MULTIOBJECTIVE OPTIMIZATION BASED PATTERN MATCHING APPROACH FOR DISCOVERING CANCER

A PROJECT REPORT

Submitted by

VINOTHINI G.A.

in partial fulfillment for the requirement of award of the degree of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering
KUMARAGURU COLLEGE OF TECHNOLOGY,
COIMBATORE 641 049
(An Autonomous Institution Affiliated to Anna University, Chennai)

APRIL 2013

MULTIOBJECTIVE OPTIMIZATION BASED PATTERN MATCHING APPROACH FOR DISCOVERING CANCER

A PROJECT REPORT

Submitted by

VINOTHINI G.A.

in partial fulfillment for the requirement of award of the degree of

MASTER OF ENGINEERING

FACULTY OF COMPUTER SCIENCE AND ENGINEERING

KUMARAGURU COLLEGE OF TECHNOLOGY
COIMBATORE 641 049
(An Autonomous Institution Affiliated to Anna University, Chennai)

APRIL 2013

BONAFIDE CERTIFICATE

Certified that this project work titled **"MULTIOBJECTIVE OPTIMIZATION BASED PATTERN MATCHING APPROACH FOR DISCOVERING CANCER"** is the bonafide work of VINOTHINI G.A. (1120128024) who carried out the research under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other students.

Prof.N.JAYAPATHI, M.Tech.,	Mrs.P.Devaki, M.E.,
HEAD of THE DEPARTMEN	SUPERVISOR
Professor,	Associate Professor,
Depart of Computer Science and Engineering,	Department of Computer Science and Engineering,
Kumaraguru College of Technology,	Kumaraguru College of Technology,
Coimbatore - 641 049.	Coimbatore - 641 049.

Submitted for the Project Viva-Voce examination held on _____.

---------- ----------

Internal Examiner **External Examiner**

ABSTRACT

Discovering cancer and occurrence of disease in a human body is a large process in the traditional method. In the existing process of disease identification there is more number of steps to find the problem and also it takes more time for process. In proposed work describes the new method of finding problem using pattern matching approach. Here finding the occurrence of disease using the method of pattern matching process and introducing the high fidelity simulation of angiogenesis and uses this as the basis for a parallel search-based approach for the discovery of cancer and occurrence of disease that inhibit white blood cell growth. The parallel search based method produces the effective result and it's very helpful in hospitals and research centers.

ஆய்வுச்சுருக்கம்

ஒரு மனித உடலில் புற்றுநோய் மற்றும் நோய் நிகழ்வு கண்டுபிடிப்பது பாரம்பரிய முறையில் ஒரு பெரிய பணியாகும். முந்தைய செயல்முறையில் நோய் கண்டறிதல் மற்றும் பிரச்சனைகளை கண்டுபிடிக்க பல நடவடிக்கைகள் ஏற்படுத்துகிறது மற்றும் அதை செயலாக்க அதிக நேரம் எடுத்துக்கொள்கிறது. வகை பொருத்துதல் அணுகுமுறை சிக்கல் கண்டறியும் புதிய முறை விவரிக்கிறது. இங்கே முறை பொருந்தும் செயல்முறையை பயன்படுத்தி இரத்த குழாய் வளர்ச்சி உயர் நம்பக உருவகப்படுத்துதலானது மற்றும் நோய் நிகழ்வு கண்டுபிடித்து வெள்ளை இரத்த செல் வளர்ச்சியை தடுக்கும். புற்றுநோய் மற்றும் நோய் நிகழ்வு கண்டுபிடிப்பு ஒரு இணை தேடல் அடிப்படையிலான அணுகுமுறையை பயன்படுத்துகிறது. இந்த வகையான வகை பொருத்துதல் அணுகுமுறை மருத்துவமனைகள் மற்றும் ஆராய்ச்சி மையங்கள் மிகவும் பயனுள்ளதாக அமையும்.

ACKNOWLEDGEMENT

I express my profound gratitude to **Padmabhusan Arutselvar Dr.N.Mahalingam, B.Sc., F.I.E, Chairman**, **Dr.B.K. Krishnaraj Vanavarayar,B.Com., BL., Co-Chairman, Mr. M. Balasubramaniam, M.Com, M.B.A, Correspondent, Mr.Sankar Vanavarayar**, **M.B.A., PGDIEM, Joint Correspondent** and **Dr.S.Ramachandran Ph.D., Principal** for providing the necessary facilities to complete my thesis.

I take this opportunity to thank **Prof.N.Jayapathi M.Tech.**, Head of the Department, Department of Computer Science and Engineering, for his support and timely motivation. Special thanks to my Project Coordinator **Dr.V.Vanitha Ph.D.**, Senior Associate Professor, Department of Computer Science and Engineering, and project committee members for arranging the project review sessions.

I register my sincere thanks to my guide **Mrs.P.Devaki M.E.,** Associate Professor, Department of Computer Science and Engineering. I am grateful for her support, encouragement and ideas. I would like to convey my honest thanks to all **Teaching** and **Non Teaching Staff** members of the department and my classmates for their support.

I dedicate this project work to my **Parents** for no reasons but feeling from bottom of my heart, without their love this work would not be possible.

-G.A.VINOTHINI

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 IMAGE PROCESSING

Disease identification in traditional method is a large tedious process. The result disease identification and the chain identified by the expert knowledge. Some time result given by the experts may mistake. Human interaction in the testing process will not produce the accurate result. To reduce the problem of accuracy the pattern matching technique used.

By using this technique easily achieve the accuracy of results and also improving the accuracy of the testing sample. In image processing method images are compared with each other. The image may be a grayscale image or a colored image.

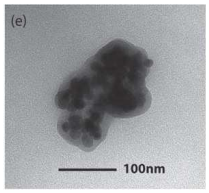

Figure 1.1 Gray scale image

1.2 PARALLEL COMPUTING

Parallel computing is a form of computation in which many calculations are carried out simultaneously. Operating on the principal that large problem can often be divided into smaller one, which is then solved concurrently.

Parallel computing programs are more difficult to write then sequential ones, because concurrency introduces several new classes of potential software bugs of which race conditions are most common. Communication and synchronization between the different subtasks are typically some of the greatest obstacles for getting good parallel program performance.

1.3 PATTERN MATCHING

Pattern matching is a family of tools for finding similar objects in different sources. In image processing, the pattern matching is used for locating a small image (called model) in a bigger one (called target image).

The simple way is to move the model in the target image and measure a similarity in each position. The position with the highest value of the similarity will be accepted as a result. The normalized correlation is used as a common metric for similarity.

The common way to reduce an amount of calculations is using the image pyramid. When using the image pyramid, the search is started on the reduced target

image, and results are improved in the next, more detailed one. This process is similar to the human vision.

Image file formats are standardized by means of organizing and storing digital images. Image files are composed of digital data in one of these formats that can rasterized for use on a computer display or printer. An image file format may store data in uncompressed, compressed, or vector formats. Once rasterized, an image becomes a grid of pixels, each of which has a number of bits to designate its color equal to the color depth of the device displaying it.

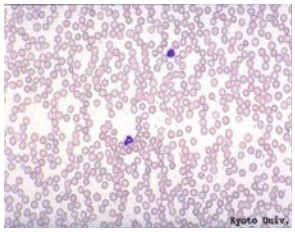

Figure 1.2 White blood cell image

White blood cells, or **leukocytes** (also spelled "leucocytes"; from the Greek word *leuko-* meaning "white"), are cells of the immune system involved in defending the body against both infectious disease and foreign materials. Five different and diverse types of leukocytes exist, but they are all produced and derived from a multipotent cell

in the bone marrow known as a hematopoietic stem cell. They live for about three to four days in the average human body. Leukocytes are found throughout the body, including the blood and lymphatic system.

The number of leukocytes in the blood is often an indicator of disease. There are normally approx. 7000 white blood cells per microliter of blood. They make up approximately 1% of the total blood volume in a healthy adult.

An increase in the number of leukocytes over the upper limits is called leukocytosis, and a decrease below the lower limit is called leukopenia. The physical properties of leukocytes, such as volume, conductivity, and granularity, may change due to activation, the presence of immature cells, or the presence of malignant leukocytes in leukemia.

The types of white blood cell images are given below. The step by step white blood cell developments are in 4 models. Each development process will generate different disease.

There are several different types of white blood cells. They all have many things in common, but are all distinct in form and function. A major distinguishing feature of some leukocytes is the presence of granules; white blood cells are often characterized as granulocytes or agranulocytes.

- **Granulocytes** (polymorphonuclear leukocytes): leukocytes characterized by the presence of differently staining granules in their cytoplasm when viewed under light microscopy. These granules are membrane-bound enzymes that act primarily in the digestion of endocytosed particles. There are three types of granulocytes: neutrophils, basophils, and eosinophils, which are named according to their staining properties.

- **Agranulocytes** (mononuclear leukocytes): leukocytes characterized by the apparent absence of granules in their cytoplasm. Although the name implies a lack of granules these cells do contain non-specific azurophilic granules, which are lysosomes. The cells include lymphocytes, monocytes, and macrophages.

