual-stage CNN as an automated leukocyte classifier. The dual-stage CNN was able to classify single leukocyte images into ten classes of myeloid and erythroid maturation series, and achieved an accuracy of 97.06%, a precision of 97.13%, a recall of 97.06%, and an F-1 score of 97.1%. The proposed method not only showed high classification performance, but also successfully classified raw images without single cell segmentation and manual feature extraction. Moreover, it demonstrated rotation and location invariance. The proposed classification algorithm was integrated with the detection algorithm for development of the automated leukocyte differential count system, which is covered in Chapter 4.

CHAPTER 4

IMPLEMENTATION OF AUTOMATED LEUKOCYTE DIFFERENTIAL COUNT SYSTEM

4.1. System Overview

The automated leukocyte differential count system was designed as a single platform with a graphical user interface, GUI, to reduce processing time and memory usage. Figure 4.1 shows the proposed steps of the automated differential count system scheme. The system was designed to minimize user involvement, so the process was fully automated.

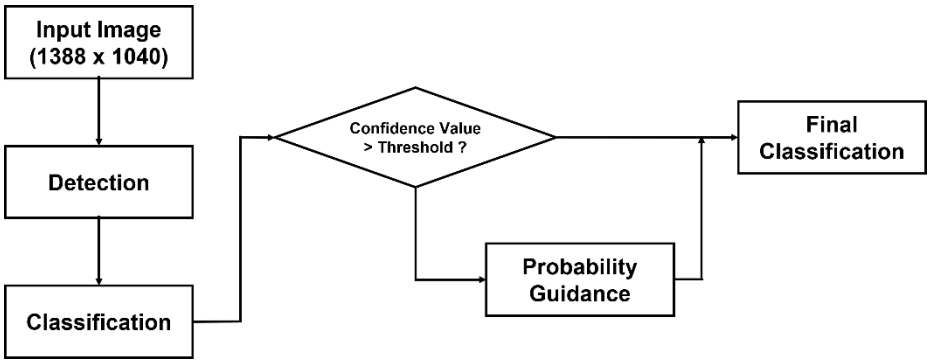

Figure 4.1 Block diagram of the automated leukocyte differential count system

The system was composed of three main processing blocks: leukocyte detection, leukocyte classification, and probability guidance. Briefly, a user loads a 1388 × 1040 digitalized bone marrow aspirate smear image of interest. The user clicks a button to operate the differential count. The system performs

the detection algorithm, which detects multiple leukocytes from the input, bone marrow aspirate smear image, and displays each corresponding points on the input image. At the end of the detection algorithm, single leukocyte sub-images are prepared as outputs, which become inputs for the classification algorithm. Then the classification algorithm classify these sub-image into ten maturation stages using the convolutional neural network (CNN). After the classification, the sub-images with low confidence values (classification probability) are re-analyzed using the probability guidance algorithm. It obtains multiple views of the sub-image by slightly adjusting the window location. The confidence values of these multiple view images are computed using the classification algorithm. The probability guidance then computes the final classification decision based on these confidence values. After the analysis, the differential count results are displayed on the GUI for the user.

In addition to the classification algorithm of Chapter 3, the detection algorithm was proposed and is described in this chapter. These two algorithms were implemented, and details of the probability guidance algorithm and the graphical user interface (GUI) are described in following sections.

4.2. Leukocyte Detection

This chapter is dedicated to multiple leukocyte detection in bone marrow aspirate smear images. It is a part of a leukocyte differential counting system that detects and yields sub-images. Before moving into the main subject, a terminology has to be clarified. Detection and localization are generally used inter changeably, but these two terms should be differentiated in image processing and deep learning. Localization is used for a process that locate a single object with in an image, whereas detection is used for a process that locate multiple objects with in an image. Main purpose of Chapter 2 is detection of leukocytes in a bone marrow aspirate smear image.

4.2.1. Introduction

Conventional machine learning follows a process of detection, segmentation, feature extraction, and classification. Usually, the performance of classification highly depends on the extracted features from segmented images. Therefore, obtaining an accurately segmented image of region of interest (ROI) is substantially crucial, because unnecessary information can be

extracted from over-segmented images and important information can be lost from under-segmented images.

There has been numerous attempts to automatically differential count leukocytes using conventional machine learning to substitute the tedious manual task to locate, identify, and count various types of leukocytes. The automated leukocyte analysis is typically preformed in four successive steps. A microscopic bone marrow image is first scanned to detect single leukocytes. Each detected leukocyte is further segmented into the nucleus and cytoplasm. After segmentation is completed, a set of features related to photometric, geometric, texture and morphological characteristic is extracted. Finally, a class label is assigned to each cell based on its feature vector [62].

Among these processes, the detection and the segmentation of a single leukocyte is the most important step since the errors of detection and segmentation progress to larger errors in feature extraction and classification. Therefore, many researches extensively focused on the detection and segmentation process to reduce errors. However, several difficulties, especially of bone marrow aspirate smear, hinder the progress of detection

and segmentation performance. One of the challenges is that heterogeneity of sample due to variation in quality of staining and acquisition. So, a boundary between nucleus and cytoplasm of a leukocyte, and a boundary between two closely adjacent leukocytes are unclear. Moreover, leukocytes exhibit wide variations of morphological characteristics. This problem is more severe in bone marrow aspirate smear compare to peripheral blood smear, since bone marrow aspirate smear contains more types of leukocytes in different maturation stages which increases divergence of morphology and size. Bone marrow aspirate smear also consists of higher cell density than peripheral blood smear, so overlapped leukocytes regions are more frequently observed. In result, a relatively few studies on bone marrow aspiration smear has been conducted and they were less successful.

Several past studies implemented several methods for segmentation, such as Hough space analysis, Gabor filters, adaptive thresholding, intensity clustering, edge detection, regions growing, support vector machine (SVM), artificial neural network (ANN), simulated visual attention [34, 63–68]. Yet, the segmentation problem has not been perfectly resolved,

especially in case of touching cells [37]. Some noteworthy relevant previous studies are briefly discussed. Liao. Q *et al.* used a simple threshold algorithm, which did not show sufficient performance [69]. Osowski *et al.* implemented watershed transformation and showed improvement, but it had over-segmentation problem [42]. Some utilized image histogram along with scale-space filtering and watershed transformation [70] and other improved performance by contrast stretching and image arithmetic [71]. Some works tried to resolve the problem of overlapped leukocytes by joining concave points using separating lines, or by eroding and growing regions retaining the shape [65, 72-74]. Several works used active contour model for leukocyte segmentation [75, 76]. Active contour was effective, especially for segregating touching leukocytes. However, a performance of active contour highly depends on initial seeding positions, which is difficult to achieve sufficient automation, so active contour was not appropriate for development of automated leukocyte detection.

Recently, Reta et al. proposed a segmentation algorithm that incorporates contextual color and texture information using Markov Random Field for semantic segmentation of leukocytes

in bone marrow images. They also tried to overcome leukocyte overlapping problem by using edge transformation from Cartesian to polar space, and linear interpolation on discontinuous points to draw a conical curve. This work achieved a segmentation accuracy of 95%, which was promising.

Unfortunately, many researches could not achieve a perfect segmentation algorithm until now. Therefore, this thesis tilted an approach by focusing only on detection of leukocyte, not on achieving perfect segmentation of a single-leukocyte. Main purpose of improving segmentation performance is to increase performance of classification. This thesis uses CNN for classification as described in Chapter 3. Strength of CNN is its characteristic of representation learning, which operates in raw or minimally preprocessed images. In result, the segmentation of conventional machine learning algorithm is replaced with detection. In this section, the first part of the automated leukocyte differential count system is described. Briefly, basic concept is similar to segmentation; however, since segmentation is error-prone so we only detect and roughly segment the region of a single leukocyte with background into a small image patch, which will be feed into CNN for classification.

4.2.2. Detection Algorithm

Detection process is to locate sub-images of a single leukocytes. Conventional machine learning required the process of nucleus and cytoplasm separation from background that includes red blood cells, platelets, and empty space, and separation of touching cells. These pre-processes were required to extract features from a single leukocytes which is fed into a classifier such as support vector machine. However, these background removal and a single-cell segmentation are neglected due to the characteristic of classifier used in this thesis, CNN. The sub-images prepared by the detection process in the chapter are used for training and testing of CNN in Chapter 3.

The detection algorithm consists of four main stages: pre-processing, saliency mapping, watershed transformation, and sub-image extraction, as shown in Figure 4.2.

