Revelation of Cancer Cells based on Microscopic Images in Blood Sample

1 Ponni, 2 Mowli, 3 Jaishree, 4 Kanishka Harini, 5 Mahalakshmi Bai
1 Faculty of Electronics and Communication Engineering, Kings College of Engineering, Punalkulam
2,3,4,5 Students of Electronics and Communication Engineering, Kings College of Engineering, Punalkulam

Abstract: At the moment, identification of blood disorders is through visual inspection of microscopic images of blood cells. From the identification of blood disorders, it can lead to classification of certain diseases related to blood. This paper describes a preliminary study of developing a detection of leukemia types using microscopic blood sample images. Analyzing through images is very important as from images, diseases can be detected and diagnosed at earlier stage. From there, further actions like controlling, monitoring and prevention of diseases can be done. Images are used as they are cheap and do not require expensive testing and lab equipments. The system will focus on white blood cells disease, leukemia. The system will use features in microscopic images and examine changes on texture, geometry, color and statistical analysis. Changes in these features will be used as a classifier input. A literature review has been done and Reinforcement Learning is proposed to classify types of leukemia. A little discussion about issues involved by researchers also has been prepared.

Keywords: White Blood Cell, Microscopic Images, Leukemia, Reinforcement Learning

I. INTRODUCTION

Medical imaging has become one of the most important visualization and interpretation methods in biology and medicine over the past decade. This time has witnessed a tremendous development of new, powerful instruments for detecting, storing, transmitting, analyzing, and displaying medical images. This has led to a huge growth in the application of digital image processing techniques [1] for solving medical problems. The most challenging aspect of medical imaging lies in the development of integrated systems for the use of the clinical sector. Design, implementation, and validation of complex medical systems require a tight interdisciplinary collaboration between physicians and engineers. Main objective of analyzing through images is to gather information, detection of diseases, diagnosis diseases, control and therapy, monitoring and evaluation [2].

At the moment, identification of blood disorders is through visual inspection of microscopic images of blood cells. From the identification of blood disorders, it can lead to classification of certain diseases related to blood. One of the most feared by the human disease is cancer. Leukemia is a type of blood cancer, and if it is detected late, it will result in death. Leukemia occurs when a lot of abnormal white blood cells produced by bone marrow. When abnormal white blood cells are a lot, the balance of the blood system will be disrupted. The existence of abnormal blood can be detected when the blood sample is taken and examined by hematologists. Microscopic images will be inspected visually by hematologists and the process is time consuming and tiring [3], [4], [5]. The process require human expert and prone to errors due to emotion disturbance and human physical capability that is of course have its own limit. Moreover, it is difficult to get consistent results from visual inspection [3]. Visual inspection also can only give qualitative results for further research [3]. Studies show that most of the recent techniques use all information about blood for e.g. number of red blood cells, hemoglobin level, hematocrit level, mean volume corpuscle and many more as the parameter for classifying diseases such as thalassaemia, cancer and etc. In order to know all information about blood, expensive testing and equipments of labs are required. Automatic image processing system is urgently needed and can overcome related constraints in visual inspection. The system to be developed will be based on microscopic images to recognize types of leukemia. The early and fast identification of the leukemia type greatly aids in providing the appropriate treatment for particular type of leukemia [30]. The currently used diagnostic methods rely on analyzing immunephenotyping, fluorescence in situ hybridization (FISH), cytogenetic analysis dan cytochemistry [18], [30]. Sophisticated and expensive laboratories are required in order to run the diagnostic methods and it have been reported to provide a high ratio of misidentification; as reported in [31]: “The best labs rely on as many as a dozen different, labor-intensive technologies, all of which require highly trained specialists. Even so, perhaps 50% of patients are misdiagnosed in regard to subtype” [30]. With this system, more images can be processed, reduce analyzing time, exclude the influence of subjective factors and increase the accuracy of identification process at the same time [7]. Inspection and classification of leukemia will be based on texture, shape, size, color, and statistical analysis of white blood cells.

This research is hoped can assist to increase efficiency globally and at the same time can benefit and be a huge contribution in medical and pattern recognition field. The main objective is to enhance algorithms that can extract data from human blood where human blood is the main source to detect diseases at earlier stage and can prevent it quickly [8]. This system should be robust towards diversity that exists among individual, sample collection protocols, time and etc [8]. It is hoped that this system can be automated in order to produce lab results quickly, easily and efficiently. The purpose of this paper is to review some
work done in blood cell recognition and to overview the proposed approach to be used in this research. In this paper, we will propose of using reinforcement learning (RL) in classifying types of leukemia. As in [9], reinforcement learning can overcome some of the problems in medical images. Medical images have very similar gray level and texture among the interested objects. Segmentation error may occur and increase. Another problem is may be lack of a sufficient number of training samples if a supervised learning technique is employed. By using RL approach, a minimum training dataset is required.

