Leukemia Detection from Blood Smears

Lokman Faivdullah, Farid Azahar, Zaw Zaw Htike, and Wei Yan Nyein Naing
Faculty of Engineering, IIUM, Kuala Lumpur, Malaysia
Email: zaw@ieee.org

***Abstract*—Leukemia is a highly fatal hematologic cancer which starts in blood-forming tissue, such as bone marrow and triggers high production of immature and abnormal shaped blood cells. Skilled medical officers are required to diagnose leukemia from blood smears. This poses a problem in remote areas and rural areas where there is a shortage of qualified medical personnel. This paper proposes a computer-aided diagnosis system that can detect and classify leukemia from blood microscopic images. Satisfactory preliminary experimental results demonstrate the efficacy of our system.**

***Index Terms*—leukemia detection, classification, scale-invariant feature transform, pattern recognition.**

I. Introduction

In this globalization era, there are several types of human diseases that have been threatening mankind. The impacts of these diseases especially cancers have severely affected the life expectancy of the human population. Leukemia is a series disease, which is among the deadliest diseases in the world. Leukemia can be difficult to diagnose. A computer-aided diagnosis system is crucial in rural areas where there is a shortage of qualified medical personnel. Computer-aided diagnosis can be achieved using computer vision and pattern recognition techniques that can recognize and diagnose leukemia from peripheral blood film images.

Leukemia is a highly fatal hematologic cancer which starts in blood-forming tissue, such as the bone marrow and trigger high production of immature and abnormal shaped blood cells [1]. Normally, bone marrow produce three distinct types of cells which are erythrocytes, leucocytes and thrombocytes. Usually, the marrow produce the precise number of cells in accordance with the need of the body. However, in leukemic patients the process becomes violated and start producing abnormal premature white blood cells (called blast) with abnormal shapes that cannot function normally and in uncontrolled manner [2], [3]. As these blasts upsurge rapidly, they interrupt the production of red blood cells and platelets in the bone marrow. Leukemia can be very broadly divided into four classes: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML).

ALL is hematological neoplasia that occurs commonly in young children, characterized by over production of abnormal leukemic cells in lymphoid organ (bone marrow) to form lymphocytes. AML is hematological malignancy which is characterized by excessive proliferation of myeloid cells in the bone marrow. This type is common among adults, especially in men [4]. The other type classes of leukemia are similar to the above two classes. The only difference is that 'acute' refers to rapid proliferation of immature blood cells and 'chronic' generally refers to proliferation that is less rapid than the 'acute' level. A patient suffering from chronic leukemia generally gets worse gradually. To diagnose leukemia, the number and the shapes of blood cells have to be investigated from microscopic blood film images.

II. Leukemia Detection Techniques

There are generally two famous techniques to diagnose leukemia: fluorescence in situ hybridization technique and flow cytometry technique.

A. FISH Technique

Fluorescence in situ hybridization [5] is a common method to detect DNA sequences by using fluorescent probes. This technique allows the detection, analysis and calculation of abnormalities of genetic structure [6], [7]. FISH technique consists of 3 main steps which are sample preparation, hybridization and fluorescence microscopy [8]-[10]. Fig. 1 illustrates the basic steps of the FISH technique.

1) Sample preparation

The first equipment used is multi-probes tagged with a different combination of fluorophore, allows simultaneous visualization of different molecular component of the cell. The second is Multi-band fluorescence microcopy, enhance in discriminating the labeled color specimens [11].

2) Hybridization

Initially, denaturation of target and sequential probes is carried out by submersing them for 5 minutes in denaturing solution which consist of chemicals such as 49ml (70%) of form amide and 7ml (30%) 2X Saline-Sodium Citrate (SSC) with pH value 5.3. This method helps them to form new hydrogen bonds to facilitate the hybridizations procedures [12]. The probes mixture is made by mixing up together the probe and target sequences. Consequently, the probe is clearly hybridized with the chromosome to form its better balancing sequences. The probes mixture is then covered by the coverslip on the slides after that excess liquid mixture is removed to prevent contamination [13].