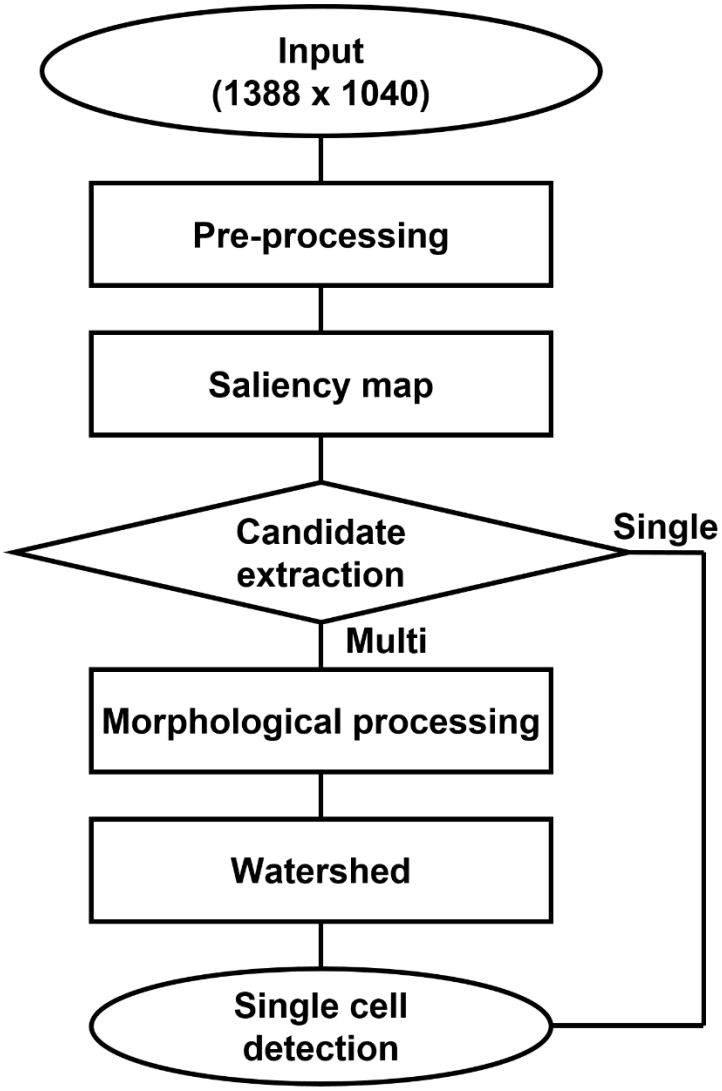

Figure 4.2 Block diagram of the detection algorithm.

First, all the input images with size of 1388 × 1040 were converted from RGB color space into HSV color space. HSV color space selected, because it provides better and similar representation of how humans perceive color than RGB color space. Images in HSV color spaces processed through a

saliency map. A saliency map is an image that is simplified representation with pixel's unique quality of an original image. It extracts region that stands out relative to its neighbors for easier analysis, which mimics the human visual attention mechanism [77−79]. Among many approaches, a spectral residual saliency map was implemented. It analyzes a log−spectrum of an input image and extracts the spectral residual of the image, which is reconstructed in spatial domain for representation of attention. The spectral residual saliency map is computed as follows in details [80]:

$$\text{Averaged amplitude spectrum}: \ \mathrm{A}(f) = \mathrm{Re}(\mathcal{F}[\mathrm{I}(x)])$$

$$\text{Phase spectrum}: \ \mathrm{P}(f) = \mathrm{Im}(\mathcal{F}[\mathrm{I}(x)])$$

$$\text{Log spectrum}: \ \mathrm{L}(f) = \log(\mathrm{A}(f))$$

$$\text{Spectral residual}: \mathrm{R}(f) = \mathrm{L}(f) - \mathrm{A}(f)$$

$$\text{Saliency map}: \mathrm{g}(x) \ * \mathcal{F}^{-1}[\exp(\mathrm{R}(f) + \mathrm{P}(f))]^2 \qquad (2.1)$$

where I(x) denotes an image, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier Transform and Inverse Fourier Transform.

Amplitude and phase of Fourier Transformed image is computed first. Based on the amplitude of Fourier Transform, the log spectrum is obtained and a spectral residual is then extracted from the log spectrum of image. At the end, the saliency map is reconstructed in spatial domain through inverse Fourier Transform. An example of saliency map in a natural image is shown in Figure 4.3.

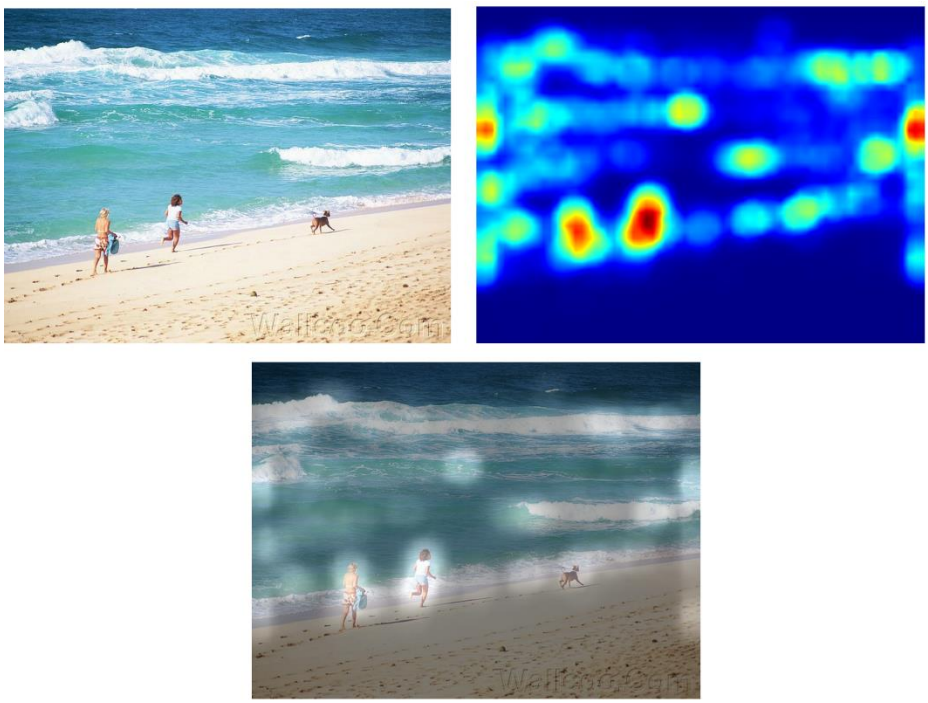

Figure 4.3 Example of saliency map applied on a natural image. (A) Original image. (B) Saliency map of the image. (C) Merged image of the original image and the saliency map.

Through saliency mapping, the leukocytes were separated from red blood cells and backgrounds of bone marrow aspirate smear image. These leukocytes were clustered and assessed for a candidate extraction. During the candidate extraction, sizes of clusters were assessed and the clusters were divided into two categories: single object and multiple object. Single objects were considered as single leukocytes, and proceeded to a sub-image extraction stage. Multiple object clusters were further processed with morphological processing and the watershed transformation. A combination of morphological image process, erosion and dilation, were conducted on the multi-object cluster to prepare for the watershed transformation.

The watershed transformation is a region-based segmentation algorithm that is widely used for separating adjacent and touching objects, which is a main issue in leukocyte detection from bone marrow aspirate smears. The watershed transform considers an image as topographic surface which is flooded by water. A pixel intensity of image is a height of landscape, and by changing a water level above water and below water are discriminated, which become segmented [81].

After watershed transformation on multiple object cluster, leukocytes in the cluster were separated into individual object and proceeded to sub-image extraction stage. The first step in sub-image extraction stage was to find the center point of each leukocyte. These center points were used to crop sub-images that contain a single leukocyte. The width and length of sub-images were fixed to 96 x 96, which is an image size that is later used for CNN. The same cropping window size was used for all leukocytes to prevent resizing of images. Size of cell is one of key morphological characteristic that is used for leukocyte classification by both hematologists and computer programs. CNN uses the same input size, so if the image size is different, images needs to be either cropped or resized. However, as mentioned, resizing a leukocyte image may contaminate features of cell characteristics. Therefore, the size of single leukocyte sub-images were fixed to 96 x 96 throughout the entire study.

4.2.3. Experimental Setup and Evaluation

The detection algorithm was implemented using MATLAB and was assessed on the same personal computer system on 64 bits Windows 7 OS with i5-3570 @ 3.40GHz CPU and 8 GB RAM. The performance of detection algorithm was assessed and evaluated on randomly selected 20 bone marrow aspirate smear images with given number of leukocytes. The output of detection algorithm was divided into four categories: correct, incorrect-missing detection, incorrect-multiple detection, and incorrect-false detection. Incorrect-missing detection indicates a case that the detection algorithm did not detect a leukocyte, incorrect-multiple detection indicates a case that the detection algorithm detected a leukocyte multiple times, and incorrect-false detection indicates a case that detection algorithm detected an object that is not a leukocyte such as background, red blood cell, and stain artifacts.

4.2.4. Results and Discussion

The detection process was demonstrated step by step, as shown in Figure 2.7 to Figure 2.9. An input image in RGB color space was converted to HSV color space. H channel showed enhanced the cytoplasm of leukocytes, and S channel enhanced the features of leukocytes' nucleus. V channel enhanced the background intensity. These channels were combined and the contrast was enhanced before converting to a binary image. This pre-process steps are demonstrated in Figure 4.4.

The binary image then proceeded with saliency mapping. The saliency map provided a simplified representation of outlines and locations of leukocytes. Figure 4.5A shows the output of saliency mapping in jet color map. Blur outlines of leukocytes were detected as unique pixels of the image. As shown in Figure 4.5B, the locations of saliency map matches with the locations of leukocytes in the original images. The outputs were clustered for candidate extraction, where single candidates and multiple candidates were divided.

Single candidates on the saliency map were processed to save sub-images of them. However, watershed the transformation was conducted to separate multiple candidates of the saliency map, as shown in Figure 4.6.