II. BACKGROUND
Blood is the main source of information that gives an indication of changes in health and development of specific diseases. Changes in the number or appearance of elements that formed will guide health condition of an individual.

2.1. Leukemia
Most blood cells produced from the cells in the bone marrow called stem cells. Bone marrow is a soft material found in the middle of each bone. Stem cells will mature and become some kind of blood cells. Each blood type has their own function. Blood components consist of:

a. Red blood cells (erythrocytes) - carry oxygen to tissues and back to the lungs with carbon dioxide.
b. White blood cells (leukocytes) - Defending the organism from infection. There are several types of white blood cells.
c. Platelets – helps blood clotting to control bleeding.
d. Plasma - The fluid in blood containing dissolved ions needed for cell function and consists of sodium, potassium, chloride, hydrogen, magnesium and iron.

When blood cells are old or damage, the cells will die and new cells will replace it [10].

Figure 1. Production Of Blood Cell [10]

Figure 1 shows that how stem cells became mature and evolve into several components of blood. They evolve into either myeloid stem cell or lymphoid stem cell. Myeloid stem cells eventually mature and became myeloid blast. This blast will form red blood cell, platelet and several types of white blood cell. Lymphoid stem cells also will mature and can form lymphoid blast and this blast will eventually form several types of white blood cells. White blood cells from myeloid blast are differs from lymphoid blast. The study will focus on leukemia because the disease is dangerous and can lead to death. For someone who has leukemia, bone marrow produces abnormal white blood cell. Compared with normal cells, abnormal white blood cell will not die when they should. Thus the number of abnormal white blood cell become numerous and interfere normal white blood cells to carry out their duties. This also causes an imbalance of blood system in the human body. Leukemia can be grouped based on how quickly this disease develops and become severe. Leukemia is either Chronic or Acute.

a. Chronic Leukemia – at earlier stage, leukemic cells can make tasks such as normal white blood cells. Gradually they will become severe chronic leukemia.
b. Acute leukemia - leukemia cells cannot make tasks like normal white blood cells. The number of leukemia cells will grow rapidly and become severe in a short time. Generally, leukemia can be divided into 4 types that are [11]:

a. Acute Lymphocytic Leukemia (ALL) – usually occurs in children aged 2-10 years. This type of leukemia is most common. It also always occur in adults.

Figure 2. Acute Lymphocytis Leukemia (All) [11]

b. Acute Myeloid Leukemia (AML) – This type of leukemia is common in children under the age of 1 year. It is extremely rare in teenagers. Even so it is mostly in adults aged 40 years.

Figure 3. Acute Myeloid Leukemia (Aml) [11]

c. Chronic Lymphocytic Leukemia (CLL) – This type of leukemia often happens to older patients. It is extremely rare in patients under the age of 40.

Figure 4. Chronic Lymphocytic Leukemia (Cml) [11]
d. Chronic myeloid leukemia (CML) – This type of leukemia can occur in all but the most common is for adults age after 45 years.

There are various aspects that make the process of recognition of blood cell images became very difficult task. Some of them are [12], [13].

- Types of leukocytes that covers a wide range of features for eg. shapes and colors in microscopic images.
- The use of different illumination for captioning image that lead to a variation of color distribution in the images
- Two neighboring cells or adjacent cells that are very similar to each other and the border point between two neighbors is not well defined
- Squashed leukocytes that appear as blurred image regions

2.2. Blood cell Research

Some research has been done in automating the process of blood cell identification and next can diagnose the patient correctly. Some of them are [14] who develop a system to identify and classify malaria parasite through microscopic images of blood cells. They use morphology approach and the major requirements in developing this system is the best techniques for segmentation of blood cell images.

[15] classify thalassaemia patients by using neural network and genetic programming. They did not use microscopic images but they use matured blood cell, platelet and reticulocyte information for eg cell percentage, hemoglobin level, hematocrit level, mean volume corpuscle, hemoglobin width distribution and etc to identify thalassaemia patients, thalassaemia traits and normal. The result acquired is using multi layer perceptron (MLP) with 2 hidden layers. Training data results have maximum accuracy 99.27%, mean accuracy 98.16% and standard deviation of 0.64%. While testing data results have maximum accuracy 88.72%, mean accuracy 84.44% and standard deviation 2.41%.