Initially, the cover glasses have to be sealed by rubber cement using suitable syringes and slides after that put it

Manuscript received July 10, 2014; revised December 10, 2014.

doi: 10.12720/jomb.4.6.488-491

into a pre-warmed humidified chamber for about 14-17 hours at 37 ˚C which is the optimal temperature of the incubator. Next, the seal of coverslips are detached gently with forceps and slides washed in post hybridization wash buffer solution (0.4 x SSC/0.3% NP-40) at 27 ˚C for 1 minute. When the slides are completely dried, 10 µL of DAPI II counterstain is applied to target area and sealed under a cover glass. These slides are required to store in the dark condition for signal record.

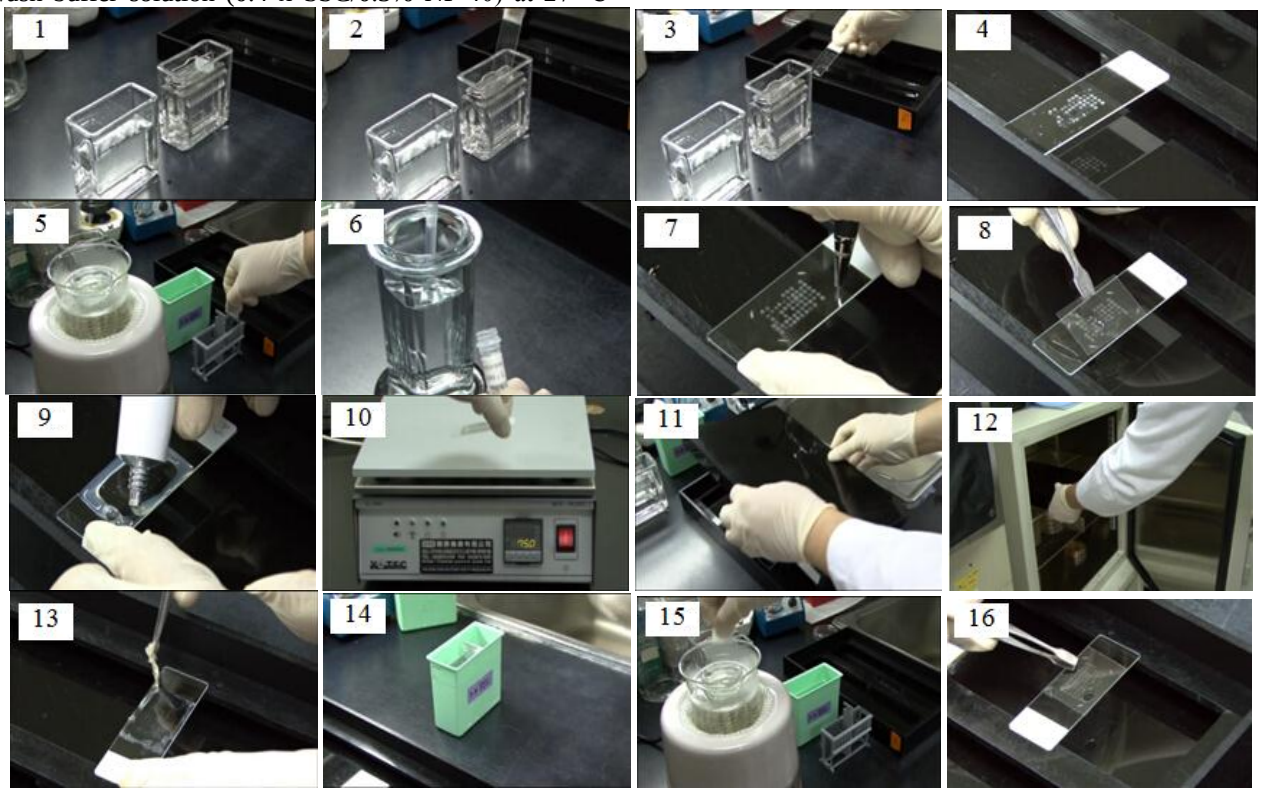

Figure 1. Basic steps of FISH technique. [5]

3) Fluorescence microscopy

Slides are then viewed under a fluorescent microscope. Signals were screened using image processing software and photo detector software. Slides are evaluated by certain criteria and condition. The room should be dark and free of fluorescence elements and the signal must be bright and compact in size. Then, while scanning the probe, it is recommended to start in the top left quadrant of the coverslip and move specifically from up to down and counting properly the number of signals detect in the boundary. By using a 400X objective, tumor images were detected. Then, the images took from all experimental slides, stored in suitable image format such as TIFF format, and recorded in a proper table for analysis.

B. Flow Cytometry

Flow cytometry as depicted in Fig. 2 refers to the way used to analyze multiple physical and chemical characteristics of particles, normally cells, conjoined with different array of fluorochromes. Using flow cytometry, the mutated cell characteristics can be diagnosed by establishing the heredity such as cytotoxic T lymphocytes (CTLs) family. This method also used to form the dual heredity in special phenotypic leukemia, uncommon co-expression of antigen or abnormal outline and exhibit similar genes [14]. Flow cytometry generally entails three steps: specimen preparation, pattern recognition, and data interpretation.

1) Specimen preparation

Two ml of cell (blood cells) is mixed in ethylene-diamine tetra-acetic acid and heparin for clotting prevention. The prevention is done by properly shaking the sample from 5 to 10 times before proceeding to the next step. The slides were tarnished in Jeiner-Wright solution for every sample. If required, cells feasibility can be tested also by using dye exclusion method which contains trypan blue. Subsequently, the cells were processed in no-cell loss method. This method lessens the breakdown of cells in the specimens. This blood lyse and wash method is the basis for red blood cells lysis. Customary cell number suggested for immune-staining is 10 million cells for all samples necessary for antibody mixture [15].

2) Pattern recognition

Each peculiar monoclonal antibody, antigen intensity expressions and percentage of cells which gives constructive results are very important for justification. Although, gating by CD45/SSC is very precise, but it is much more convenient to use the conventional side and onward scatter SSC/FSC gating. Also, sequential and back gating can be used as an optional choice. For a situation of AML M0 and M1, blasts population with FSC and SSC had established modest intensity of expression for CD45; This CD45 and SSC produces discrete blasts number merging with growing granulocytic elements for the circumstances of myeloid

leukemia with granular variation. In the case of monocytic leukemia, intersections in the region of the normal monocytes and the blasts are detected [16]. Lymphoblast displayed damaging CD45 with low SSC in case of acute lymphoblastic leukemia [17].

3) Data interpretation

The blasts' ratio expressing each antigen tested usually considered as positive whenever any markers shows blast presence is equal or more than 20%. However, 20% as the limit point is subjective and changeable. Alternative technique of examining is by calculating antigen expression intensity i.e. measured fluorochrome antibody conjugate binding capacity. There are four different ranges that show either poor(+), moderate (++) and strong (+++) which starts with first log value between 100 and 101 for the weak range upto the strong range of fluorescence intensity is consecutively for second, third and fourth decades. Occasionally, because of the diverse discoloration intensity, bimodal may appear in the malicious group of cells [13].

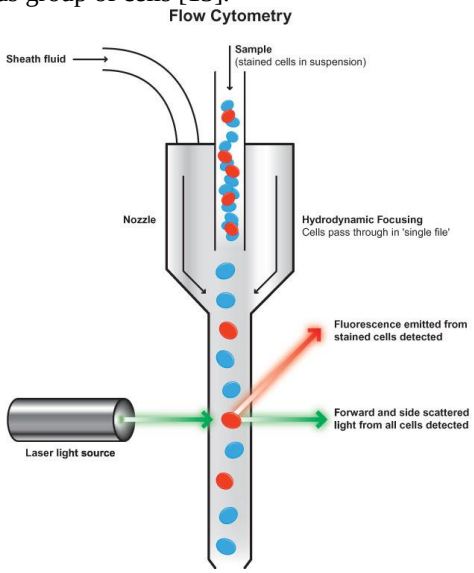

Figure 2. Flow cytometry [18].

III. Methodology

The goal of leukemia detection is to predict, given a blood smear image, whether the sample comes from a patient with leukemia and what type of leukemia the patient is suffering from. We proposed a three-layered framework that consists of feature extraction, coding, and classification as shown in Fig. 3.

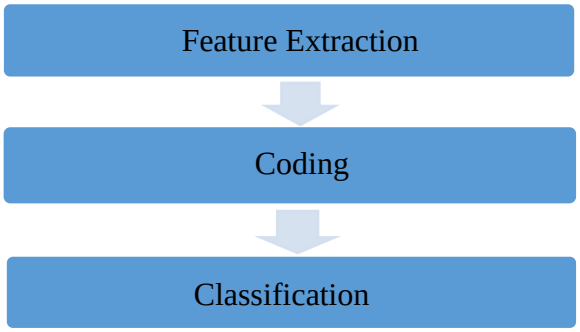

Figure 3. Block diagram of proposed leukemia detection system.

A. Feature Extraction

We employ dense Scale-Invariant Feature Transform (DSIFT) as features because it is efficient and invariant to illumination, scale, shift, and rotation. We set the spacing of the dense grid to 6 pixels, and the sampling window to 16×16 pixels. We employ 16 histograms to allow preservation of geometric information of images.

B. Coding

The complexity of any machine learning classifier depends upon the dimensionality of the input data [19]-[27]. To lower the complexity, coding processing is carried out on the DSIFT descriptor to reduce dimensionality. For a given dataset, we construct a codebook of 1500 vectors using k-means clustering [9]. Each cluster center corresponds to a "word" in codebook. After coding, spatial pyramid matching techniques is then employed to preserve the spatial relationship between local descriptors. The basic idea is to divide the image into a few rectangle blocks and construct the histogram in the block. We employed 3 layers of spatial pyramid [28].

C. Classification

Coded feature vector is then fed to a support vector machine classifier (SVM) [29] to classify leukemia types. Since there are multiple types of leukemia and an SVM is essentially a binary classifier, we employ a one-vs-all technique to make it a multi-class classifier [30].

IV. Experimental Results

We collected 100 blood images corresponding to each of the four leukemia classes (acute lymphoblastic leukemia, acute myeloid leukemia, chronic lymphocytic leukemia, and chronic myeloid leukemia) from the Internet. For each leukemia class, 60 images were used for training while the remaining 40 images were used for testing. We obtained an overall accuracy of 79.38% in classifying leukemia from blood images.

V. Conclusion

We have presented a machine vision based approach to detect leukemia from blood microscopic images. Given a blood microscopic image, the system predicts whether the blood microscopic image comes from a patient with leukemia and the type of leukemia the patient has. We have performed experiments on a dataset of leukemia blood microscopic images. The experimental results are quite promising. As future work, we would like to perform optimization of the system parameters to further boost the performance of the system. "

References

[1] P. Younger, "Human diseases and conditions," *Reference Reviews*, vol. 24, pp. 40-41, 2010.
[2] M. Greaves and F. Alexander, "An infectious etiology for common acute lymphoblastic leukemia in childhood?" *Leukemia*, vol. 7, pp. 349-360, 1993.
[3] A. Quint ás-Cardama and J. E. Cortes, "Chronic myeloid leukemia: Diagnosis and treatment," in *Mayo Clinic Proceedings*, 2006, pp. 973-988.

[4] Y. O. Huh and S. Ibrahim, "Immunophenotypes in adult acute lymphocytic leukemia: Role of flow cytometry in diagnosis and monitoring of disease," *Hematology/oncology Clinics of North America*, vol. 14, pp. 1251-1265, 2000.
[5] C. O'Connor, "Fluorescence in situ hybridization (FISH)," *Nature Education*, vol. 1, pp. 171, 2008.
[6] B. Lerner, *et al.*, "Segmentation and classification of dot and non-dot-like fluorescence in situ hybridization signals for automated detection of cytogenetic abnormalities," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, pp. 443-449, 2007.
[7] Y. L. Wang, *et al.*, "Fluorescence microscopy of living cells in culture Part A. fluorescent analogs, labeling cells, and basic microscopy," in *Methods in Cell Biology*, D. L. T. Yu-Li Wang and K. W. Jeon, Eds., ed: Academic Press, vol. 29, 1988, p. 3.
[8] K. R. Castleman and B. S. White, "Dot count proportion estimation in FISH specimens," *Bioimaging*, vol. 3, pp. 88-93, 1995.
[9] T. Kanungo, *et al.*, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881-892, 2002.
[10] C. O. de Solórzano, *et al.*, "Automated FISH spot counting in interphase nuclei: Statistical validation and data correction," *Cytometry*, vol. 31, pp. 93-99, 1998.
[11] K. R. Castleman, T. P. Riopka, and Q. Wu, "FISH image analysis," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 15, pp. 67-75, 1996.
[12] Fluorescence in Situ Hybridization. (2013). [Online]. Available: www.abnova.com
[13] J. Nath and K. L. Johnson, "A review of fluorescence in situ hybridization (FISH): Current status and future prospects," *Biotechnic & Histochemistry*, vol. 75, pp. 54-78, 2000.
[14] D. Pinkel, *et al.*, "Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4," *Proceedings of the National Academy of Sciences*, vol. 85, pp. 9138-9142, 1988.
[15] R. Saxena and H. Anand, "Flow cytometry in acute leukemia," *Indian Journal of Hematology and Blood Transfusion*, vol. 24, pp. 146-150, 2008.
[16] D. T. Nguyen, *et al.*, *Flow Cytometry in Hematopathology: A Visual Approach to Data Analysis and Interpretation*, Springer, 2008.
[17] E. G. Weir and M. J. Borowitz, "Flow cytometry in the diagnosis of acute leukemia," in *Seminars in hematology*, 2001, pp. 124-138.
[18] A. B. Lyons and C. R. Parish, "Determination of lymphocyte division by flow cytometry," *Journal of Immunological Methods*, vol. 171, pp. 131-137, 1994.
[19] Pathology Student: A Shortlist of CD Marker. (2013). [Online]. Available: http://www.pathologystudent.com/?p=7963
[20] E. E. M. Azhari, *et al.*, "Brain tumor detection and localization in magnetic resonance imaging," *International Journal of Information Technology Convergence and Services*, vol. 4, 2014.
[21] E. E. M. Azhari, *et al.*, "Tumor detection in medical imaging: A survey," *International Journal of Advanced Information Technology*, vol. 4, 2014.
[22] Z. Z. Htike and S. L. Win, "Recognition of promoters in DNA sequences using weightily averaged one-dependence estimators," *Procedia Computer Science*, vol. 23, pp. 60-67, 2013.
[23] Z. Z. Htike and S. L. Win, "Classification of eukaryotic splice-junction genetic sequences using averaged one-dependence estimators with subsumption resolution," *Procedia Computer Science*, vol. 23, pp. 36-43, 2013.
[24] N. A. Mohamad, *et al.*, "Bacteria identification from microscopic morphology using naïve bayes," *International Journal of Computer Science, Engineering and Information Technology*, vol. 4, 2014.
[25] N. A. Mohamad, *et al.*, "Bacteria identification from microscopic morphology: A survey," *International Journal on Soft Computing, Artificial Intelligence and Applications*, vol. 3, 2014.
[26] S. L. Win, *et al.*, "Cancer recurrence prediction using machine learning," *International Journal of Computational Science and Information Technology*, vol. 6, 2014.
[27] S. L. Win, *et al.*, "Gene expression mining for predicting survivability of patients in early stages of lung cancer," *International Journal on Bioinformatics & Biosciences*, vol. 4, 2014.
[28] S. L. Win, *et al.*, "Cancer classification from DNA microarray gene expression data using averaged one-dependence estimators," *International Journal on Cybernetics & Informatics*, vol. 3, 2014.
[29] Y. Xiong, "Image classification," *CSCI 5280 Image Processing and Computer Vision*, 2014.
[30] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