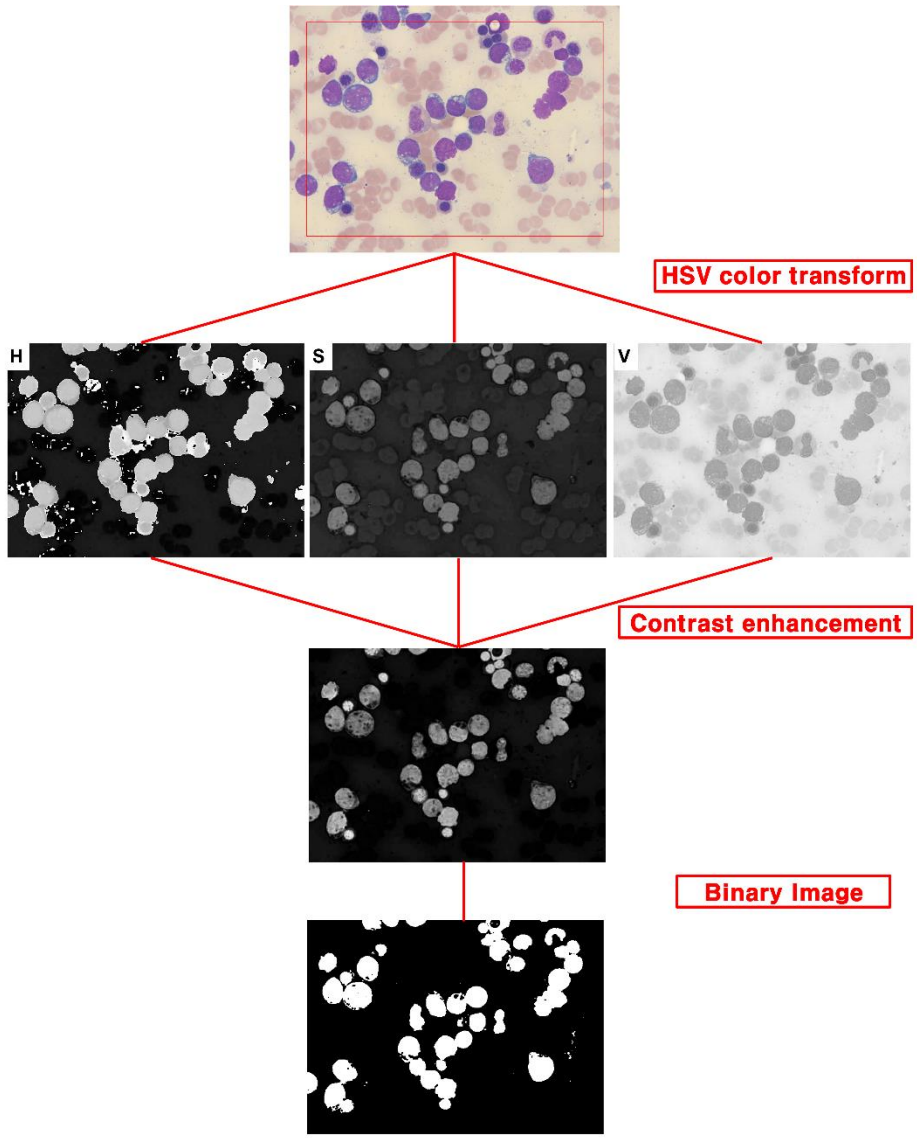

Figure 4.4 Example of stepwise outputs of pre-processing.

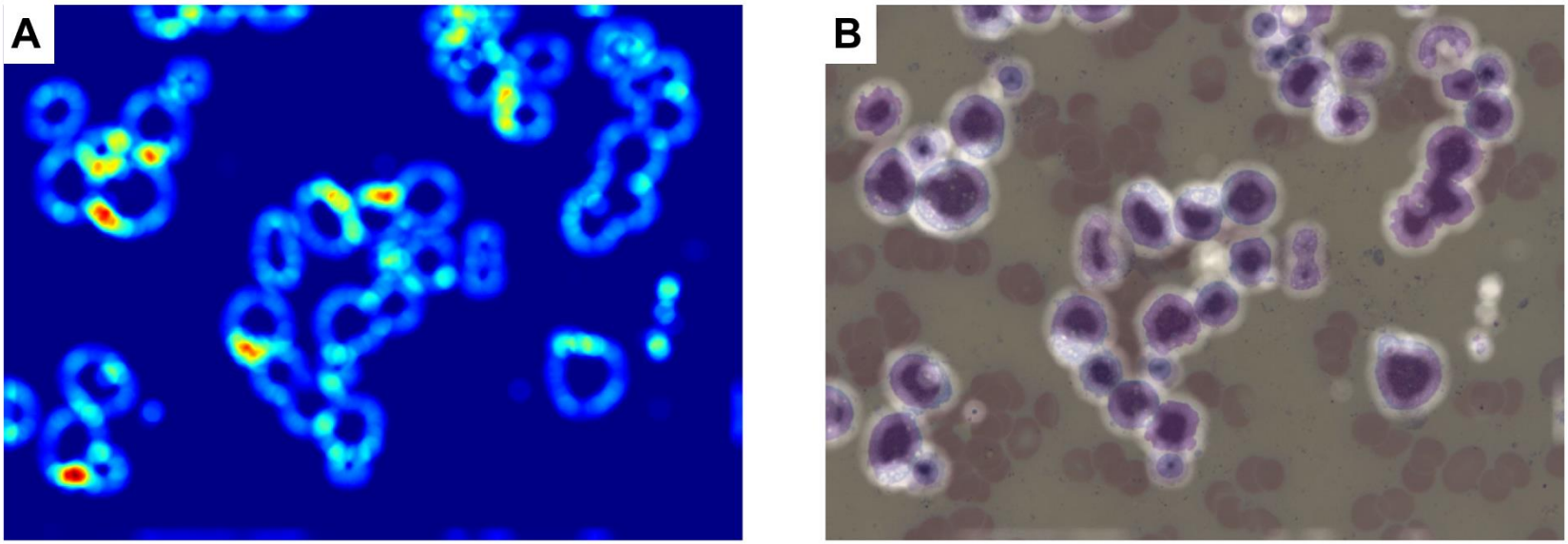

Figure 4.5 (A) Example of the output saliency map. (B) Merged image of the saliency map and the corresponding input image.

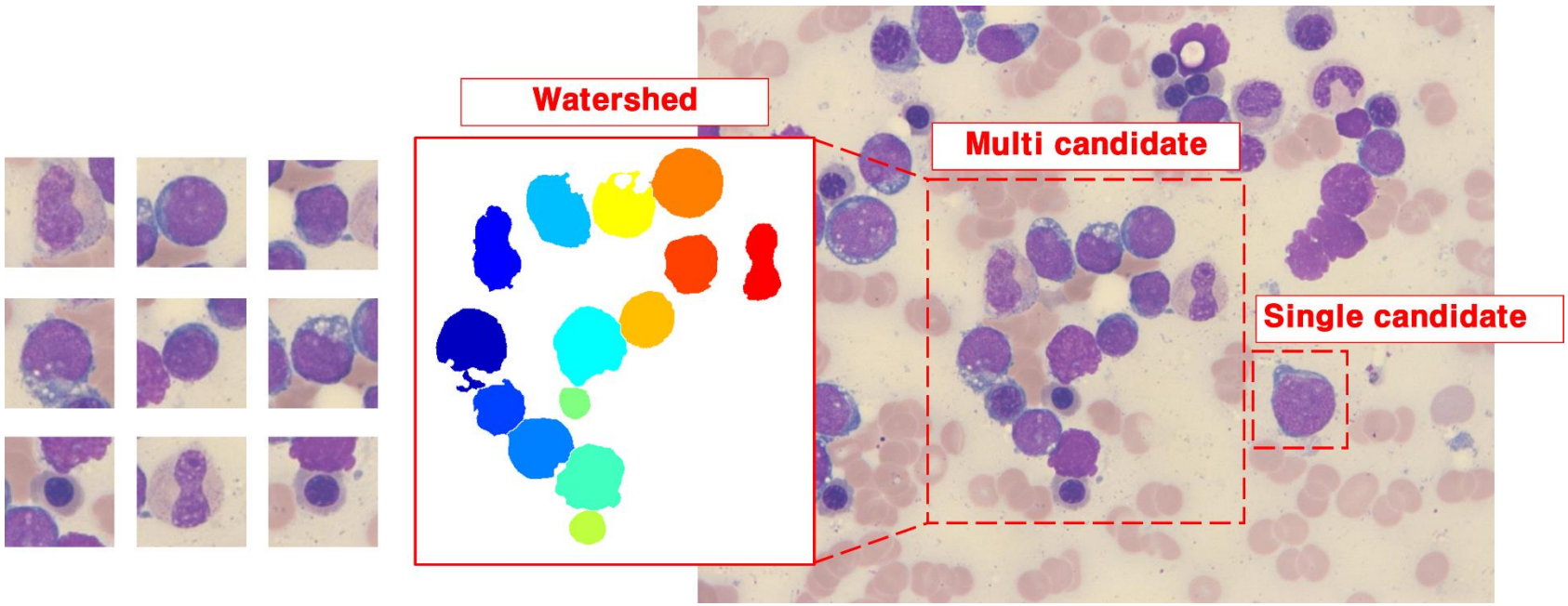

Figure 4.6 Example of processing single leukocyte sub-image. For a multi-candidate cluster, watershed transformation was utilized to separate each leukocyte into a single leukocyte sub-images. A single candidate was considered as a single leukocyte sub-image.