While [4] develop a system called Leuko and they use textural information to increase differences among leukocytes. They used textural parameters based on gray level occurrence matrices (GLCM) that are energy, inertia, homogeneity and correlation. Feature selection is very important in developing Leuko system that is data reduction can be done so that classification algorithms can learn quickly and accurately. Classifier can standardize better from available data, results are easier to understand as well as reduce the time.

[23] then built a system to recognize leukemia cell by using bone marrow images. The system was built by using Support Vector Machine (SVM) classifier and exploit features in blood cell images that related to texture, geometry and statistical analysis. They stress on generation and selection of features to get the best recognition. Textural parameters that have been used are mean value, angular second momentum, contrast and entropy. Geometrical parameters are radius, perimeter, area, filled area, compactness, concavity, concavity points and symmetry. While parameters for statistical analysis are mean value and standard deviation for nucleus and cytoplasm, mean and standard deviation for gradient matrix, skewness and kurtosis for image and gradient matrix. Error of training data is 11.87%, errors of testing data is 21.13%. They just select 30 best features and this produce error of training data to 13.07% and errors of testing data to 18.71%.

[6] use modified k nearest neighbour (KNN) to classify malaria parasite for microscopic images of blood cell and the results are so good with error 0.01%. [7] then, they use artificial neural network 3 layers and 4 layers to identify malaria and thalassaemia. They use microscopic images of blood cells and apply image processing technique for e.g. image enhancement, edge erosion, color, size normalizing and many more. They found out that artificial neural network with 3 layers give the best results that is with error 2.7454e(-0.005) and rate of correct classification is 86.54%.

[5] use EM-Algorithm to recognize types of leukocyte. First, the image pattern is changed into a lower dimensional space by using principal component analysis. EM-algorithm is used to get the parameters of the Gaussian functions to model the probability distribution function of each class of cells. The images are classified by getting the class conditional densities using Bayes’ theorem. Classification is done by choosing the class with the highest probability.

[18] have detected ALL by using fractal features that is hausdorff dimension to classify a lymphocytic cell into normal lymphocyte or lymphoblast. [18] also used fuzzy based segmentation technique to extract WBC nucleus from blood microscope images using color based clustering. At the end, they use SVM to classify. The accuracy of 93% was observed. [19] use global contrast stretching to enhance the images. By doing this, the visual aspect of blast cells can be increased and they do the segmentation based on HSI color space.

There are several applications in medical images that use reinforcement learning. [9] and [20] have used reinforcement learning (RL) in their work. They use RL in
order to overcome some problems in medical images. Medical images have very similar gray level and texture among the interested objects. Segmentation error may occur and increase. Another problem is may be lack of a sufficient number of training samples if a supervised learning technique is employed. By using RL approach, a minimum training dataset is required.

[9] use Q-learning to segment Computed Tomography (CT) images. They use cranial CT images. They found that they are able to segment simultaneously an image into some distinct regions. The images are divided into several sub images. Reinforcement learning (RL) agent will choose an action for sub images to change and updates the Q-matrix. There is an evaluator that compares the results and gives RL agent a reward. Segmentation accuracy that they achieve is above 95%.

[20] use RL approach to segment ultrasound images of prostate. They also use Q-learning and the results show high potential for applying reinforcement learning in medical segmentation. Their method is to control the local threshold and the post processing parameter by using a RL agent.

[21] extract kidney region as a preprocessing of kidney disease detection. They use abdominal X-Ray CT images. Q-learning is used within the rough kidney region and kidney contour edge is detected. However, there are a few error margins with an actual contour and it is corrected by snake method. The success probability is quite low that is 53%.

Another RL application is used by [22]. They use CT images of lung to classify lung nodules either benign or malignant. They use 3D geometric nodules characteristics to guide classification. The obtained results are very encouraging and show that the RL classifier can effectively classify the benign or malignant nodules based on CT images.

III. RESEARCH METHODOLOGY

From the literature, it is found that typical steps for the process of automating blood recognition are as in Figure 6.

