

**Identification of Abnormal Red Blood Cells and Diagnosing Specific Types of Anemia Using Image Processing and Support Vector Machine**

by

**Paul Daniel C. Divina**

**John Philip T. Felices**

A Thesis Report Submitted to the School of EE-ECE-CpE in Partial

Fulfilment of the Requirements for the Degree

Bachelor of Science in Computer Engineering

Mapúa University

June 2018

**Chapter 1**

**INTRODUCTION**

Blood is an important liquid for human because it performs important functions that supports human life. The general components of human blood are red blood cells (RBCs), plasma, platelets, and white blood cells (WBCs). RBCs occupies the largest volume in blood with an approximately 40%. Plasma is a fluid component containing melted salts and proteins. While, Platelet cells are similar particles but are smaller than WBCs and RBCs. WBCs occupy smaller blood volume but is larger than RBCs. (Biradar et al., 2015; Deligiannidis and Arabnia, 2014; Xia and Wu, 2015). Red blood cells, or erythrocytes, are the most abundant cells in the blood of human being body. An RBC is circular in front view and biconcave in side view. In terms of size, it is 7.5 μm in diameter and 2 μm in thickness. These cells work as oxygen carriers throughout the human body (Pinsky et al., 2007; Rakshit and Bhowmik, 2013). In analyzing a person’s RBC its health can be identified. A decrease in the number of RBCs and an abnormality in RBC’s shape is a clear indicator of presence of blood related disorders (Mazalan et al., 2013). Presence of abnormal blood cells indicate various conditions like, Thalassemia, Anemia, Chronic inflammation, liver dysfunction, vitamin deficiency, etc. (Venkatalakshmi and Thilagavathi, 2013; Bunn, 2011).

Anemia is the prominent condition globally. Where a 1.62 billion people or 24.8% of the population are affected by Anemia (Garg et. al, 2016). There are various types of Anemia such as Sickle Cell Anemia, Iron-deficiency Anemia, Vitamin deficiency Anemia, Megaloblastic Anemia, and Hereditary Spherocytosis. Anemia is a condition which develops

when the blood lacks adequate healthy RBCs or even hemoglobin (Sheng and Li, 2010). The diagnosis of Anemia requires some laboratory test such as complete blood cell count, measuring the level of hemoglobin, blood smear analysis, bone marrow examination and other test that a doctor will recommend. To count the number of RBCs automated and manual techniques are performed. The use of Hemocytometer involves the manual counting of blood cells. Another manual method is through blood smear analysis, it is done by placing the smear under the microscope. With the use of microscope counting of cells and morphology of RBCs can be identified (diameter, shape, color, and inclusions). However, Manual techniques leads erroneous results and medical laboratory technicians are put under stress and Anemia can be diagnosed. (Acharya & Kumar, 2017). To solve this, a computer-aided system to automate the detection and identification of RBC in blood smear image with the use of image processing is proposed. This is done by acquiring RBC images and using global threshold to extract regions of RBC. After that noise and holes in the RBCs are removed by using image processing filters. Next, information based from the RBCs’ geometrical properties are extracted. Eventually, with the use of Artificial Neural Network (ANN) classifier the RBCs were classified as normal/abnormal (Tomari, et.al, 2014). There is also a proposed system that based on the feature properties of RBC, visual features were clustered into two main groups, namely shape and texture groups. With the use hybrid neural network based classifier, the extracted images from RBC will determine if the RBC is normal or abnormal. (Lee & Chena, 2014). Additionally, a study was also conducted to classify RBC diseases such as Hereditary Spherocytosis, Hypochromic Anemia, Iron Deficiency Anemia, Megaloblastic Anemia, Normochromic Anemia, and Sickle Cell Anemia in a few seconds. The proposed process is an image processing techniques algorithm is used to separate RBC from other components and extract the blood smear image, to support the process RBCs are categorized into 11 types and 8 attributes. It achieved an accuracy of 98 percent correct identification of type of Anemia when the obtained results were compared with evaluation done by medical experts. (Acharya & Kumar, 2017).

Previous studies focused mainly on classification of RBCs whether it is normal or abnormal and did not specifically identified what kind of abnormal RBC are present. The previous studies lack the ability to classify different RBCs based on morphological features and be able to diagnose specific types of anemia using the identified RBC parameters.

The general purpose of the study is to design and develop a system that would be able to identify morphological abnormalities in red blood cells through the use of image processing and support vector machine. The specific objectives include the following: (1) To create a device that would be able to visually inspect the shapes and variation in sizes of red blood cells from a blood specimen taken for testing; (2) To identify morphological abnormalities in red blood cells using Sobel Edge Detection and Watershed Transform Segmentation; (3) To classify the identified abnormal blood cells into specific types of anemia using Support Vector Machine (SVM).

With this study, it will automate the process and help identify diseases related to the red blood cells more efficiently since it will lessen human intervention of analyzing the specimen. It will be beneficial to doctors, pathologist, hematologist, medical technicians, and other medical related professions once it is implemented to hospitals and other medical institutions. The device will be able to provide help in analyzing RBC parameters and diagnosing various blood abnormality disorders in a much shorter period of time. In that way, preventive measures and proper medication can be immediately provided to the patient that was diagnosed which is necessary to avoid worsening the condition.

The study focuses on the detection of morphological abnormalities of the red blood cells. There are several blood morphological abnormalities that can be depicted from the size and structure of red blood cells but the study is limited to five different types which includes echinocytes, elliptocytes, spherocytes, dacrocytes, and sickle cells. There are lot of blood disorders that may arise when these abnormal blood cells are present to the human body. The study is towards diagnosing specific types anemia based on the morphological abnormalities of RBCs. Different types of Anemia that are focused on the study are Iron deficiency anemia, Myelophthisic anemia and such anemia that falls under Hemolytic anemia which includes Thalassemia, Microangiopathic hemolytic anemia, Hereditary spherocytosis and Sickle cell anemia. Other anemia conditions that are not covered in this study are Sideroblastic anemia, Megaloblastic anemia, and Aplastic anemia. With the study having a limited scope and not being able to cover all possible blood disorders means it cannot be used to conclude a final diagnosis but it serves and can be utilized as a medical device that can be used as a guide in determining possible outcomes.

**Chapter 2**

**REVIEW OF RELATED LITERATURE**

**2.1 Anemia**

According to Keohane, Smith, and Walenga (2016), anemia is defined operationally as a reduction in the hemoglobin content in a person’s blood. This condition is caused by decrease in the number of red blood cells (RBCs) and hematocrit. A decrease in hemoglobin concentration or the number of RBCs does result into a decreased oxygen transport throughout the body tissues. The hemoglobin that is contained in the human’s RBCs is required to transport and deliver oxygen from the lungs to the rest of the body. When there is insufficient amount of oxygen to the body, many tissues and organs can be badly affected. People with anemia commonly experience fatigue and weakness and lack of energy to do tasks.