The proposed detection algorithm showed good qualitative detection results allowing sub-image extraction of single leukocytes. It was able to differentiate touching leukocytes in high density bone marrow aspirate smear images. Detailed analysis was conducted on randomly selected 20 samples of the original image set. The subsets consists of images with a size of 1388 × 1040 pixels that have different cell densities and color variations from their staining and acquisition process.

An example of large bone marrow aspirate smears and corresponding extracted sub-images are demonstrated in Figure 4.7. The detection algorithm excluded leukocytes on boundaries of the image, because a full 96 × 96 image cannot be obtained due to occlusion problem. Moreover, the detection algorithm was able to perform detection in variation conditions of color, contrast, and leukocyte density. It also detected leukocytes that were touching each other. However, it often missed some leukocytes when the clusters of multiple leukocytes were containing too large.

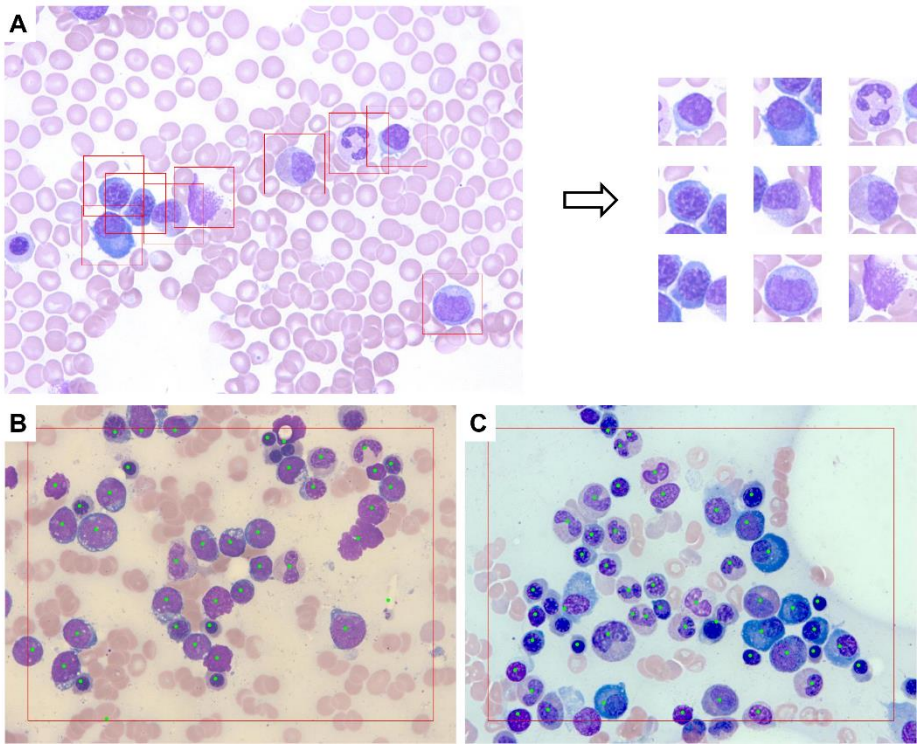

Figure 4.7 Output of detection algorithm: (A) Bounding boxes on input image (left), and corresponding sub-images (right). (B) and (C) Example showing performance of the detection algorithm on images with color, contrast, and leukocyte density variation.

The result of assessment and evaluation of the detection algorithm is summarized in Table 2.1. It provides a numerical count of each evaluation category, an accuracy for each sample image, and elapsed process time. The detection algorithm achieved the average accuracy of 96.09% and the average

processing time of 2.98 seconds per a sample. This means that it can detect 100 leukocytes, which is the current counting number used in the clinical hematology laboratories, in 18.98 seconds and can process 300 leukocytes within a minute. The system can not only increase efficiency, but also provide statistically more reliable information from random sampling of leukocytes in a smear [15].

Throughout the samples, the incorrect detection-false detection was the most frequently observed errors. Usually, red blood cells and stain artifacts were detected in this category. This problem can be easily solved from the classification algorithm by classifying these red blood cells and artifacts as a background class and excluding them from the differential count.

The incorrect detection-missing detection also occurred frequently. However, the overall count of the missing detection would be negligible when over 300 to 500 leukocytes are examined as compare to the current counting number of 100.

The most severe problem of the detection algorithm was the incorrect detection-multiple detection of the same leukocyte.

This error was usually observed on band neutrophils and segmented neutrophils due to their banded nucleus shapes. This may interfere obtaining a correct differential count. A possible solution to this problem is to make fine adjustments to the watershed transformation. However, a user cannot adjust the parameter every usage, so alternatively better solution to this issue may be excluding one of the double counts by comparing two similar adjacent images. However, the incorrect detection–multiple detection cases were rarely observed, the problem was not considered seriously and may be neglected in the future when a large sample of detection is performed.

Table 4.1 Evaluation results of the detection algorithm

Image ID	True number of leukocytes	Correct detection	Accuracy (%)	Incorrect detection			Time elapsed (sec)
				Missing detection	Multiple detection	False detection	
Sample 252	9	9	100	0	0	1	2.94
Sample 251	8	8	100	0	0	1	2.71
Sample 250	8	8	100	0	0	1	2.67
Sample 249	14	12	85.71	2	0	1	3.15
Sample 248	5	5	100	0	0	1	2.55
Sample 247	5	5	100	0	0	3	2.79
Sample 244	9	9	100	0	0	0	2.64
Sample 230	9	9	100	0	0	2	2.83

Sample 220	9	9	100	0	0	0	2.95
Sample 211	7	7	100	0	0	2	2.63
Sample 210	10	10	100	0	0	1	2.73
Sample 006	51	46	90.20	3	2	0	3.80
Sample 009	37	30	81.08	7	0	1	3.49
Sample 030	18	18	100	0	0	0	2.87
Sample 142	12	12	100	0	0	0	2.71
Sample 147	24	23	95.83	1	0	0	3.33
Sample 143	18	15	83.33	3	0	1	3.18
Sample 099	21	19	90.48	1	1	0	2.98
Sample 109	21	21	100	0	0	0	2.95
Sample 113	41	39	95.12	2	0	2	3.63

4.3. Automated Leukocyte Differential Count System

4.3.1. Implementation of Detection and Classification Algorithm

As stated in Chapter 2 and Chapter 3, the detection algorithm was developed with MATLAB on Window 7 OS, and the classification algorithm was developed with Torch7 deep learning framework on Linux OS. Therefore, the detection and classification algorithms needed to be translated into a uniform language for integration. Usually one language is translated into one another, but both algorithms needed to be translated for this study. Python was chosen as the base language, because not only Deep Learning framework is available, but also it is the most compatible language for GUI development and image processing on Linux OS. Therefore, MATLAB and Torch7, which is based on Lua, were translated into Python, and PyTorch was used for the deep learning framework. PyTorch provides two high-level features: tensor computation with GPU and Deep Learning network built on an autograd system. Strength of PyTorch is that it can be used with other python

packages, such as numpy and scipy that are widely used for image processing. Some of useful MATLAB functions used in the detection algorithms were re-implemented in Python.

4.3.2. Graphical User Interface Design

Graphical user interface of the automated leukocyte differential count system was designed for the convenience of users, as shown in Figure 4.8.

The design of GUI mostly focused on simplification of operation step. A user only needed to load a digitalized bone marrow aspirate image of interest and click two buttons for the system operation. The mechanism embedded in the GUI followed the process described in Figure 4.1. The GUI was implemented with PyQT4.