1. **Image Acquisition**
   Blood image from slides will be obtained from nearby hospital with effective magnification.
2. **Preprocessing**
   During image acquisition and excessive staining, the images will be disturbed by noise. The noise may be due to illumination or shadows that make region of interest (ROI) appear as blurred image region. Background will be excluded since our ROI will be white blood cells. During this preprocess, image enhancement will be done as the contrast enhancement technique is capable to improve the medical image quality [19].
3. **Segmentation**
   Segmentation of white blood cell (WBC) and determine ROI that is nucleus for WBC only. This is because in leukemia cell images, the cytoplasm is scanty [18]. So, focus will be on nucleus of WBC only. Determination the types of WBC should be done from the nucleus. Only lymphocytes and myelocytes should be considered and need to determine them whether they are blast cells or not. Others like neutrophils, basophils and eosinophils should be excluded. Once the blast cells are determined, then proceed to the next step. Sub images containing nucleus only will be considered. This is to reduce errors since there are similar color scales in WBCs with other blood particles.
4. **Feature Extraction**
   The most important problem in generation of features of blood cells that characterize them in a way enabling the recognition of different blast types with the highest accuracy [23]. The features to be used are for nucleus of lymphocytes and myelocytes:
   - Geometrical Features – which includes area, radius, perimeter, symmetry, concavity, compactness, solidity, eccentricity, elongation, form factor will be obtained.
   - Texture Features – which includes homogeneity, energy, correlation, entropy [18], contrast, angular second moment will be obtained.
   - Color Features – the RGB color spaces will be transformed into HSV or L*a*b color spaces. Their mean color values will be obtained.
   - Statistical Features – the mean value, variance, skewness, kurtosis of the histograms of the image matrix and the gradient matrix for RGB or HSV or L*a*b color space (whichever appropriate) will be obtained.

Based on [19], ALL is small, blast cells are uniform, cytoplasm is scanty, round and usually contains single nucleoli inside nucleus. While in AML, the blasts are larger and irregular form and usually multiple nucleoli with the presence of Auer rode. [24] said that, the WBC appears rather darker than the background while red blood cell (RBC) appears in an intermediate intensity level. [25] indicates that white cells are the darker elements in images with RBC appear to be pale. Platelets are much smaller than white and red cells.
5. Classification

Classification is the task of assigning to the unknown test vector to a known class. In this step, a reinforcement learning algorithm is proposed. The RL approach will classify the types of leukemia into ALL, AML, CLL and CML. The basic model of RL is as in Figure 7 [26].

5.1 Supervised learning such as neural network or linear function when tabular representation is not feasible especially in continuous state space. Most application use Q-learning as this technique is the most popular in RL. Q-learning is a soft policy which means that its Q-values approximate the optimal Q-values that is Q(s,a), regardless of exploration. Q-values will be stored in Q-matrix. Q(s,a) is the expected sum of future payoffs, r, obtained by taking action, a, from state, s. The optimal action will be the one with the highest Q-value. Q(s,a) will be updated based on the experience as follows:

\[ Q(s,a) = (1-\alpha)Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a')] \]

where \( \alpha \) is the learning rate and \( 0 < \gamma < 1 \) is the discount factor. As a conclusion, the research methodology can be viewed as in Figure 8.

IV. DISCUSSION

Things that need to be discussed are to resolve some issues about the blood cells. One of the issues is the problem on the blood cell itself. [28] claim that their system fail in classification processes for some of the blood cells. Some of the cells can be deformed to arbitrary shape due to environment pressure. [3] takes note on their algorithms that does not separate overlapping cells. Overlapping cells may be joined cells which are caused by diseases. [13] said that the most difficult problem is the recognition between two neighboring cells in their development line since the cells are very similar and the border point between two neighbors is not well defined. But overlapping problem can be solved using watershed techniques used by [28] in their research.

Another issue is on collection of data. Images of blood samples should be sufficient enough so that generalization properties can be exhibited and be able correctly classify unseen data [29]. Lacking of samples means that only few principal components can be used to represent the data [5]. [6] claims that the ideal way is separated data sets should be used for each stage. However, sometimes cross validation or bootstrap sampling will be used since obtaining large number of samples is difficult. Approach of RL in classification step is hoped can minimize the problem of insufficient data.

All the issues present by researchers need to take into consideration when we are to build the system. We should try to overcome by applying appropriate techniques.
V. CONCLUSION

This research involves detecting the types of leukemia using microscopic blood sample images. The system will be built by using features in microscopic images by examining changes on texture, geometry, colors and statistical analysis as a classifier input. The system should be efficient, reliable, less processing time, smaller error, high accuracy, cheaper cost and must be robust towards varieties that exist in individual, sample collection protocols, time and etc. Information extracted from microscopic images of blood samples can benefit to people by predicting, solving and treating blood diseases immediately for a particular patient.

REFERENCES:


