Volume 5, Issue 5, May 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Cancer Detection Technique using Image Processing: A Review

Mini Puri, Jyoti Mann
South Point Institute of Technology & Management
Haryana, India

Abstract— In this paper, two steps were used to improve the appearance of the acquired bone marrow slide images. These techniques include create the dataset using the cancerous images and make decision for images (whether cancerous or not). Use of both techniques together produced good results. Several features were extracted to test whether an image is cancerous or not. These features include: mean, standard deviation and variance for red channel, green channel and blue channel separately. Then, the test image is also processed for these features. Finally based upon the calculated features of the dataset and that of the test image, we make our decision if the given image is cancerous or not.

Keywords— Cancer Detection, Dataset, Image Processing, Cancerous Images.

I. INTRODUCTION

Cancer known medically as malignant neoplasia, is a broad group of diseases involving unregulated cell growth. In cancer, cells divide and grow uncontrollably, forming malignant tumors, which may invade nearby parts of the body. The cancer may also spread to more distant parts of the body through the lymphatic system or bloodstream. Not all tumors are cancerous; benign tumors do not invade neighboring tissues and do not spread throughout the body. There are over 200 different known cancers that affect humans. [1]

The causes of cancer are diverse, complex, and only partially understood. Many things are known to increase the risk of cancer, including tobacco use, dietary factors, certain infections, exposure to radiation, lack of physical activity, obesity, and environmental pollutants, these factors can directly damage genes or combine with existing genetic faults within cells to cause cancerous mutations. Approximately 5–10% of cancers can be traced directly to inherited genetic defects. Many cancers could be prevented by not smoking, eating more vegetables, fruits and whole grains, eating less meat and refined carbohydrates, maintaining a healthy weight, exercising, minimizing sunlight exposure, and being vaccinated against some infectious diseases.[1]

Cancers are classified by the type of cell that the tumor cells resemble and are therefore presumed to be the origin of the tumor. These types include:

- **Carcinoma**: Cancers derived from epithelial cells. This group includes many of the most common cancers, particularly in the aged, and includes nearly all those developing in the breast, prostate, lung, pancreas, and colon.
- **Sarcoma**: Cancers arising from connective tissue (i.e. bone, cartilage, fat, nerve), each of which develops from cells originating in mesenchymal cells outside the bone marrow.
- **Lymphoma and leukemia**: These two classes of cancer arise from hematopoietic (blood-forming) cells that leave the marrow and tend to mature in the lymph nodes and blood, respectively. Leukemia is the most common type of cancer in children accounting for about 30%.
- **Germ cell tumor**: Cancers derived from pluripotent cells, most often presenting in the testicle or the ovary (seminoma and dysgerminoma, respectively).
- **Blastoma**: Cancers derived from immature "precursor" cells or embryonic tissue. Blastomas are more common in children than in older adults.[2]

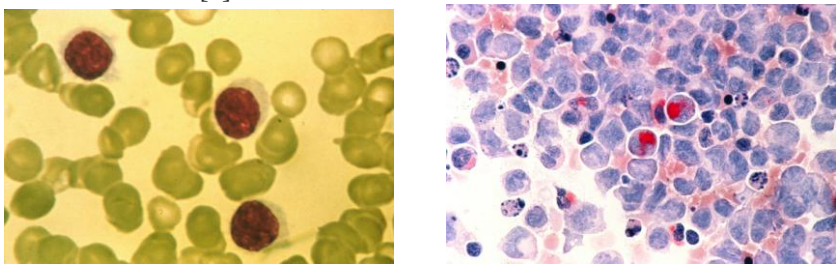

Figure 1: Cancerous Images

II. LITERATURE SURVEY

A. Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images

The early identification of acute lymphoblastic leukemia symptoms in patients can greatly increase the probability of recovery. Nowadays the leukemia disease can be identified by automatic specific tests such as Cytogenetics and Immunophenotyping and morphological cell classification made by experienced operators observing blood/marrow microscope images. Those methods are not included into large screening programs and are applied only when typical sysssfmptoms appears in normal blood analysis. The Cytogenetics and Immunophenotyping diagnostic methods are currently preferred for their great accuracy with respect to the method of blood cell observation which presents undesirable drawbacks: slowness and it presents a not standardized accuracy since it depends on the operator's capabilities and tiredness. Conversely, the morphological analysis just requires an image -not a blood sample- and hence is suitable for low-cost and remote diagnostic systems. The presented paper shows the effectiveness of an automatic morphological method to identify the Acute Lymphocytic Leukemia by peripheral blood microscope images. The proposed system firstly individuates in the blood image the leucocytes from the others blood cells, then it select the lymphocyte cells (the ones interested by acute leukemia), it evaluates morphological indexes from those cells and finally it classifies the presence of the leukemia.[3]

B. Cell Stem Cell Regional localization within the bone marrow influences the functional capacity of human HSCs

Numerous studies have shown that the bone marrow (BM) niche plays a key role in mouse hematopoietic stem cell (HSC) function and involves contributions from a broad array of cell types. However, the composition and role of the human BM HSC niche have not been investigated. Here, using human bone biopsies, we provide evidence of HSC propensity to localize to endosteal regions of the Trabecular Bone Area (TBA). Through functional xenograft transplantation, we found that human HSCs localizing to the TBA have superior regenerative and self-renewal capacity and are molecularly distinct to those localizing to the Long Bone Area (LBA). In addition, osteoblasts in the TBA possess unique characteristics and express a key network of factors that regulate TBA vs. LBA localized human HSCs in vivo. Our study reveals that BM localization and architecture play a critical role in defining the functional and molecular properties of human HSCs.[4]

C. Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification

The complete blood count (CBC) is widely used test for counting and categorizing various peripheral particles in the blood. The main goal of the paper is to count and classify white blood cells (leukocytes) in microscopic images into five major categories using features such as shape, intensity and texture features. The first critical step of counting and classification procedure involves segmentation of individual cells in cytological images of thin blood smears. The quality of segmentation has significant impact on the cell type identification, but poor quality, noise, and/or low resolution images make segmentation less reliable. We analyze the performance of our system for three different sets of features and we determine that the best performance is achieved by wavelet features using the Dual-Tree Complex Wavelet Transform (DT-CWT) which is based on multi-resolution characteristics of the image. These features are combined with the Support Vector Machine (SVM) which classifies white blood cells into their five primary types. This approach was validated with experiments conducted on digital normal blood smear images with low resolution.[5]

D. Automated classification in digital images of osteogenic differentiated stem cells

The study of stem cells has received considerable attention in forming many different tissue types, and gives hope to many patients as it provides great potential for discovering treatments and cures to many diseases such as Parkinson's disease, schizophrenia, Alzheimer's disease, cancer, spinal cord injuries and diabetes. This study was concerned with developing algorithms that analyses microscope images of stem cells harvested from the bone marrow or dental pulp of a rabbit, expanded in the laboratory at the Tissue Engineering Center in Alexandria, Egypt, and then transplanted into subcutaneous pouches of the rabbit. The research aimed to detect automatically as soon as osteogenic differentiated stem cells were ready to be implanted in the defective parts, thereby avoiding the cells becoming damaged by bacterial infection. A further requirement was that the algorithms would not use traditional (chemical) markers which eventually lead to the sample being discarded as it dies after adding the marker. A total of 36 microscopy images were obtained from seven separate experiments each lasting over 10 days, and the clinicians visually classified 18 images as showing not-ready osteogenic differentiated stem cells and the remaining images showing a variety of cells ready for implantation. The ready cells typically appeared as a colony, or spread all over the image interconnecting together to form a layer. Initially, image pre-processing and feature extraction techniques were applied to the images in order to try and identify the developing cells, and a t-test was applied to the total cell area in each image in an attempt to separate the not-ready and ready images. While there was a significant difference between not-ready images and the ready images which showed the colony shaped characteristics, there was no significant difference between not-ready images and ready images with the spreading interconnecting layer shape, and so more sophisticated classification techniques were investigated. As the differentiated stem cells are effectively texture based images, each of the 36 images were divided into quadrants to give a total of 144 images to increase the image dataset. Several sets of texture parameters were derived from the grey-scale histogram statistics, Grey-Level Co occurrence Matrix (GLCM), and Discrete Cosine Transform (DCT) spatial frequency components of the images. Some of these parameters were used with traditional classification techniques including cross-correlation, and Euclidean distance measures to try and classify the texture

