

Disease Detection using Blood Smear Analysis

Pragati Sharma
Dept. of CSE
MSIT, GGSIPU
New Delhi

ABSTRACT

Blood smear analysis is an important diagnostic test which is performed to diagnose an array of diseases. The count of various blood cells and their morphological properties are the main focus of this test. Manual analysis of blood smears is time consuming and laborious. By automating this process and ultimately narrowing the scope of possible diseases, a considerable amount of time can be saved. This may in turn help medical staff as well as the patients. In this paper an automated technique for blood smear analysis using image processing is proposed to discern the blood cell count and blood cell properties. The results of image processing are then employed to generate a neuro-fuzzy system capable of predicting possible diseases.

Keywords

Blood Smear, Image Processing, Neuro-Fuzzy System

1. INTRODUCTION

Apart from these another important blood entity is platelet. A blood smear is essentially is a thin film of blood which is spread on a glass slide and is studied by pathologist under a microscope. It is usually done in combination with CBC (complete blood count). Human blood consists of two types of cells, Red blood corpuscles (RBC) and white blood corpuscles (WBC). Red blood cells or erythrocytes are the cells which are present in abundance in our blood. They are the reasons why our blood is red. The primary task of RBC's is delivering oxygen to various parts of our body.

Abnormality in their number and structure is indicative of various types of anaemia and microangiopathic diseases. WBC's or leukocytes are cells which constitute the immune system of the body. They fight back infections and foreign bodies. Any abnormality in their count and appearance could be due to immunodeficiency diseases, leukaemia etc. or simply an infection or allergic reaction. Platelets or thrombocytes despite not being proper cells play an important role in our body. They are responsible for clotting blood at sites of injury, both external and internal. An abnormal amount of platelets is indicative of blood clotting disorders, dengue, etc.

Manual analysis of blood smear involves examining blood smear slides under the microscope. Examination focusses on various aspects some of which are the number of various types of cells, their shape, size, if they have some foreign bodies or not, their colour is right or not etc. This makes this technique prone to human error. This process can be automated by using image processing.

Image Processing is the technology of manipulating the groups of bits (or pixels) of the image to enhance the quality of the image or create different perspectives or to extract information from the image digitally, with the help of computer algorithms. In this input is an image, a series of images, or a video, such as a photograph; the output may be either an image or some parameter of an image. By applying

various operations like grey scale conversion, thresholding, boundary analysis, image segmentation etc. useful information from blood smear images can be obtained. Our aim will be to harness the image processing capacity of MATLAB in order to obtain the necessary information required for disease detection.

Neuro fuzzy system is a system that uses learned algorithm from neural networks to predict various parameters of fuzzy system. The human like reasoning style of fuzzy systems is incorporated by NFS (neuro fuzzy system). The architecture of the system is feed forward layered neural network having input variables, fuzzy rules and output variables. The results obtained from image processing would then be used to predict possible diseases.

2. RELATED WORK

Analysis of microscopic medical images is an important task involving both physicians and computer scientists. One of the essential and active areas of research is the problem of counting blood cells (CBC) which is used as screening test to check these disorders as infections, allergies, problem with clotting, and it helps diagnosing and managing a large number of diseases.[1]

Essentially an efficient method using the image processing technique (image enhancement, segmentation and feature extraction) has been constructed method isolates and determines the type of all red blood cells in the smear which could be either normal or abnormal. Taking into consideration that abnormal red blood cells indicate to the associated blood anaemia. Focus was directed to the process of classifying types of abnormal cells. [2]

Leukocytes are categorized into five groups in a peripheral blood microscopic image. A differential count of these various types of cells is used to determine the presence of an infection in the human body. Precise boundary of leukocyte and cytoplasm are necessary for extracting features of them. In fact texture, colour, size and morphology of nucleus and cytoplasm make differences among different kinds of leukocytes. [3]

The counting and classification of blood cells allows the evaluation and diagnosis of a vast number of diseases. Through the analysis of white blood cells (WBCs) the ALL - Acute Lymphocytic Leukaemia, a blood cancer that can be fatal if left untreated, can be detected.

Nowadays the morphological analysis of blood cells is performed manually by skilled operators. This involves numerous drawbacks, such as slowness of the analysis and a non-standard accuracy, dependent on the operator skills. In literature there are only few examples of automated systems in order to analyse and classify the blood cells, most of which only partial. [4]

There has been an exponential intensification in the classification of medical images by different classifiers and algorithm. Different algorithms have been used to classify the medical images. Effective medical image can play an important role in aiding in diagnosis for healthcare students by explaining with this image will help them in their studies as well. Data mining is the process of discovering meaningful new correlations. Pattern and trends by shifting through large amount of data stored in data base. [5]

The measure of WBC and RBC Cells are very important for the doctor to diagnose various diseases such as anaemia, leukaemia etc. So, precise counting of blood cells plays very important role. The old conventional method used in hospital laboratories involves manual counting of blood cells using a device called Haemocytometer. But this process is extremely monotonous, time consuming, and leads to inaccurate results. Even though hardware solutions such as the Automated Haematology Counter exists, they are very expensive machines and unaffordable in every hospital laboratory. [6]

3. METHODOLOGY

The steps followed during the process were as follows:

3.1 Image Pre-processing

Image Processing is the technology of manipulating the groups of bits (or pixels) of the image to enhance the quality of the image or create different perspectives or to extract information from the image digitally, with the help of computer algorithms. In this input is an image, a series of images, or a video, such as a photograph; the output of image processing may be either an image or some parameters of the image.

3.1.1 Greyscale Conversion

The first step of pre-processing the image is converting the normal image (RGB) into greyscale. This step is necessary because further image processes can be done only on binary images and there is no way to convert a coloured image to binary images, thus it is necessary to convert them into greyscale. An in-built method of MATLAB was used in order to convert the normal image to greyscale. As shown in figure 1

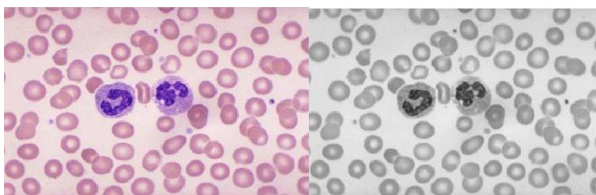


Figure 1: Original image (L) greyscale image

3.1.2 Noise Removal and Image Conversion to Binary

Next step in pre-processing of image is removal of noise from the image because images are prone to a variety of types of noise. Noise removal is necessary because it results in pixel values of the image that do not reflect the true intensities of the real scene.

Image thresholding is a simple, yet effective, way of partitioning an image into a foreground and background. A Histogram shape-based method of thresholding was used, where, for example, the peaks, valleys and curvatures of the smoothed histogram are analysed. After removing all the noise from the image our next step is to convert our image to binary. Converting to binary is used in order to find a Region of Interest, a portion of the image that is of interest for further

processing. The intention is binary, "Yes, this pixel is of interest" or "No, this pixel is not of interest". The results are shown in figure 2.

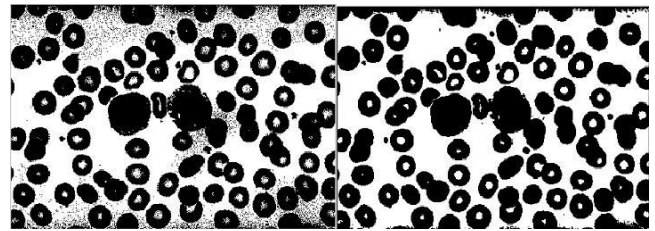


Figure 2: Binary image before noise removal (L) and after noise removal (R)

3.2 Counting Of Cells

In order to obtain the count of the various types of cells, the pre-processed fill was used in the holes for each cell. This is done because boundaries of the cells are used in order to count the number of entities in the image. Filling holes will prevent count of extra non- existent boundaries. After this, distinct boundaries of the objects were identified and systematically cell components were removed from the image based on their size. Hence, the count of WBC's, RBC's and platelets were obtained as present in the image. This process is shown in figure 3.

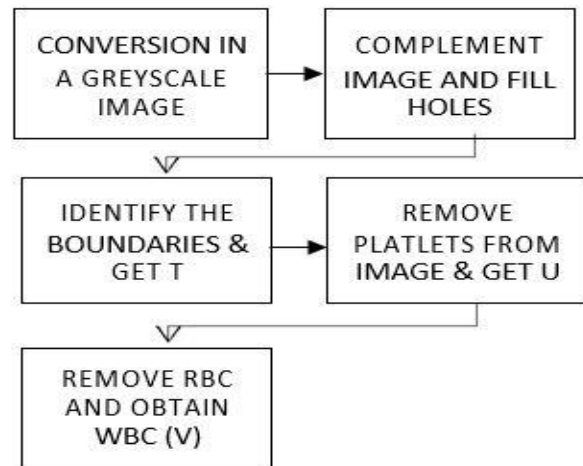


Figure 3: Flowchart for counting

$$T = RBC + WBC + PLATELET COUNT$$

$$U = RBC + WBC COUNT$$

$$NO. OF WBC'S = V$$

$$NO. OF RBC'S = U - V$$

$$NO. OF PLATELETS = T - (U + V)$$

3.3 RBC Feature Extraction

Starting with the pre-processing of image and converting it into a noise free greyscale image, then complement the image and remove small holes and identify individual objects in an image. After this remove platelets and WBC's. Then remove overlapping of cells and incomplete cells at the boundary. Identify abnormally sized and shaped RBC's.

This is demonstrated in figure 4.

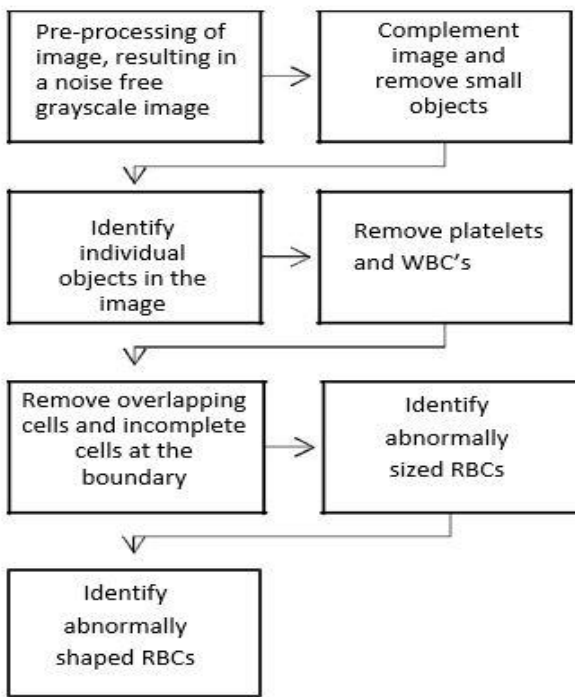


Figure 4: Flowchart for RBC feature detection

- An FIS system is generated with appropriate set of inference rules for mapping the inputs (Blood count, sickle cell's presence, types of WBCs) with the Output (Disease).

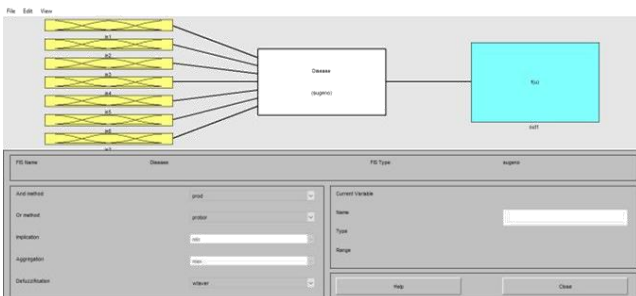


Figure 5: The generated FIS after sub-clustering

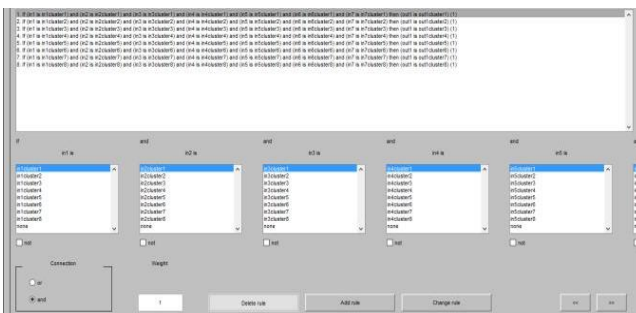


Figure 6: If-then Rules of FIS

- After, training the system is validated with the testing data.