Anemia affects an estimated 1.62 billion people worldwide (Keohane, Smith, and Walenga, 2016). Anemia is a common condition that affects both men and women and does not depend on any age bracket, race or ethnicity. However, some people have increased their risk of developing anemia based on the lifestyle and conditions that they have. These include people with diets that has poor content in iron and vitamins, chronic diseases such as cancer, kidney disease, diabetes, inflammatory bowel disease, a family history of inherited anemia, chronic infections such as HIV or tuberculosis, and those who have had significant blood loss from injury or surgery. Anemia should not be thought of as a disease but rather as a manifestation or effect of an underlying disease or deficiency of certain substances to the body. It is highly advised that the cause of a patient’s anemia should be investigated properly to provide further understanding of the situation.

**2.1.1 Some Common Types of Anemia**

Based on the book “Hematology Clinical Principles and Applications Fifth Edition” by Keohane, Smith, and Walenga (2016), the following are some common types of anemia.

1. **Iron-deficiency Anemia**

Iron deficiency anemia is a common type of anemia wherein the symptoms are usually related to the decrease in the red blood cells and the level of hemoglobin. Iron is part of the “heme” which is the protein in RBC that binds to oxygen for it to be transported throughout the body. Iron is needed to produce healthy red blood cells. Insufficient iron in the body leads to decrease in the production of hemoglobin which in return leads to producing a smaller and hypochromic or paler red blood cells.

1. **Pernicious Anemia and Other B Vitamin Deficiencies**

Pernicious anemia occurs when the body is incapable of absorbing enough vitamin B12. This happens when there is not enough “intrinsic factor” produced which is a protein produced in stomach that binds with vitamin B12 for it to be absorbed. When the body cannot absorb enough vitamin B12, it cannot produce enough normal RBCs which will eventually lead to anemia.

1. **Hemolytic Anemia**

Hemolytic anemia does happen when the RBCs are being destroyed but there is not enough replacement produced from the bone marrow. Thus, leads to decreased number of red blood cells in the blood which lessens the capability of the blood to supply oxygen throughout the body.

1. **Anemia Caused by Chronic Diseases**

Some chronic diseases can also cause anemia. When these diseases such as kidney disease and chronic infections is present in the body, the inflammatory response of the body is stimulated. The ability of the bone marrow to respond is then decreased. These various conditions in the long term can decreased the production of RBCs in the body.

**2.1.2 Laboratory Diagnosis of Anemia**

Based on the book “Hematology Clinical Principles and Applications Fifth Edition” by Keohane, Smith, and Walenga (2016), the following are the standard laboratory testing that are practiced in diagnosing anemia.

1. **Complete Blood Count with Red Blood Cell Indices**

To detect the presence of anemia, the medical laboratory professional performs a complete blood count (CBC) using an automated hematology analyzer to determine the RBC count, hemoglobin concentration, hematocrit, RBC indices, white blood cell count, and platelet count. The RBC indices include the mean cell volume (MCV), mean cell hemoglobin (MCH), and mean cell hemoglobin concentration (MCHC). The most important of these indices is the MCV, a measure of the average RBC volume in femtoliters (fL).

1. **Reticulocyte Count**

The reticulocyte count serves as an important tool to assess the bone marrow’s ability to increase RBC production in response to an anemia. Reticulocytes are young RBCs that lack a nucleus but still contain residual ribonucleic acid (RNA) to complete the production of hemoglobin. Normally, they circulate peripherally for only 1 day while completing their development. The adult reference interval for the reticulocyte count is 0.5% to 2.5% expressed as a percentage of the total number of RBCs. The newborn reference interval is 1.5% to 6.0%, but these values change to approximately those of an adult within a few weeks after birth.

1. **Peripheral Blood Smear Examination**

A blood smear examination is a blood test that is used to check for abnormalities in the blood cells. The test provides visualization of the three types of cells which are the red blood cells, white blood cells, and platelets. The analysis starts from collection of blood sample, preparation of peripheral blood smear and evaluation of blood cells under a microscope. The variations of features such as size and shape and number of occurrence of each cell were observed and recorded in the test. The pathologists have different approaches of inspecting for each type of cell. The red blood cells are inspected by observing uniform distribution, shape, size, central pallor area. The total count and differential count, nucleus to cytoplasm ratio presence of granules, nuclear lobulation, mature and immature cells are observed in white blood cells. The count, size, and clumps are checked on platelets. Any difference or deviation from the normal appearance of the blood components are assumed to be under abnormal condition which will require further testing and evaluation to produce a concrete diagnosis. The result may vary based on different circumstances. It practically depends on the instruments and methods used in analyzing the sample and mainly on the skill and experience of the technician conducting the test.

1. **Bone Marrow Examination**

The cause of anemia can usually be determined from health history, physical examinations, and peripheral blood test. In case that the cause is undetermined, a bone marrow examination is suggested. A bone marrow examination evaluates hematopoiesis which can determine if there are abnormal cells in the bone marrow. The important results that could be collected from the test include abnormal cellularity, lack of iron on iron stains of bone marrow, and presence of granulomata, fibrosis, infectious agents and tumor cells.

1. **Other Laboratory Tests**

Other laboratory tests that could help in determining and establishing the cause of anemia include the routine urinalysis and microscopic examination and analysis of stool. Urinalysis helps detect hemoglobinuria, hematuria, and hemosiderin. The analysis of stool is used to detect occult blood or intestinal parasites.

**2.2 Morphology of Normal Red Blood Cells**

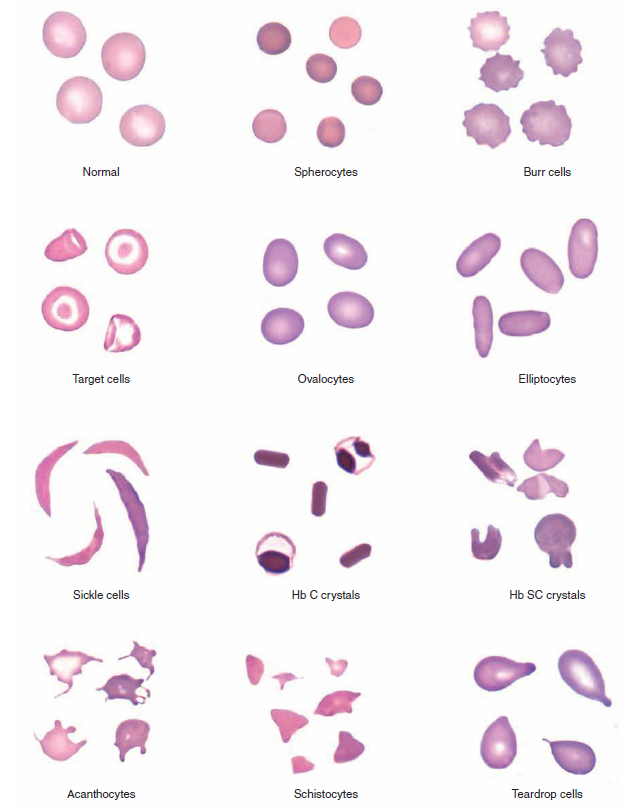
The normal red blood cells are biconcave disc or disc-shaped cell. The center of the cell is compressed so it is thinner than the edge. A single red blood cell is typically 7.2 microns but the normal range measures 6.2 to 8.2 microns. The typical width is 2.0 microns with a range of 2.0 to 2.5 microns. The term commonly used in describing RBCs that has normal size is “normocytic”. The body produces enough RBCs each day with the old cells being removed by the liver or spleen. A normal RBC can approximately survive for 120 days in the blood circulation.

**2.3 Morphological Abnormalities of Red Blood Cells**

The red blood cells are the most abundant cells that is present in a human’s body. Its normal shape is biconcave and disk-shaped. Any abnormality in the shape of RBC may indicate the presence of a disease (Dalvi and Vernekar, 2016). These abnormalities may be differed based on variation in size (anisocytosis), variation in shape (poikilocytosis), and variation in color. Table 2.1 shows the tabulated description of abnormal red blood cells and the commonly associated diseases to it. The acquired data is based on the book “Hematology Clinical Principles and Applications Fifth Edition” by Keohane, Smith, and Walenga (2016).

**Table 2.1** Description of Red Blood Cells Abnormalities and Commonly Associated Diseases

|  |  |  |
| --- | --- | --- |
| **RBC Abnormality** | **Cell Description** | **Commonly Asssociated Disease** |
| Macrocyte | Large RBC (.8 mm in diameter), MCV .100 fL | Megaloblastic anemia  Myelodysplastic syndrome  Chronic liver disease  Bone marrow failure  Reticulocytosis |
| Microcyte | Small RBC (,6 mm in diameter), MCV ,80 fL | Iron deficiency anemia  Anemia of chronic inflammation  Sideroblastic anemia  Thalassemia/Hb E disease and trait |
| Spherocyte | Small, round, dense RBC with no central pallor | Hereditary spherocytosis  Immune hemolytic anemia  Extensive burns (along with schistocytes) |
| Elliptocyte, ovalocyte | Elliptical (cigar-shaped), oval (egg-shaped), RBC | Hereditary elliptocytosis or ovalocytosis  Iron deficiency anemia  Thalassemia major  Myelophthisic anemias |
| Stomatocyte | RBC with slit-like area of central pallor | Hereditary stomatocytosis  Rh deficiency syndrome  Acquired stomatocytosis (liver disease, alcoholism)  Artifact |
| Sickle cell | Thin, dense, elongated RBC pointed at each end;  may be curved | Sickle cell anemia  Sickle cell–b-thalassemia |
| Hb C crystal | Hexagonal crystal of dense hemoglobin formed within the  RBC membrane | Hb C disease |
| Hb SC crystal | Fingerlike or quartz-like crystal of dense hemoglobin  protruding from the RBC membrane | Hb SC disease |
| Target cell (codocyte) | RBC with hemoglobin concentrated in the center and  around the periphery resembling a target | Liver disease  Hemoglobinopathies  Thalassemia |
| Schistocyte (schizocyte) | Fragmented RBC due to rupture in the peripheral circulation | Microangiopathic hemolytic anemia\* (along with microspherocytes)  Macroangiopathic hemolytic anemia  Extensive burns (along with microspherocytes) |
| Burr cell (echinocyte) | RBC with blunt or pointed, short projections that are usually  evenly spaced over the surface of cell; present in  all fields of blood film but in variable numbers per field† | Uremia  Pyruvate kinase deficiency |
| Teardrop cell (dacryocyte) | RBC with a single pointed extension resembling a  teardrop or pear | Primary myelofibrosis  Myelophthisic anemia  Thalassemia  Megaloblastic anemia |



**Figure 2.1** Red blood cells (RBCs): varied RBC shapes and inclusions. Hb, Hemoglobin. (Modified from Rodak BF, Carr JH: Clinical hematology atlas, ed 4, St Louis, 2013, Elsevier, Saunders.)

**2.4 Important Morphological Features of Red Blood Cells**

In the study conducted by Dalvi and Vernekar (2016) on “Computer Aided Detection of Abnormal Red Blood Cells”, they specified different features or parameters that are significant in classifying abnormal red blood cells. They used Decision Tree and Artificial Neural Network (ANN) classifiers. Another study from Acharya and Kumar (2017) on their research about “Identification and Red Blood Cell Classification using Computer Aided System to Diagnose Blood Disorders” which uses image processing for feature extraction included some other attributes to classify RBCs. Features used on both researches are:

1. **Area** - number of pixels enclosed inside the RBC’s boundary.
2. **Perimeter** - Sum of pixels around the boundary of the RBC.
3. **Diameter** – It is the ratio between area and perimeter.

Diameter = (4\* Area)/perimeter **(2.1)**

1. **Shape Geometric Feature** – The RBCs will be usually oval in shape. Normal cells, Hypochromic Macrocyte will be circular in shape which is a special oval with equal diameter. The factor is computed using

SGF = Length of the Major Axis of RBC/ Length of the Minor Axis of RBC **(2.2)**

1. **Form Factor** – It is a measure of irregularity. It is given as the ratio between Area and perimeter. Form factor value is computed using

Form Factor = (4\*3.14\* area)/perimeter^2 **(2.3)**

Additional RBCs features used in Acharya and Kumar’s research include:

1. **Area Proportion** - Area Proportion: It is the ratio between the central pallor area to the cell area. It is used to differentiate a Hypochromic cell from the other cell as the Hypochromic cell will have larger central pallor.

Area Proportion = Central Pallor Area / Cell Area (**2.4)**

1. **Deviation** – It is the ratio between the shape geometric factor and area of the object. The equation to compute it is given

Deviation = SGF / Cell Area (**2.5)**

1. **Central Pallor** – The central pallor is the bionconcave disk shape at the center of RBC. This feature suggests if the cell has a central pallor or not so it is set to either 1 or 0.

**2.5 Image Processing**

Image Processing is a technique to enhance raw images. Also, it refers to processing of digital image, using a digital computer it can remove any noise or irregularities present. Image processing mainly deals with image acquisition, image enhancement, image segmentation, feature extraction, image classification etc. (Chitrasdevi and Srimathi, 2014).

The standard method to count and identify RBC is a microscopic-based evaluation. Despite its long clinical success, this method requires an expertise to manually classify the cells which is tedious, time-consuming and qualitative process (Venkatalakshmi and Thilagavathi, 2013). Furthermore, this method is prone to inaccuracy, inconsistency and due to its poor reliability, it may lead to false diagnosis (Tomari et.al, 2014).

To overcome the problem, automated detection using image processing is proposed it is less time consuming and error free than manual detection (Biswas et.al, 2016). Through image processing it will aid doctors in analyzing and visualizing images to understand morphological abnormalities in RBCs (Hamouda et. al, 2012).

**2.6 Image Filtering**

Noise Filtering is a pre-processing step that aims to remove various types of noises or unwanted regions from the images. Also, to enhance the image for a better view of red blood cells in the blood sample.

**2.7 Image Segmentation**

In image processing, the separation of homogenous area in different regions is known as image segmentation. Blood cell segmentation is a key part to analyze different blood cells it helps to identify and diagnose various pathological conditions. Further, counting the different cell, shapes measurement and abnormal cell detection in blood smears. The main objective of the blood cell segmentation is border detection or boundary detection of cells. These regions based information about the cells are incorporate for the further process.

A simple segmentation technique is an intensity based segmentation with a gray or color threshold value, which may not detect the objects correctly if many objects have almost same gray values. Morphology based segmentation technique can be used to segment objects whose shapes may be clearly defined, but maybe unsuccessful when the objects have intricate shapes (Kulkarni, 2012, Sheerba et.al, 2012).

**2.8 Edge Detection**

Edge detection is an image processing technique for finding the boundaries of objects within images. An edge is the separation between an object and its background, and points out the boundary between overlapping object. There are several types of edge detectors like Canny Operator, LoG Operator, Prewitt Operator, Robert’s Operator, Sobel Operator and Zerocross Operator (Rakshit and Bhowmik, 2013).

**2.9 Sobel Edge Detection Method**

Sobel method performs 2-D spatial gradient measurement on an image and so emphasizes regions of high spatial frequency that correspond to edges. The convolution mask of Sobel operator which are used to obtain the gradient magnitude of the image from the original (Biswas et.al, 2016).

Sobel method is a discrete differentiation operator, computing an approximation of the gradient of images function. In theory, the operator consists of pair of 3x3 convolution kernels (Sobel et. al, 1968 and Yu et.al, 2017).

**2.10 Watershed Transform**

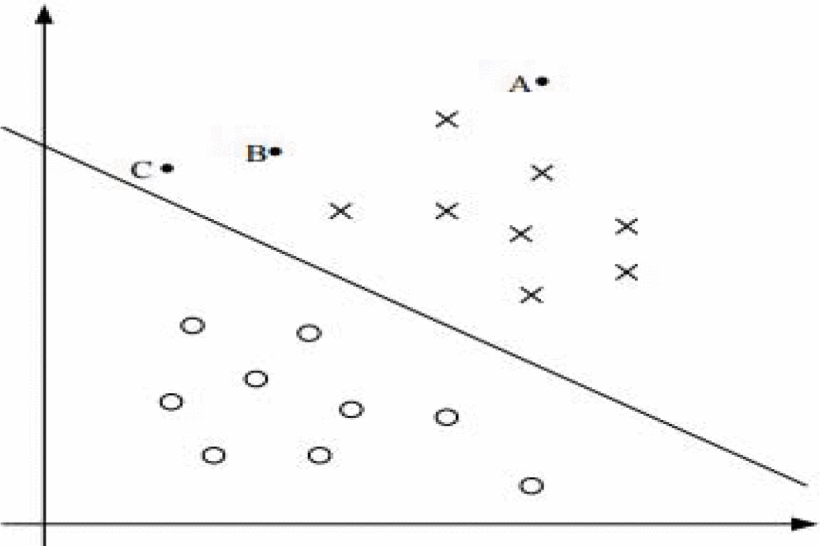
Separating touching objects in an image is one of the most difficult challenges in image processing. A solution to this problem is to employ watershed transform. This method tries to search "catchment basins" and "watershed ridge lines" in an image by treating it as a surface; where dark pixels are low and light pixels are high (Fadhel et.al, 2017).

According to Ghane, Vard, Talebi, and Nematollahy (2017) on their study Segmentation of White Blood Cells from Microscopic Images using a Novel Combination of K-means Clustering and Modified Watershed Algorithm, the researchers stated that Watershed Transform is a good segmentation technique that can segment overlapping objects whose regional maxima can be identified. Especially, in a blood smear images which consist of different components that needs to be removed.

**2.11 Support Vector Machine**

Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, and are used for classification (machine learning) and regression analysis. The goal of an SVM model is to predict which category a particular subject or individual belongs to, based on training set examples.

SVM has an algorithm called hyperplane for it to be able to identify and analyze patterns. Its main concept is to separate different classification data for different objects. Also, it is defined as the decision boundary for SVM. Figure 2.2 shows how SVM are classified. The ‘x’ and ‘o’ are the different data for an object and the line drawn between them is the hyperplane. (Anna Monica M. De Los Reyes, Anna Camille A. Reyes, Jumelyn L. Torres, Dionis A. Padilla, Jocelyn Villaverde, 2016)



**Figure 2.2** SVM Graphical Representation

In the study of Jameela Ali Akrimi, Loay E. George, Azizah Suliman and Abdul Rahim Ahmad (2014) about Classification Red Blood Cells Using Support Vector Machine, image processing techniques that use the optimization segmentation and mean filter play an important role in obtaining the geometric, texture and color features related to RBC images by using a photo imaging microscope. The support vector machine, which is an advanced kernel-based technique, is used to classify RBC whether it is normal or abnormal. The study achieved a very good accuracy rates with validation measure of sensitivity of 100%, specificity of 0.998% and Kappa of 0.9944.

**Feature Extraction**

Features of acquired images of RBC will help in classifying a normal or abnormal RBC. These features are shown in Table 2.2.

**Table 2.2** List of Features of RBC

|  |  |  |
| --- | --- | --- |
| ID | Features | Description |
| 1 | Area (A) | * Sum of pixels enclosed by cell boundary |
| 2 | Perimeter (P) | * Sum of perimeter pixels |
| 3 | Diameter (D) | * Ratio between area and perimeter * Area/(4\*perimeter) |
| 4 | Shape Geometric Factor (SGF) | * Proportion of peripheral oval’s diameter * SGF = Large\_diameter/small\_diameter |
| 5 | Deviation Value (DV) | * ratio between the shape geometric factor and area of the cell * SGF/A |
| 6 | Central Pallor (CP) | * 1 if central pallor is present * 0 if there is no central pallor |

**Table 2.3** Classification of Different RBC

Table 2.3 shows the different type of RBC tested in the study. The healthy RBC and 5 abnormal RBC are studied.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class ID | Name | A | P | D | SGF | DV | CP | Image |
| 1 | Normal RBC |  |  | >6.0  And  <8.0 | <1.2 |  | 1 |  |
| 2 | Echinocytes (Burr Cell) |  |  | <6 | <1.2 |  | 1 |  |
| 3 | Elliptocytes (Elongated) |  |  |  | >1.8 |  | 1 |  |
| 4 | Dacrocytes (Tear Drop cell) |  |  |  | <1.8  And  >1.2 |  | 1 |  |
| 5 | Spherocytes |  |  |  | <1.2 |  | 0 |  |
| 6 | Sickle Cell |  |  |  | >1.2 | >0.20 | 0 |  |

* If SGF is larger than 1.2, cell has elongation (elliptocytes) and if SGF is less than 1.2 cell has a circular shape.
* Deviation value of sickle cell is greater than 0.20

**Table 2.4** Classification of Anemia Conditions Bases on RBC

Table 2.4 shows a summary of what Anemia condition does this RBCs have.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class ID | Name | Iron Defi-ciency Anemia | Thalassemia | Microan-giopathic Hemolytic Anemia | Hereditary Spher-ocytosis | Myelo-phthisic Anemias | Sickle Cell Anemia |
| 1 | Normal RBC |  |  |  |  |  |  |
| 2 | Echinocytes (Burr Cell) |  |  | ✓ |  |  |  |
| 3 | Elliptocytes (Elongated) | ✓ | ✓ |  |  |  | ✓ |
| 4 | Dacrocytes (Tear Drop cell) |  | ✓ |  |  | ✓ |  |
| 5 | Spherocytes |  |  |  | ✓ |  |  |
| 6 | Sickle Cell |  |  |  |  |  | ✓ |

A check mark (✓) means a presence of Anemia conditions in this type of abnormal RBC.

**Statistical Analysis**

To be able to determine the accuracy of the built device, applying confusion matrix will greatly help. Accuracy is the measure of how often the system is correct in classifying. Confusion matrix, maybe labelled as confusing or being confused, but relatively it is simple to understand. A confusion matrix is a table that describes the performance of a machine or mostly classifier with a set of data for which some values are known. The rates in this study to be used are base by the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

* True positives (TP): Case that the classifier correctly identified an abnormal RBC.
* True negatives (TN): Instance, that the classifier correctly stated that there is no abnormal RBC.
* False positives (FP): Occurrence, of predicting that the classifier detected an abnormal RBC even though there weren’t any abnormalities.
* False negatives (FN): Event, when the classifier predicted that there is no RBC abnormality even though there was an abnormality.

**Chapter 3**

**METHODOLOGY**

**Identification of Abnormal Red Blood Cells and Diagnosing Specific Types of Anemia Using Image Processing and Support Vector Machine**

**Research Methodology**



**Figure 3.1** Methodology

Figure 3.1 shows the constructive research methodology which was adapted by the researchers in doing the study. This type of research method aims to solve the problem while also providing scientific and theoretical contribution. The constructive research steps involve: (1) selecting a relevant problem; (2) obtaining a comprehensive understanding of the area of study; (3) designing a solution to the problem; (4) demonstrating the viability of the solution; (5) linking the results back to the theory and demonstrating their practical contribution; (6) examining the general usability of the result. The researchers have selected the problem about the manual analysis of blood smear to examine a blood with anemia. The process is time-consuming, laborious, and prone to error wherein an automated system can solve these problems. Researchers obtained a comprehensive understanding of the area of study by reviewing several related literatures that could help in developing a solution to the problem. After gathering information, the researchers come up with a solution and was able to create the conceptual framework which involves using image processing techniques such as Sobel operator, Watershed algorithm and Support vector machine. The development of the hardware and software was then designed. For the hardware development, the components to be used are listed including the use of raspberry pi and camera module and then possible hardware setup was constructed. The software development will be done by applying different processes and algorithms that was picked for the system. After developing both hardware and software, implementation will be done to create the system. Then, data gathering will follow which involves testing the system. The system will be trained with different samples. Conclusion will be made based on the data and results produced by the system.

**Conceptual Framework**



**Figure 3.2** Conceptual Framework

On figure 3.2 shows the conceptual framework of the system. An image of blood specimen taken for testing will be used as input. The inputted image will then proceed to multiple processes including image pre-processing, edge detection, region selection, and feature extraction. These were done using three different image processing algorithms namely Sobel operator, Watershed algorithm, and Support Vector Machine. At image pre-processing, the image is taken as input into the system and converted into grayscale by eliminating the hue and saturation components and retaining the luminance. Edge detection is the process of identifying discontinuities in images. This is done to see the demarcation between the blood and the background and the boundary between overlapping images. The region selection will measure properties of the image regions such as area, centroid and bounding box. At feature extraction, the features of the red blood cells have been extracted to be able to classify the cells into different types of anemia. After all processes were executed, the output that will be produced are the summary of the blood morphological characteristics and its anemia classification based on the result of the gathered characteristic using the support vector machine algorithm. The characteristics that will be included in the summary are area, perimeter, diameter, shape geometric factor (SGF), deviation value and central pallor. Based on these characteristics, the system will classify six different types of anemia namely Iron-deficiency anemia, Thalassemia, Microangiopathic hemolytic anemia, Hereditary spherocytosis, Myelophthisic anemia, and Sickle cell anemia.

**HARDWARE DEVELOPMENT**

Raspberry Pi

LCD Monitor

Microscope

Wireless Keyboard with touchpad

Raspberry Pi Camera

Power Supply

**Figure 3.3** Block Diagram of the System

Figure 3.3 shows the block diagram of the components used in the design of hardware. A microscope is needed to view the microscopic morphology of RBCs. The Raspberry Pi Camera is needed to capture the image from the sample and it will be sent to the Raspberry Pi. A lighting will be used to provide enough lighting for the setup which is already provided on the microscope. The Raspberry Pi will be the main part of the system which it will identify the morphology of RBC and classify Anemia conditions. The acquired image, identified morphology and classification of RBC will be outputted in the monitor. To control the Raspberry Pi a wireless keyboard and mouse is needed. Lastly, a power source will be used to supply power with devices that needed electricity.