relative to the first image (not-ready) in each experiment and the other images (not-ready and ready) in the experiment. The success rate using cross-correlation was 70%, and 68% for the Euclidean distance approach. Secondly, intelligent classification techniques using Artificial Neural Networks (ANN) were considered, using the various texture parameters as inputs to a feed-forward 1-hidden layer MLP using Back-propagation of Errors for training. The ANN approach gave the better results, with 77% using the grey-scale histogram statistics, 73% for GLCM, and 92% for the DCT with 70 spatial frequency components. It was observed for each of the experiments that images became classified as ready for implantation after approximately 10 days, and then remained ready for the rest of the experiment.[6]

III. METHODOLOGY

The whole implementation is broadly divided into 2 parts:

1. Create the dataset using the cancerous images
2. Make decision for images (whether cancerous or not)

In the first part, various features are extracted from the sample images. In this project, a total of 35 sample images are taken as input. This sample size can be changed. More number of sample images tends to increase the accuracy of the decision. Also, after extraction of features, range is set for each feature. This range will be used in step 2 for classification.

In step 2, range from step one is taken into account. The test image is inputted (for which the decision is to be made). Features are extracted for this image. These features are then tested with the range of features as calculated from step 1. This is how the decision is made whether an image is cancerous or not.

IV. FUTURE SCOPE

The first step in any future study will involve obtaining a larger image dataset. Also, neural network classifiers can be used for decision making. The larger dataset will enable the recommended ANN classifier to be investigated more thoroughly and its performance reassessed. Alternative classifiers could also be investigated including the *k*-Nearest Neighbors Classifier and the Bayesian Classifier as these have also been successful in classifying biomedical data, and time did not permit an evaluation of the classifiers on this image dataset in this study

V. CONCLUSION

This research involves detecting leukemia (a form of cancer) using microscopic blood sample images. The present system uses microscopic images and extracts its features such as changes on texture, geometry, colors and statistical analysis and then uses them as an input to the classifier. The image processing technique that is employed here has been able to understand the infected cells present in Red Blood Cells (RBC) in case of the sick/infected patient. 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. The use of such microscopic images of blood samples can be used to provide information about diseased patient more quickly. The image processing techniques used has helped us to better understand the sickle-cells present in Red Blood Cells (RBCs) in case of sickle-cell patient.

REFERENCES

[1] http://en.wikipedia.org/wiki/Cancer

[2] Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M (2005). "Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors".*Lancet* **366** (9499): 1784–93. doi:10.1016/S0140-6736(05)67725-2. PMID 16298215

[3] F Scotti (2005). " Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images". IEEE International Conference on Computational Intelligence for Measurement Systems and Applications.pp-96-101.

[4] Mickie Bhatia. " Cell Stem Cell Regional Localized Within The Bone Marrow Influences the functional Capacity Of Human HSCs. " McMaster Stem Cell and Cancer Research Institute Hamilton, CANADA." CELL-STEM-CELL-D-12-00621R3".

[5] M. Habibzadeh, A. Krzyżak, T. Fevens, "Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification", Journal of Theoretical and Applied Computer Science, vol. 7, no. 1, pp. 20-35, 2013.

[6] Abdelgawadbirry, R 2013, Automated classification in digital images of osteogenic differentiated stem cells , PhD thesis, University of Salford, Manchester.