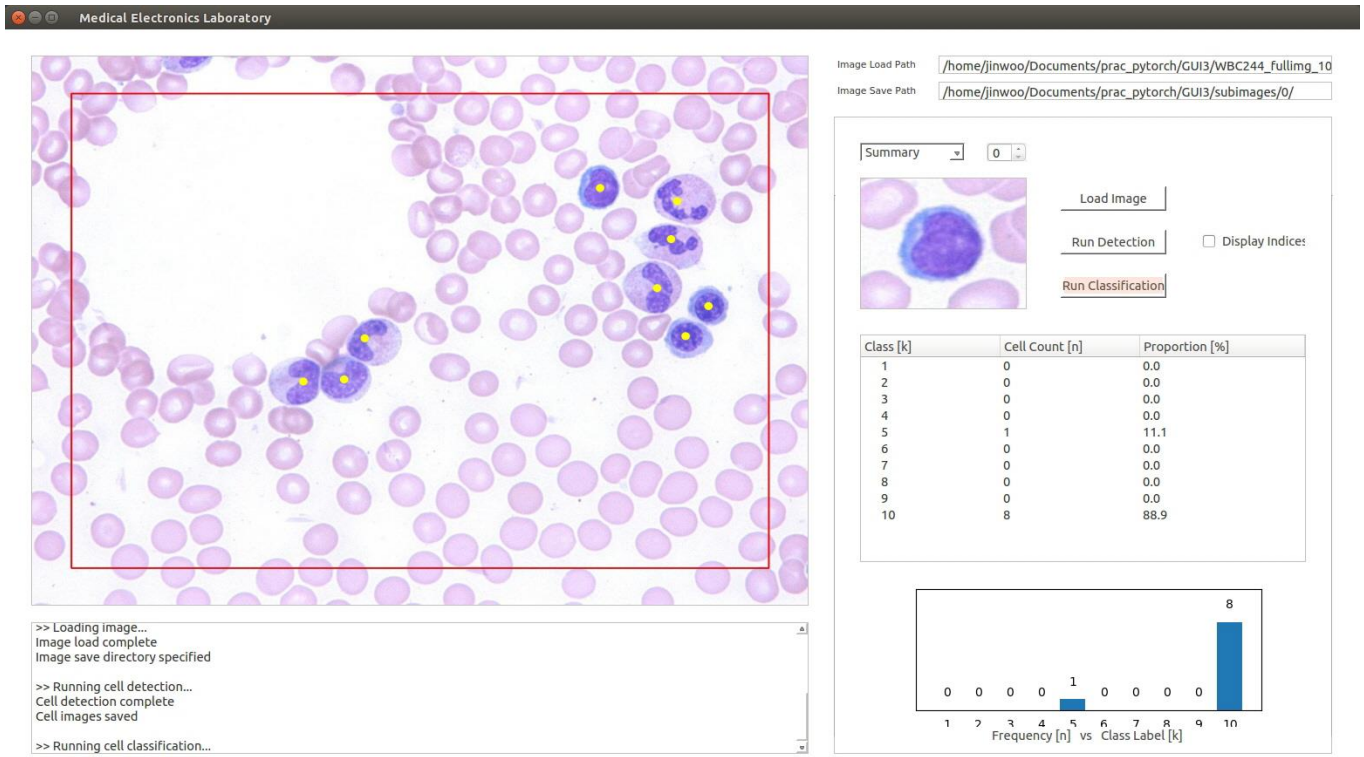

Figure 4.8 Screenshot of the automated leukocyte differential count system GUI.

4.3.3. Probability Guidance Algorithm

In developing and integrating the proposed system, a problem has been brought up related to the performance. Some of detected leukocyte sub-images were off the center or were not applicable for the classification algorithm, even though the classification algorithm showed strength of location invariance. Therefore, a probability guidance algorithm was proposed to fine-tuned errors that occurred during transition from detection to classification, as shown in Figure 4.9.

After the classification of a single leukocyte was processed, the probability guidance algorithm was executed if the confidence value of the classification was below a threshold value of 80%. The probability guidance algorithm extracted multiple sub-images of the leukocyte in slightly different coordinates and computed classification confidence values of the multiple sub-images. Final classification decision was made based on these confidence values.

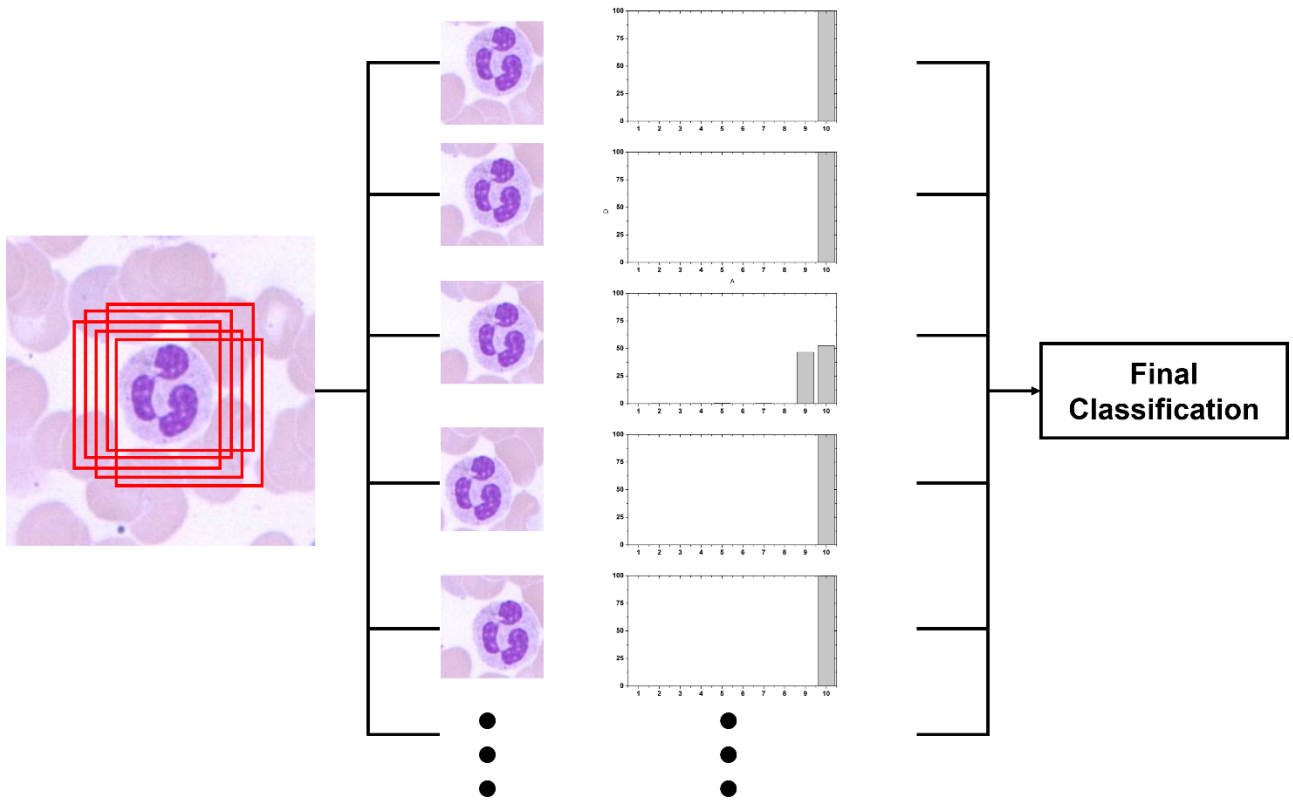

Figure 4.9 Description of probability guidance algorithm.

4.3.4. Experimental Setup and Evaluation

The automated leukocyte differential count system was implemented with Python based PyTorch Deep Learning framework and was assessed on the same personal computer system on Linux OS with Intel CPU i7 6700 @ 3.40GHz, RAM 16GB, and GPU NVIDIA GTX 980 4GB. The system was assessed on 50 bone marrow aspirate smear images with given number of leukocytes and corresponding labels. The performance of the propose system was evaluated with the same metrics for detection and classification.

4.3.5 Results and Discussion

The network was re-trained on PyTorch and was successfully trained. Figure 4.10 demonstrates details of training networks. The accuracy of train set and validation set increased over 90% and converged after 50 epochs. A confusion matrix of validation set was generated to evaluate class-wise performance. A progression of the confusion matrix was generated in GIF file over 150 epochs of training, and the file was attached as Appendix A.1.

Appendix A.1 A progression of the confusion matrix over 150 epochs during training.

A demonstration of the integrated system was recorded in a video and was attached as Appendix A.2.

Appendix A.2 A video demonstration of the integrated system.

During the overall assessment the integrated system showed the similar performance as those from detection and classification algorithms. Among 50 bone marrow aspirate smear images, total 300 leukocytes of 10 target maturation stages were present.

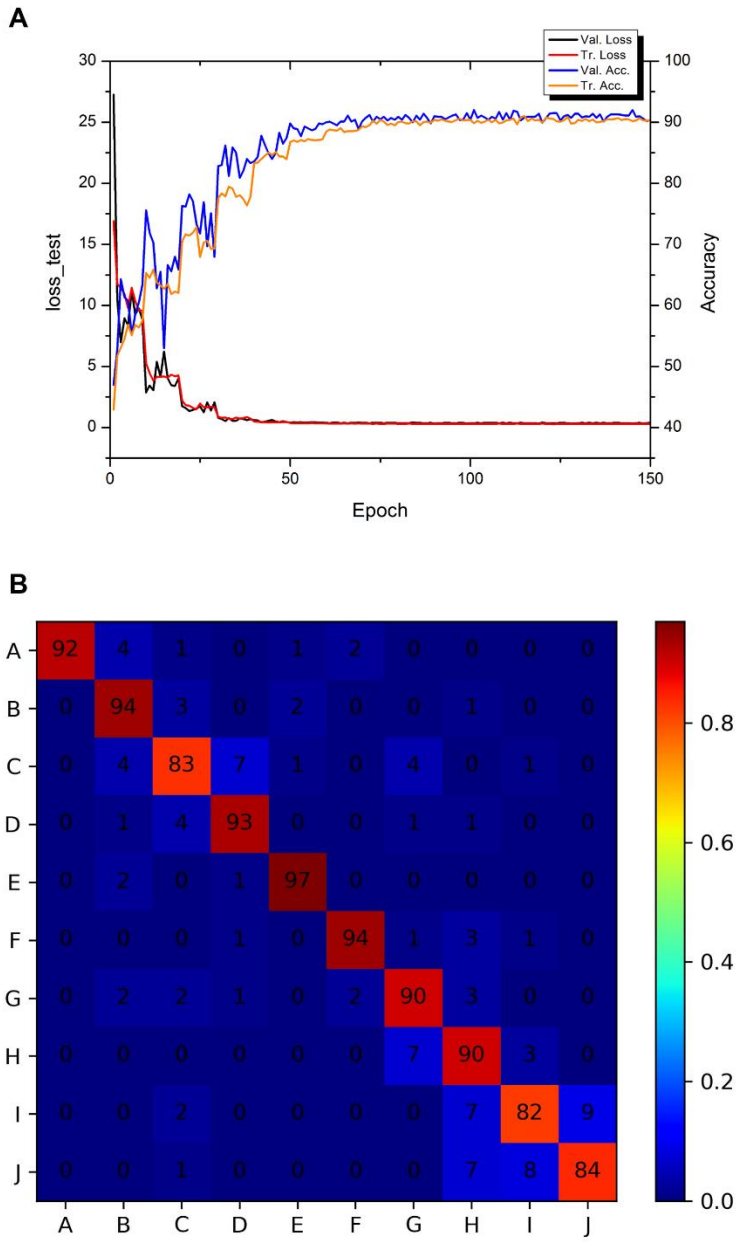

Figure 4.10 Details of training networks in PyTorch. (A) Graph of accuracy and loss of training and validation accuracy and loss during training of the network. (B) Confusion matrix of the network.