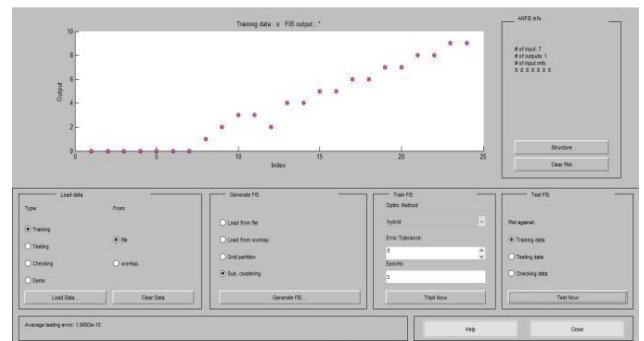


Figure 7: Testing vs Training data plot

3.4 ANFIS

NeuroFuzzy approach as a cumulative one is precise and effective in calculating the Software Quality Prediction. In trained and tested in MATLAB:

- Load the Training data in neuro-fuzzy system using MATLAB toolbox.
- The training of the data is done with the help of the Sub-clustering of datasets and application of hybrid algorithm.

The Overall process followed is depicted in figure 9.

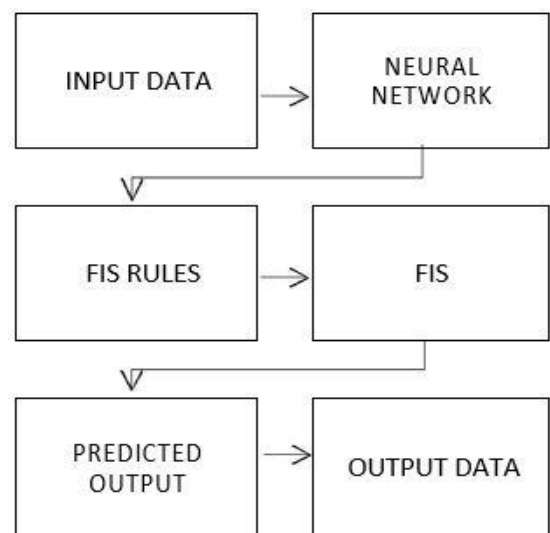


Figure 8: Flowchart of ANFIS

4. CONCLUSION

A successful mapping was done between the Symptoms that acted as Inputs and the Diseases. The Inputs considered included RBC Count, Platelet Count, WBC Count, Sickle cells' presence and Types of WBC (Basophil, Monocyte, and Lymphocyte). The detected disease were Thalasemia, Polycythemia vera, Sickle Cell anemia, Leukemia, Measles, Dengue, Kidney, Hypertension and Malaria. A comprehensive analysis of the diseases helped us to reach to the determining factors for every disease. Thus, Disease Detection was accomplished by analysis of the blood smear, Counting each type of blood cell, Extracting the features of RBCs, Determining the abnormalities in WBCs along with their type and finally, creation of an FIS for achieving the eventual mapping.

5. REFERENCES

- [1] A. K. T. F. Mehdi Habibzadeh, "Application of Pattern Recognition Techniques for the Analysis of Blood Smear Images," *Journal of Medical Informatics and Technologies*, vol. 18, 2011.
- [2] B.A.I. Khawaldeh, "Developing a Computer-based Information System to Improve the Diagnosis of Blood Anemia," Middle East University, 2013.
- [3] M.M.a.A.E. Mohammad Hamghalam, "Leukocyte segmentation in Giemsa-stained image of peripheral blood smears based on active contour," in *IEEE*, 2009.
- [4] C.D.R.Lorenzo Putzu, "White Blood Cells Identification and Classification from Leukemic Blood Image," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, 2013.
- [5] S. S. Niranjana Chatap, "Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour," *IOSR Journal of Computer Engineering*, vol. 16, no. 5, pp. 79-87, 2014.
- [6] M. C. P. Thejashwini M, "Counting of RBC's and WBC's Using Image Processing Technique," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, no. 5, pp. 2948-2953, 2015.