**Table 3.1** Tabulated List of Materials

|  |  |
| --- | --- |
| **List of Materials** | **Description** |
| Raspberry Pi | * The microcontroller to be used on the system. * Full sized HDMI port for display output * 4 USB 2.0 port for mouse and keyboard input |
| Raspberry Pi Camera Module V2 | * The main camera used for image capture. * This camera will be mounted to the eyepiece of the microscope. * 8 MP camera * Compatible with the Raspbian OS |
| Wireless Keyboard with Touchpad | * Wireless keyboard connected to the Raspberry Pi for inputting alphanumeric characters. |
| LCD Display | * Used as a display for the Raspberry Pi |
| Power Supply | * The standard Raspberry Pi power supply will be used which supplies a steady 5V and 2.5A. * Capable of supporting other peripherals connected to the Raspberry Pi. |
| Microscope | * A microscope with 100X – 1200X magnification will be used for the system. |

Table 3.1 shows the materials used to construct the hardware design of the research. It also shows brief discussion of the component’s features and the main purpose as a part of the system.

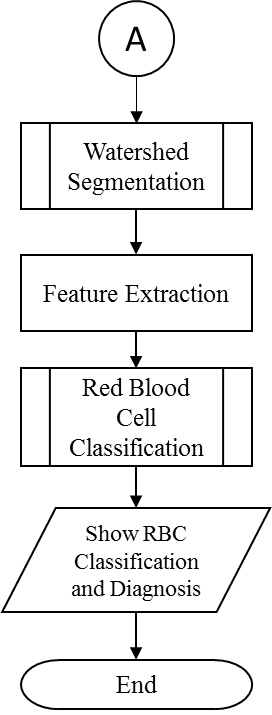
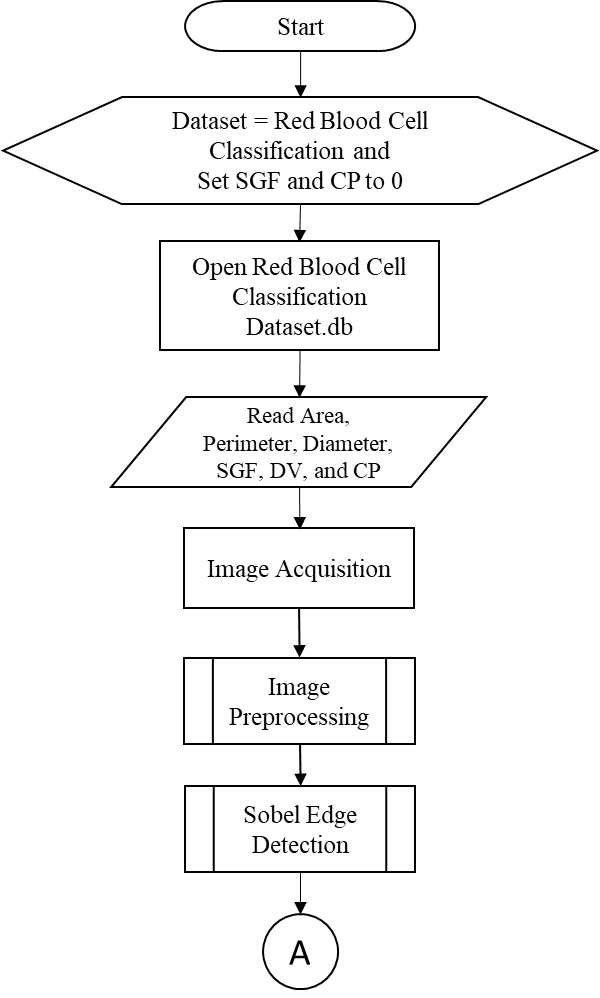
**Proposed Hardware Set-up**



**Figure 3.4** Hardware Set-up

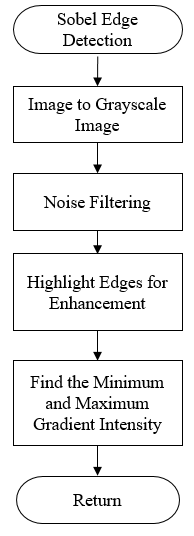
Figure 3.4 shows the proposed hardware set-up for the system. As shown on the figure, the camera will be mounted on the eyepiece of the microscope to be able to capture the microscopic view of the blood specimen. The box contains the main components of the system which are the raspberry pi and LCD display. They are enclosed to a casing for it to look more compact and as protection to the hardware components. The box will be designed in a manner that the ports will be easily accessed to connect peripheral devices such as the wireless keyboard and mouse.

**SOFTWARE DEVELOPMENT**

****

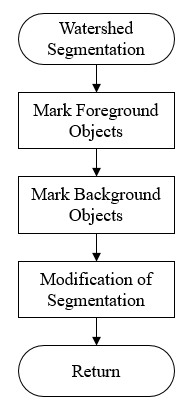
**Figure 3.5** Flowchart of the Main Program

Figure 3.5 show the Flowchart of the Main Program, first initialize the name of the dataset and SGF and CP to 0. After that, open the dataset Red Blood Cell Classification and read the dataset. Next, is capturing images of acquired RBC samples. Followed by detecting edges using Sobel Algorithm. Followed by, segmenting the RBC with Watershed Segmentation. Next, processing the image feature extraction will be done such as getting the RBC’s Area, Perimeter, Diameter, Shape Geometric Factor, Deviation Value and Central Pallor. It will now classify what kind of RBC the sample is and its diagnosis will be outputted.



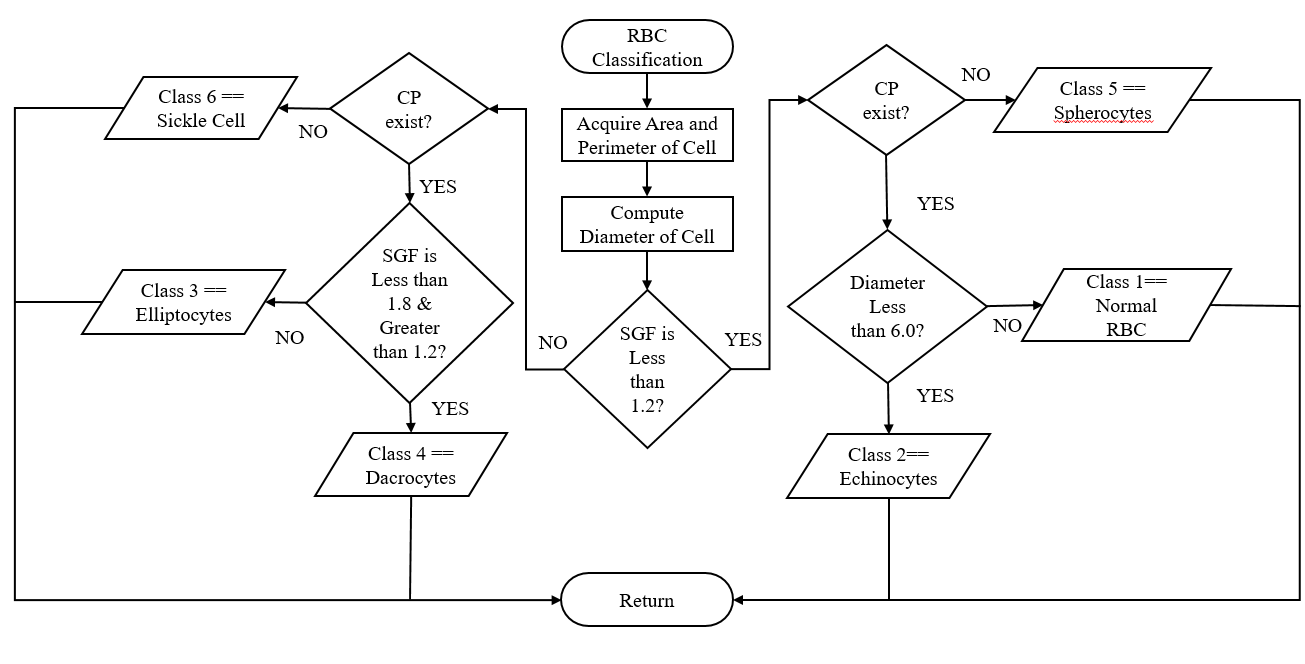
**Figure 3.6** Flowchart of Sobel Edge Detection

Figure 3.6, Flowchart of Sobel Edge Detection, is done after preprocessing. First, the image will now be converted to Grayscale. Converting to Grayscale is needed to easily locate edges for the Edge Detection Process and Segmentation process. Then with Sobel filtering out noise or removing unwanted regions of the image. Next, highlighting Edges for enhancement, these edges will be highlighted to emphasize the boundary of cell. Finding the minimum and maximum gradient intensity is last process for it to find the direction of the largest increase from light to dark and the rate of change in that direction.



**Figure 3.7** Flowchart of the Watershed Segmentation

Figure 3.7, Flowchart of the Watershed Segmentation, after applying Sobel Edge detection, segmenting the image can now be done. First marking foreground objects in the image which are connected blobs of pixels inside each of the foreground objects. Next is marking the background objects which are pixels not used in any part of the object. Lastly, modifying of segmentation to only have minima at the background and foreground marker locations.



**Figure 3.8** Flowchart of Red Blood Cell Classification

Figure 3.8, Flowchart of Red Blood Cell Classification, first acquiring the Area and Perimeter of the cell and get the Diameter. Next, comparing the Shape Geometric Factor if it has less 1.3 SGF then it has a circular shape if not it has an elongation. If the SGF is less than 1.3 then it will go to another decision which will as if there is a Central Pallor in the cell, if there no CP then it is a Spherocytes which is a Class 5. If CP exist, a decision which will ask if diameter is less than 6.0 if it is less than it Echinocytes which is labeled as Class 2. And if not, it is a Class 1, Normal RBC. Moreover, if the SGF of cell is greater than 1.3 it will ask again if there is a Central Pallor if there is no Central Pallor and it has an elongation greater than 1.3 then it is a Sickle Cell which is a Class 6. Next, a decision asking if SGF is less than 1.8 but greater than 1.2 if it is not in the range then it is a Class 3 Elliptocytes and it is in the range then it is a Class 4 Dacrocytes.

**Proposed Graphical User Interface (GUI) of the System**

Analyze Red Blood Cell

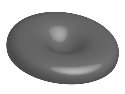
**Red Blood Cell Classification**

User Manual

About

Exit

Summary of Red Blood Cell and Anemia Conditions



**Figure 3.9** Propose GUI of the System

Figure 3.9, Flowchart of the propose GUI of the system it has 5 options, Analyzing a red blood cell means undergoing the main process of the system. Next, A User Manual, a guide on how to use the system. Summary of Red Blood Cell and Anemia Conditions it is an overview of the RBC that can be tested in the machine as well with its features and morphology and shows different kinds of anemia condition. The About, is the information about the system and its inventor. Lastly, exiting or shutting down the machine.

**Data Gathering Procedure**

For the preparation of specimen, the researchers will request for different blood sample at the Philippine General Hospital (PGH). For research purposes, students could request for blood sample on the hospital which takes 1-2 months for approval.

a.) The camera module will be mounted on the eyepiece of the microscope.

b.) The blood sample will be placed on the stage of the microscope. A total of 120 samples will be tested with at least 20 sample for each classification.

c.) The camera will take pictures of the blood and then send it to the microcontroller to apply different processes.

d.) Image pre-processing, edge detection, region selection, and feature extraction will be applied to the image using Sobel operator and watershed algorithm.

e.) Measurement of parameters such as area, perimeter, diameter, and shape geometric feature (SGH) will be extracted and be used by the support vector machine to determine the classification of the sample.

f.) The LCD display will show the result including the parameter values and the classification.

**Testing Tables**

**Training Procedure**

**Table 3.5** Training of SVM for Class 1 RBC

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| RBC Sample # | Area | Perimeter | Diameter | Shape Geometric Factor | Deviation Value | Central Pallor |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| . |  |  |  |  |  |  |
| . |  |  |  |  |  |  |
| 20 |  |  |  |  |  |  |

**Table 3.6** Training of SVM for Class 2 RBC

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| RBC Sample # | Area | Perimeter | Diameter | Shape Geometric Factor | Deviation Value | Central Pallor |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| . |  |  |  |  |  |  |
| . |  |  |  |  |  |  |
| 20 |  |  |  |  |  |  |

For Table 3.5 and 3.6, it will train the SVM by inputting diagnosed blood smear images of different classes in separate trainings and teach the machine to understand normal and abnormal RBC. As well, determine the 7 features to successfully classify a RBC. The desired data for each feature is shown in Table 2.2 and 2.3. Same training process will be done for the remaining classes 3-6.

**Table 3.7** Testing and Classifying of RBC in the Constructed Machine

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| RBC Sample # | A | P | D | SGF | DV | CP | Class | Anemia Condition |
| 1 |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |
| . |  |  |  |  |  |  |  |  |
| . |  |  |  |  |  |  |  |  |
| 20 |  |  |  |  |  |  |  |  |

Table 3.7 shows the Table of testing and classifying RBC in the constructed machine; 20 samples will be used for different classes. It will then identify its features, classify its class and diagnose an Anemia condition base on the morphology and features of RBC. The features and Anemia diagnosis is based on Table 2.3 and 2.4 respectively.

**Statistical Analysis**

**Table 3.8** Confusion Matrix for the Blood Sample Classification

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Predicted Classification of Blood Sample** | | | | | | | |
|  | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Class 7 | Unknown |
| **Actual Classification**  **of Blood Sample** | Class 1 |  |  |  |  |  |  |  |  |
| Class 2 |  |  |  |  |  |  |  |  |
| Class 3 |  |  |  |  |  |  |  |  |
| Class 4 |  |  |  |  |  |  |  |  |
| Class 5 |  |  |  |  |  |  |  |  |
| Class 6 |  |  |  |  |  |  |  |  |
| Class 7 |  |  |  |  |  |  |  |  |
| Unknown |  |  |  |  |  |  |  |  |

Table 3.8 shows the confusion matrix for the blood classification. This table will be used to evaluate the ability of the system to successfully classify the RBC samples. For each class, 20 samples will be tested, and the result will be tabulated. The cases True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are used to interpret the values in a confusion matrix. In this study these cases mean the following: a.) True positive (TP), the system correctly identified the RBC sample class; b.) True negative (TN), the system correctly identified that the sample does not belong to a certain class; c.) False positive (FP), the system identified that the RBC sample belongs to certain class but it belongs to another class; d.) False negative (FN), the system identified that the RBC sample does not belong to a certain class but it actually does. The total for each case are required in getting the accuracy of the system.

Accuracy is the measure of how often the system is correct in classifying. The formula of the accuracy of the system:

**Accuracy of the System = (3.1)**

**References**

Acharya, V., & Kumar, P. (2017). Identification and red blood cell classification using computer aided system to diagnose blood disorders. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi:10.1109/icacci.2017.8126155

Akrimi, J. A., Suliman, A., George, L. E., & Ahmad, A. R. (2014). Classification red blood cells using support vector machine. Proceedings of the 6th International Conference on Information Technology and Multimedia. doi:10.1109/icimu.2014.7066642

Anna Monica M. De Los Reyes, Reyes, A. C., Torres, J. L., Padilla, D. A., & Villaverde, J. (2016). Detection of Aedes Aegypti mosquito by digital image processing techniques and support vector machine. 2016 IEEE Region 10 Conference (TENCON). doi:10.1109/tencon.2016.7848448

Bain, B. J. (2017). Rodaks Hematology: Clinical Principles and Applications. By E.M.Keohane, L.J.Smith and J.M.Walenga (Eds), Elsevier, St Louis, 2016. British Journal of Haematology, 177(4), 662-662. doi:10.1111/bjh.14687

Biradar, N., Dewal, M.L., Rohit, M.K., (2015). Speckle noise reduction in b-mode echocardiographic images: a comparison. IETE Tech. Rev. 32, 435–453.

Biswas, Soumen, and Dibyendu Ghoshal. (2016). "Blood cell detection usingthresholding estimation based watershed transformation with Sobel filter in frequency domain." Procedia Computer Science 89: 651-657.

Bunn HF. (2011).Approach to the anemias. In: Goldman L, Schafer AI, eds. Cecil Medicine. 24th ed. Philadelphia, Pa: Saunders Elsevier; chap 161

Chitradevi, B., Srimathi, P. (2014). An Overview on Image Processing Techniques International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11,

Dalvi, P. T., & Vernekar, N. (2016). Computer aided detection of abnormal red blood cells. 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). doi:10.1109/rteict.2016.7808132

Deligiannidis, L., Arabnia, H., (2014). Emerging Trends in Image Processing, Computer Vision and Pattern Recognition.

Sheeba, Nagar, Thamburaj, Mammen, (2012). “Segmentation of Peripheral Blood Smear Images Using Tissue-Like P Systems”, International Journal of Natural Computing Research, 3(1), 16-27, January-March

Fadhel, M., Humaidi, A. J., & Razzaqoleiwi, S. (2017). Image processing-based diagnosis of sickle cell anemia in erythrocytes. 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT). doi:10.1109/ntict.2017.7976124

Garg, Nigam, Agrawal, Nigam and Agrawal (2016). “Iron Carboxymaltose: A Safe and Effective Molecule to Combat Anemia in Pregnancy”, International Journal of Current Research and Academic Review ISSN: 2347-3215 Volume 4 Number 2 pp. 124-130

Ghane N., Vard A., Talebi A., Nematollahy P. (2017). :Segmentation of White Blood Cells from Microscopic Images using a Novel Combination of K-means Clustering and Modified Watershed Algorithm”, Journal of Medical Signals and Sensors, 7(2), pp. 92-101.

Hamouda, Khedr, and Ramadan. (2012). “Automated Red Blood Cell Counting” International Journal of Computing Science, VOL. 1, NO. 2, FEBRUARY, ISSN (Oline): ISSN (Print): 2164-1366,

Kulkarni, (2012). Color Thresholding Method for Image Segmentation of Natural Images, I.J. Image, Graphics and Signal Processing, pp. 28-34.

Lee, H., & Chen, Y. P. (2014). Cell morphology based classification for red cells in blood smear images. Pattern Recognition Letters, 49, 155-161. doi:10.1016/j.patrec.2014.06.010

Mazalan, Mahmood, Razak, (2013). “Automated Red Blood Cells Counting in Peripheral Blood Smear Image Using Circular Hough Transform”, First International Conference on Artificial Intelligence, Modelling & Simulation.

Pinsky, Brochard and Mancebo, (2007). “Applied Physiology in Intensive CareMedicine”. Springer. 229-238.

Rakshit, P., & Bhowmik, K. (2013). Detection of Abnormal Findings in Human RBC in Diagnosing Sickle Cell Anaemia Using Image Processing. Procedia Technology, 10, 28-36. doi:10.1016/j.protcy.2013.12.333

Sheng, D., &. Li, C. (2010). Intelligent Optimization Methods for HighDimensional Data Classification for Support Vector Machines. Intelligent Information Management, 2(6), 2-14, 16-33.

Sobel, Irwin, and Gary Feldman. (1968) "A 3x3 isotropic gradient operator for image processing." a talk at the Stanford Artificial Project: 271-272

Tomari, R., Zakaria, W. N., Jamil, M. M., nor, F. M., & Fuad, N. F. (2014). Computer Aided System for Red Blood Cell Classification in Blood Smear Image. Procedia Computer Science, 42, 206-213. doi:10.1016/j.procs.2014.11.053

Venkatalakshmi, B. and K. Thilagavathi. (2013). Automatic red blood cell counting using Hough transform. IEEE Conference on Information & Communication Technologies (ICT). 2013.

Xia, Y., Wu, X.-J., (2015). Adaptive ball particle filter and its application to visual tracking. IETE Tech. Rev. 32, 462–470

Yu, Xiaolong, et al. (2017) "Image edge detection based tool condition monitoring with morphological component analysis." ISA transactions 69: 315-322.