The system correctly detected 290 out of 300 leukocytes (96.7%), and correctly classified 274 out of 290 detected leukocytes (94.5%). Therefore, the overall performance of the system was 91.3% accuracy, which was slightly lower than the expected overall performance of 93.26%.

The system was able to detect leukocytes within a range of 160 milliseconds to 630 milliseconds and to classify leukocytes within a range of 250 milliseconds to 740 milliseconds respect to a number of analyzed leukocytes in a single bone marrow aspirate smear image. Usually, an expert hematologist can examine a smear, where 100 leukocytes are differential counted, within six minutes. However, the system required 40 to 50 milliseconds to detect and classify a single leukocyte on average, which indicates that the system would be able to process 100 leukocyte analysis within four to five seconds. This is 72 to 90 times faster performance than a human. As stated before, another goal of this thesis is to increase the number of leukocyte analyzed for each sample to increase statistical reliability of diagnosis result through the automated system. With the current efficient of the system, 300 leukocyte analysis per a smear can be done in 12 to 15 seconds and 500

leukocyte analysis per a smear can be done in 20 to 25 seconds, which are still over 18 times more efficient than a human. However, in order to achieve fully automated differential count, a hardware that can automatically acquire smear images should be integrated to the proposed system. With advances of imaging system such as slide scanner and automated microscope, this could be easily accomplished. Moreover, the system required about 3.56 seconds of initial loading time of the trained network. This may be negligible once the system is loaded, but it should be considered in the future for convenience.

During the assessments of the system, some misclassifications were observed, because the smear image included maturation stages that were not trained during the classification algorithm. Therefore, the evaluation of classification performance was insufficient. As mentioned in Chapter 3, obtaining a complete dataset of all maturation stages is essential future work that should be done first in order to put the system in clinical hematology laboratories.

4.4. Conclusion

An automated leukocyte differential count system was developed. The system consisted of detection process and classification process. Therefore, additional to the classification algorithm of Chapter 3, the detection algorithm was developed.

The proposed leukocyte detection algorithm demonstrated the promising result. The saliency map and the watershed transformation was utilized to detect and obtain single leukocyte sub-images from 1388 × 1040 high density bone marrow aspirate smear images with numerous touching leukocytes. The proposed detection algorithm achieved an average accuracy of 96.09%. The average processing time was 2.98 seconds per a sample, which means that the detection algorithm is capable of processing 100 leukocytes in 18.98 seconds.

Since, two algorithms were developed in different languages, these were re-implemented in Python. This also reduced the performance time of detection algorithm. Additionally, the probability guidance algorithm was proposed to link effectively

link two algorithms and to increase the overall performance. Moreover, the system set up the fully automated and user-friendly interface with GUI.

The automated leukocyte differential count system successfully detected multiple leukocytes and classified them into ten different maturation stages. It also fine-tuned errors that occurred from integrating the detection and classification algorithms with the probability guidance algorithm. The system achieved 4-5 seconds per 100 leukocytes and the accuracy of 91.3% (detection accuracy: 96.7% and classification accuracy: 94.5%) in differential count performance. The system proposed a new paradigm of differential count and hematology laboratories.

CHAPTER 5

THESIS SUMMARY AND FUTURE WORK

5.1. Thesis Summary and Contributions

An automated leukocyte differential count system using the dual-stage convolutional neural network (CNN) was developed. The system consists of leukocyte detection algorithm and classification algorithm, and user-friendly graphical user interface (GUI) was designed for integration. The system was able to detect and classify more than ten classes of leukocytes, and showed strength of location and rotation variance. This thesis successfully overcame challenges of bone marrow aspirate smears.

First, the classification algorithm was developed to classify sub-images from the detection algorithm into ten maturation stages of leukocytes, including myeloid series and erythroid series. This thesis overcame several challenges of leukocyte classification in bone marrow aspirate smear. The maturation of leukocytes are continuous, so it is hard to define discrete standard. Also, the density of leukocytes in bone marrow is high, so many leukocytes touch other leukocytes, RBCs, or backgrounds, which makes segmentation and feature extraction of a single leukocyte difficult. Due to this problem, the

conventional machine learning was incapable of processing raw or minimally pre-process images However, the dual-stage CNN successfully learned representative features of leukocytes from images with background and touching leukocytes. It also successfully classified leukocytes of ten maturation stages with accuracy of 97.06%. Moreover, it showed strength of rotation and location invariance.

Next, the detection algorithm was developed to detect multiple leukocytes from a large bone marrow aspirate smear image and to crop them into single leukocyte sub-images for the classification. For the purpose, 200 digitalized bone marrow aspirate smear images were collected with help of hematologists at Seoul National University Hospital. Unlike previous studies that used conventional machine learning, the proposed system used CNN that can classify raw or minimally pre-processed images. Therefore, the segmentation process of the conventional machine learning was neglected, and only detection was considered. The detection algorithm utilizes saliency mapping and watershed transformation. Due to simplified process without error prone segmentation step, the

detection algorithm showed outstanding performance of 96.09% overall accuracy.

Lastly, these detection and classification algorithms were integrated into a single system with GUI. The detection process were re-implemented in Python with OpenCV library, and the classification process was re-implemented from Torch7 to PyTorch deep learning framework. Moreover, the probability guidance algorithm was proposed. Although, the detection algorithm successfully detected a single leukocyte and the classification algorithm successfully classified location invariantly, there were some cases that the location of leukocyte was not suitable for classification. For these cases, the probability guidance algorithm introduced multiple views of a single leukocyte and computed confidence value of each view. The average of the confidence values was used to make a final classification decision. The system was able to process leukocyte detection and classification with 91.3% of accuracy within 40 to 50 milliseconds for a single leukocyte, which was up to 90 times faster performance than a human. The system operation time discarded the time required for acquisition of bone marrow aspirate smears, so it was not a fair comparison

between the system and a human. However, the acquisition time may be negligible when it is minimized by changing an acquisition environment. With advances in acquisition systems, many systems, such as a slide scanner, can acquire wide range of samples with high resolution. By using these systems, a bone marrow aspirate smear needs to be acquired once for analysis, and many samples can be acquired automatically to increase throughput. Further discussion regarding the issue of high throughput is covered in the following section.

5.2. Future Work

The automated leukocyte differential count system successfully performed automated leukocyte differential counting and has proven the potential of CNN in classification of multiple leukocyte maturation stages. This thesis proposes a new paradigm in automated blood disease diagnosis and shows a potential of deep learning, especially CNN, in biomedical image processing and automated systems. However, the system can be improved by resolving a few issues. Here, some possible improvements, that should be considered in the future, are suggested.

First, the system should include the complete list of leukocyte maturation stages. This thesis demonstrated the differential count of ten leukocyte maturation stages, including six myeloid maturation series and four erythroid series, among many leukocyte types. These two series were first selected because these have more sub-stages of maturation, which increases a chance of mutation and abnormality. Actually, these maturations were frequently observed and myeloid series are directly related to acute myeloid leukemia and chronic myeloid leukemia.

However, for the completion of list, basophil, eosinophil, monocyte, plasma cell, immature lymphocyte, lymphoblast, and lymphocyte should be included, and the lymphoid series should be considered the first. Analysis of lymphoid series, that include lymphoblast, immature lymphocyte, and lymphocyte, is essential since it is directly related to acute lymphocytic leukemia and chronic lymphocytic leukemia. Furthermore, the system should be able to distinguish the abnormal morphology of each leukocyte. Distribution of leukocyte is the fundamental information for diagnosis, but cytology of leukocyte will provide more information for diagnosis. Developing a system with the complete list of maturation stages and cytology can be accomplished once the data is provided, which is the difficult part. The dataset for this thesis was collected for a year, but other maturation stages, that were mentioned above, were less frequently observed and collected. However, this problem will be eventually resolved and the system can be upgraded as soon as the data is collected. Once the system with complete list is developed, the system should be validated in clinics by directly comparing the performance with expert hematologists.

Another suggestion is an upgrade of acquisition environment for higher throughput. The proposed system was developed for 1388 × 1040 bone marrow aspirate smear images and 96 × 96 single leukocyte sub-images. With advances in acquisition technologies including virtual microscopy, scanning microscopes and cameras, acquisition of larger area in peripheral blood smear and bone marrow aspirate smear would be possible [82]. Using these cameras with higher resolution would increase a number of leukocyte entrapped inside a single image. This would enable the proposed system to detect and assess more leukocyte in one sight, and would reduce analysis time. Moreover, higher resolution images would depict more detailed features of cells, such as clear boundaries of cytoplasm and nucleus, and clear granularity of leukocytes. This would result in higher classification accuracy. However, there is an obstacle in indefinitely increasing resolution and size of images for the automated leukocyte differential count system. As described in Chapter 3, CNN learns representations from pixels of images. Therefore, increasing resolution and size results in increased required computing memory and process time. In this thesis, NVIDIA GTX 980 4GB was used, which is not enough

memory to process high resolution and size images and requires a long time to converge during training. Therefore, this obstacle can be resolved by implementing the system on a computer with higher computing power such as NVIDIA TITAN X 12GB and/or multiple GPUs in parallel. However, the image resolution and size should be still considered along with the computing system.

Another inconvenience is a long initial loading time of the system. It was built based on Python language, which is known to be light in computation. The proposed system does not require much computation for detection and classification process, 160 to 630 milliseconds and 250 to 740 milliseconds respectively, but it requires a long time for initial loading of trained CNN network, 3.56 seconds. This problem will become more severe when the two above recommendations are satisfied, because the system will require a larger CNN network that increases size of saved trained CNN network. The problem can be partially by upgrading a computing system, but this solution does not fundamentally resolve the issue. A possible solution to this problem is developing a new protocol for saving a trained network that could reduce a loading time. Moreover,

the size of CNN network should be carefully selected, modified, or redesigned, so that unnecessarily large features are not saved. This ancillary issue, however, does not affect the performance of the system, and can be neglected after the program is initially loaded, since the trained network is initialized only once at the start of the program. However, this should be considered for convenience of users.

Lastly, an algorithm that can robustly transform cell sizes depending on acquired images is recommended. A size of leukocyte is one of the most important features that is considered for the classification. Therefore, the sub-images in this thesis were prepared with a single acquisition system in uniform size of 96 x 96 to avoid resizing of images. However, different acquisition systems have different size and resolution of output images, so the size of leukocytes will be different in images that were acquired from different acquisition systems. In order to utilize the proposed automated differential count system on different images from other acquisition systems, the size of leukocytes must be calibrated respect to different magnification and resolution of images. For example, a 96 x 96 sub-image of leukocyte from a higher resolution image would

look smaller, so the image needs resizing. However, CNN only can compute a 96 x 96 image size and this causes a problem. Therefore, an algorithm that can robustly transform leukocyte image size respect to the input size of the CNN should be developed to utilize the system in many clinical hematology laboratories.

Bibliography

1. Janeway CA, Travers P, Walport M, Shlomchik MJ. Immunobiology: the immune system in health and disease. 1997;1.
2. Theil KS. Bone Marrow Processing and Normal Morphology. Laboratory Hematology Practice. 2012;22:279-99.
3. Briggs C, Machin SJ. Automated Platelet Analysis. Laboratory Hematology Practice. 2012;5:48-58.
4. d'Onofrio G, Zini G. Analysis of bone marrow aspiration fluid using automated blood cell counters. Clinics in laboratory medicine. 2015;35(1):25-42.
5. Stetler-Stevenson M, DiGiuseppe JA, Arthur DC. Myelodysplastic Disorders. Laboratory Hematology Practice. 2012;25:331.
6. DiGiuseppe JA, Mnayer L. Acute Leukemias. Laboratory Hematology Practice. 2012;26:345.
7. Allan RW, Al-Quran SZ, Li Y, Braylan RC. Lymphoproliferative Disorders. Laboratory Hematology Practice. 2012;27:364.
8. Richendollar BG, Cook JR. Plasma Cell Disorders. Laboratory Hematology Practice. 2012;28:381.
9. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA: a cancer journal for clinicians. 2016;66(1):7-30.
10. Park E-H, Lee H, Won Y-J, Ju HY, Oh C-M, Ingabire C, et al. Nationwide statistical analysis of myeloid malignancies in Korea: incidence and survival rate from 1999 to 2012. Blood research. 2015;50(4):204-17.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436.
12. Da Costa L. Digital image analysis of blood cells. Clinics in laboratory medicine. 2015;35(1):105-22.
13. Chabot-Richards DS, George TI. White blood cell counts: reference methodology. Clinics in laboratory medicine. 2015;35(1):11-24.
14. Davis BH, Barnes PW. Automated Cell Analysis: Principles. Laboratory Hematology Practice. 2012;3:26.
15. Béné M-C, Lacombe F. Differential Leukocyte Analysis. Laboratory Hematology Practice. 2012;4:33.
16. Peterson P, McNeill S, Gulati G. Cellular Morphologic

Analysis of Peripheral Blood. Laboratory Hematology Practice. 2012;2:10-25.
17. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. arXiv preprint arXiv:151207108. 2015.
18. LeCun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, et al., editors. Handwritten digit recognition with a back-propagation network. Neural Information Processing Systems (NIPS); 1989.
19. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-324.
20. Krizhevsky A, Sutskever I, Hinton GE, editors. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems; 2012.
21. Zeiler MD, Fergus R, editors. Visualizing and understanding convolutional networks. European conference on computer vision; 2014: Springer.
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.
23. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015.
24. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016.
25. Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging. 2014;1(3):034003-.
26. Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Transactions on Medical Imaging. 2016;35(5):1313-21.
27. Cheng J-Z, Ni D, Chou Y-H, Qin J, Tiu C-M, Chang Y-C, et al. Computer-Aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in CT scans. Scientific reports. 2016;6.
28. Rao P, Fereira NA, Srinivasan R, editors. Convolutional

neural networks for lung cancer screening in computed tomography (CT) scans. Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on; 2016: IEEE.
29. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Transactions on Medical Imaging. 2016;35(5):1196-206.
30. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402-10.
31. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annual Review of Biomedical Engineering. 2017;19.
32. Kaur M, Singh Rana AP, Kapoor S, Puri A. Diagnostic value of bone marrow aspiration and biopsy in routine hematology practice. J Clin Diagn Res. 2014;8(8):FC13-6. doi: 10.7860/JCDR/2014/9823.4760.
33. Devet HCW, Koudstaal J, Kwee WS, Willebrand D, Arends JW. Efforts to Improve Interobserver Agreement in Histopathological Grading. Journal of Clinical Epidemiology. 1995;48(7):869-73. doi: Doi 10.1016/0895-4356(94)00225-F.
34. Mohammed EA, Mohamed MM, Far BH, Naugler C. Peripheral blood smear image analysis: A comprehensive review. J Pathol Inform. 2014;5(1):9. doi: 10.4103/2153-3539.129442.
35. Bain BJ. Diagnosis from the blood smear. N Engl J Med. 2005;353(5):498-507. doi: 10.1056/NEJMra043442.
36. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. IEEE Rev Biomed Eng. 2009;2:147-71. doi: 10.1109/RBME.2009.2034865.
37. Theera-Umpon N, Dhompongsa S. Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification. IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society. 2007;11(3):353-9.
38. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al. Proposals for the classification of the

acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol. 1976;33(4):451-8.
39. Osowski S, Siroic R, Markiewicz T, Siwek K. Application of Support Vector Machine and Genetic Algorithm for Improved Blood Cell Recognition. IEEE Trans Instrum Meas. 2009;58(7):2159-68. doi: 10.1109/Tim.2008.2006726.
40. Reta C, Altamirano L, Gonzalez JA, Diaz-Hernandez R, Peregrina H, Olmos I, et al. Segmentation and Classification of Bone Marrow Cells Images Using Contextual Information for Medical Diagnosis of Acute Leukemias. PLoS One. 2015;10(6):e0130805. doi: 10.1371/journal.pone.0130805.
41. Beksac M, Beksac MS, Tipi VB, Duru HA, Karakaş MÜ, Çakar AN. An artificial intelligent diagnostic system on differential recognition of hematopoietic cells from microscopic images. Cytometry. 1997;30(3):145-50.
42. Osowski S, Markiewicz T. Support vector machine for recognition of white blood cells in leukemia. Kernel methods in bioengineering, signal and image processing. 2007;4:93-123.
43. Siroic R, Osowski S, Markiewicz T, Siwek K, editors. Support vector machine and genetic algorithm for efricient blood cell recognition. Instrumentation and Measurement Technology Conference Proceedings, 2007 IMTC 2007 IEEE; 2007: IEEE.
44. Theera-Umpon N, Gader PD, editors. Training neural networks to count white blood cells via a minimum counting error objective function. Pattern Recognition, 2000 Proceedings 15th International Conference on; 2000: IEEE.
45. Theera-Umpon N, editor White blood cell segmentation and classification in microscopic bone marrow images. International Conference on Fuzzy Systems and Knowledge Discovery; 2005: Springer.
46. Escalante HJ, Montes-y-Gómez M, González JA, Gómez-Gil P, Altamirano L, Reyes CA, et al. Acute leukemia classification by ensemble particle swarm model selection. Artificial intelligence in medicine. 2012;55(3):163-75.
47. Staroszczyk T, Osowski S, Markiewicz T, editors. Comparative analysis of feature selection methods for blood cell recognition in leukemia. International Workshop on Machine Learning and Data Mining in Pattern Recognition; 2012: Springer.
48. Lee LH, Mansoor A, Wood B, Nelson H, Higa D, Naugler C. Performance of CellaVision DM96 in leukocyte classification.

J Pathol Inform. 2013;4:14. doi: 10.4103/2153-3539.114205.
49. Kratz A, Bengtsson H-I, Casey JE, Keefe JM, Beatrice GH, Grzybek DY, et al. Performance Evaluation of the CellaVision DM96 System. American Journal of Clinical Pathology. 2005;124(5):770-81. doi: 10.1309/xmb9k0j41lhlatay.
50. Briggs C, Longair I, Slavik M, Thwaite K, Mills R, Thavaraja V, et al. Can automated blood film analysis replace the manual differential? An evaluation of the CellaVision DM96 automated image analysis system. Int J Lab Hematol. 2009;31(1):48-60. doi: 10.1111/j.1751-553X.2007.01002.x.
51. Cornet E, Perol JP, Troussard X. Performance evaluation and relevance of the CellaVision DM96 system in routine analysis and in patients with malignant hematological diseases. Int J Lab Hematol. 2008;30(6):536-42. doi: 10.1111/j.1751-553X.2007.00996.x.
52. Simonyan K, Zisserman A, editors. Two-stream convolutional networks for action recognition in videos. Advances in neural information processing systems; 2014.
53. Sermanet P, Kavukcuoglu K, Chintala S, LeCun Y, editors. Pedestrian detection with unsupervised multi-stage feature learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2013.
54. Mikolov T, Karafiát M, Burget L, Cernocký J, Khudanpur S, editors. Recurrent neural network based language model. Interspeech; 2010.
55. Kainz P, Burgsteiner H, Asslaber M, Ahammer H. Training echo state networks for rotation-invariant bone marrow cell classification. Neural Computing and Applications. 2016:1-16.
56. Wei Y, Xia W, Huang J, Ni B, Dong J, Zhao Y, et al. CNN: Single-label to multi-label. arXiv preprint arXiv:14065726. 2014.
57. Huang C, Li Y, Change Loy C, Tang X, editors. Learning deep representation for imbalanced classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016.
58. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research. 2014;15(1):1929-58.
59. Cornbleet PJ. Clinical utility of the band count. Clinics in

laboratory medicine. 2002;22(1):101-36.
60. He K, Zhang X, Ren S, Sun J, editors. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision; 2015.
61. Al-Janabi S, Huisman A, Van Diest PJ. Digital pathology: current status and future perspectives. Histopathology. 2012;61(1):1-9.
62. Park J. Automatic white blood cell differentiation. University of Missouri-Columbia; 2002.
63. Zheng X, Zhang Y, Shi J, Yu Y, editors. Analysis of leukemia development based on marrow cell images. Image and Signal Processing (CISP), 2011 4th International Congress on; 2011: IEEE.
64. Zheng X, Zhang Y, Shi J, Yu Y, editors. A new method for automatic counting of marrow cells. Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on; 2011: IEEE.
65. Hengen H, Spoor SL, Pandit MC, editors. Analysis of blood and bone marrow smears using digital image processing techniques. Medical Imaging; 2002: International Society for Optics and Photonics.
66. Boser BE, Guyon IM, Vapnik VN, editors. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory; 1992: ACM.
67. Sjöström PJ, Frydel BR, Wahlberg LU. Artificial neural network-aided image analysis system for cell counting. CYTOMETRY-NEW YORK-1999. p. 18-26.
68. Pan C, Park DS, Yoon S, Yang JC. Leukocyte image segmentation using simulated visual attention. Expert Systems with Applications. 2012;39(8):7479-94.
69. Li Y, Zhu R, Mi L, Cao Y, Yao D. Segmentation of white blood cell from acute Lymphoblastic Leukemia images using dual-threshold method. Computational and Mathematical Methods in Medicine. 2016;2016:12.
70. Jiang K, Liao Q-M, Xiong Y. A novel white blood cell segmentation scheme based on feature space clustering. Soft Computing. 2006;10(1):12-9.
71. Madhloom H, Kareem S, Ariffin H, Zaidan A, Alanazi H, Zaidan B. An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold. Journal of Applied Sciences. 2010;10(11):959-66.

72. Won CS, Nam JY, Choe Y. Segmenting cell images: a deterministic relaxation approach. Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis: Springer; 2004. p. 281-91.
73. Dorini LB, Minetto R, Leite NJ, editors. White blood cell segmentation using morphological operators and scale-space analysis. XX Brazilian Symposium on Computer Graphics and Image Processing; 2007: IEEE.
74. Wang W, Song H, editors. Cell cluster image segmentation on form analysis. Natural Computation, 2007 ICNC 2007 Third International Conference on; 2007: IEEE.
75. Ongun G, Halici U, Leblebicioglu K, Atalay V, Beksaç M, Beksaç S, editors. Feature extraction and classification of blood cells for an automated differential blood count system. Neural Networks, 2001 Proceedings IJCNN'01 International Joint Conference on; 2001: IEEE.
76. Ongun G, Halici U, Leblebicioglu K, Atalay V, Beksac M, Beksac S, editors. An automated differential blood count system. Engineering in Medicine and Biology Society, 2001 Proceedings of the 23rd Annual International Conference of the IEEE; 2001: IEEE.
77. Intriligator J, Cavanagh P. The spatial resolution of visual attention. Cognitive psychology. 2001;43(3):171-216.
78. Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research. 2000;40(10):1489-506.
79. Itti L, Koch C. Computational modelling of visual attention. Nature reviews Neuroscience. 2001;2(3):194.
80. Hou X, Zhang L, editors. Saliency detection: A spectral residual approach. Computer Vision and Pattern Recognition, 2007 CVPR'07 IEEE Conference on; 2007: IEEE.
81. Roerdink JB, Meijster A. The watershed transform: Definitions, algorithms and parallelization strategies. Fundamenta informaticae. 2000;41(1, 2):187-228.
82. Hsu D, Lee SH. Digital Imaging in Hematology. Laboratory Hematology Practice. 2012;55:707-18.

초 록

백혈구 백분율 분석은 진단 검사과의 필수 검사 방법으로 다양한 혈액 질환의 진단에 있어 중요한 정보를 제공한다. 하지만 혈액 샘플 간 및 샘플 내 다양성이 존재하여 이 검사 방법을 통한 정확한 진단을 위해서는 고도로 숙련된 의사나 혈액 학자를 필요로 한다. 또한 이 검사 방법은 모두 수기로 진행됨에 따라 지루하고 많은 시간과 비용을 필요로 하여 자동화 시스템 개발에 대한 높은 요구가 있어왔다. 자동화 시스템이 실제 임상 진단 검사과에서 사용이 되기 위해서는 골수 도말 표본에서 백혈구의 분화과정을 세분화 할 수 있어야 한다. 하지만 골수 도말 표본이 가지고 있는 복잡한 특성 때문에 컴퓨터 비전, 이미지 처리 및 기계 학습에서 어려운 문제로 여겨져 왔으며 자동화 시스템 개발에 여러 어려움이 존재한다. 먼저 골수 내 백혈구의 밀도가 높기 때문에 많은 백혈구들이 서로 닿아 있는 문제가 있다. 또한 백혈구의 여러 분화 단계는 전문가도 어려움을 겪을 만큼 차이가 적고 사람 간 및 내 형태 및 크기의 다양성이 존재한다. 도말 표본의 염색 품질에서도 차이를 보일 수 있다.

이 논문에서는 이러한 문제점을 극복한 이중 합성곱 신경망을 이용한 자동 백혈구 백분율 분석 시스템이 제안되었다. 이 시스템은 높은 정확성과 객관성을 보여야 하며 높은 처리율과 효율성을

가져야 한다. 또한 골수 도말이 가지고 있는 어려움들을 극복해야 한다. 이를 위하여 백혈구 검출 및 분류 알고리즘이 개발 및 통합되었다. 먼저 골수 도말 표본의 현미경 사진에서 백혈구 위치 및 단일 백혈구 사진을 검출 하기 위하여 세일리언시 맵과 워터쉐드 변환을 활용하였고, 백혈구 분화 과정을 분류하기 위하여 복잡한 백혈구의 글로벌 및 로컬 한 특징들을 학습할 수 있는 이중 합성곱 신경망을 제안 하였다. 마지막으로 확률 유도 알고리즘을 통하여 이 두 검출 및 분류 알고리즘을 통합하였다. 개발된 시스템은 골수 도말에서 마이엘로이드와 에리스로이드 계열 검출 및 분류를 통하여 성능을 평가하였다.

본 연구를 위하여 총 200 개의 큰 (1388 × 1040) 골수 도말 표본 사진과 2,323 개의 작은 (96 × 96) 단일 백혈구의 사진을 모았고 이를 토대로 검출 및 분류 알고리즘을 개발 및 성능 평가 하였다. 96.09%의 검출 정확성을 보였으며 97.06%의 분류 정확성을 보여줌으로 현존하는 시스템 중 가장 많은 백혈구 분류하고 가장 높은 정확성을 보였다. 두 알고리즘을 통합한 시스템은 100 개의 백혈구 분석을 4-5 초 이내에 완료하였으며 91.3%의 전체 정확성을 보였다. 이 시스템의 소개로 인하여 혈액 진단검사에 있어 새로운 패러다임을 제시하고 의료영상에서의 딥러닝의 가능성을 보여주었다. 또한, 이 시스템을 통하여 한 환자에서 더 많은 백혈구의 분석을 가능하게 할 것이며 이는

진단에 있어 통계적으로 더 유의미한 정보를 제공 할 것으로 예상된다.

핵심어: 백혈구 백분율, 합성곱 신경망, 골수도말, 자동 백혈구 백분율 분석 시스템, 컴퓨터 보조 시스템
학번: 2013-30303