Neutrophils

Eosinophils

Basophiles

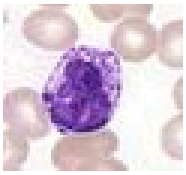

Figure 1.3 Blood cell images

1.4 LITERATURE SURVEY

1.4.1. A CELLULAR AUTOMATON MODEL FOR TUMOUR GROWTH IN INHOMOGENEOUS ENVIRONMENT

The authors (T. Alarcona, H.M. Byrneb, and P.K. Maini) proposed the cellular automation approach for performing the tumour growth in inhomogeneous environment. The advent of proteomics has brought with it the hope of discovering novel biomarkers. These can used to diagnose diseases, predict susceptibility and

monitor progression. Disease diagnostics using proteomic patterns is a revolutionary method to detect early-stage cancer. This effort has focused upon the mass spectral identification of the thousands of proteins.

Although many factors contribute to the heterogeneity of the environment in which cells proliferate, and focus our attention on blood flow. Despite the highly organized structure of vasculature in normal tissue, as compared to the chaotic appearance of vascular beds in cancer tissue the blood flow distribution in normal tissue is, for a number of reasons, strongly inhomogeneous.

The diagnostic models generated from proteomic patterns to provide highly specific results in testing. The most important result is that environmental inhomogeneity restricts dramically the ability of malignant colonies to grow and invade healthy tissue.

The major disadvantages are the identification of a specific biomarker may not provide any direct insight into how a disease may arise or progress. Diagnosis using proteins involves more number of comparisons.

1.4.2 OPTIMAL THRESHOLD FOR HEPATIC BLOOD VESSELS BASED ON STRUCTURAL ANALYSIS AND CANCER DETECTION

The author (Thomas *P* Conrads, and etal) proposed that, the conventional method of detecting hepatic cancer on the basis of three-dimensional images. It is

difficult to detect cancers with a three-dimensional structure. This approach proposed a method of detection of cancer on the basis of three-dimensional hepatic blood vessel regions which are extracted by threshold processing.

Recently, X-ray CT (computed tomography) scanners have been progressing rapidly in the reduction of imaging time and the improvement of resolution. Secondary detection detects cancers in the region not extracted in the above primary detection. Two kinds of region remain.

(1) A region may be separated from the blood vessel, because the density of the cancer is higher than threshold value (t3) but the density of the blood vessel connected to the cancer is low.

(2) The cancer is in the early stage and has a lower density than threshold *t*3, even though it is spherical. The secondary detection process detects the above two kinds of cancer regions. Cancer in the early stage has a low density, and detection is not easy even by human observation.

The density is not uniform in the cancer region, and it is desirable to eliminate small regions with high densities which is considered as noise. Consequently, three-dimensional smoothing is applied. Then, step 1 is applied if the density is at least t3, and step 2 is applied if it is lower.

The main advantage of this approach is to achieve threshold values using the edge detection of an image.3-Dimensional structural analysis for blood vessel, thinning algorithm for extraction blood vessel.

The cancer detection was attempted in CT images of seven cases, including cases of early stage cancer. Computation time is high for extraction and structural analysis of the blood vessel region. For practical application, it is desirable to examine more cases, and to adjust the feature parameters.

1.4.3 HUMAN–COMPUTER INTERACTION FOR THE GENERATION OF IMAGE PROCESSING APPLICATIONS

The author (Yong-Gyu Lee and Gilwon Yoon) proposed that, the development of customized image processing applications is time consuming and requires high level skills. The human computer interaction method described the design of an interactive application generation system oriented towards producing an image processing software programs.

The description is focused on two models which constitute the core of the human–computer interaction. Here the human interaction is main objective in finding of disease in pattern matching approach. By using the human interaction processing for the pattern matching approach has the following advantages.

i) Processing objective is to segmenting the images into regions corresponding to individual cells.

ii) Application ontology is generated using the formulation model identities and end user information.

But there is a major disadvantage in the existing work, still a cognitive gap between the evaluation of the resulting images and the correction of the query contents with new descriptors and new descriptor confidences.

1.4.4 AUTOMATED LOCALISATION OF THE OPTIC DISC, AND BLOOD VESSELS FROM DIGITAL COLOUR IMAGES

The authors (Chanjira Sinthanayothin, James F Boyce, and etal) proposed that, an automated localisation method for digital image. Automatically recognize the main components of the digital colour images in blood vessels were identified by means of a multilayer neural net and the inputs were derived from a principal component analysis (PCA) of the image and edge detection.

The main features of a fundus retinal image were defined as the optic disc, fovea, and blood vessels. The methods described for their automatic recognition and location. The optic disc was located by identifying the area with the highest variation in intensity of adjacent pixels.

Blood vessels were identified by means of a multilayer perception neural net, for which the inputs were derived from a Principal Component Analysis (PCA) of the image and edge detection of the first component of PCA.

The fovea was identified using matching correlation together with characteristics typical of a fovea—for example, darkest area in the neighborhood of the

optic disc. The main components of the image were identified by an experienced ophthalmologist for comparison with computerized methods.

These approaches first study the optic disc, blood vessels, and fovea was accurately detected. The identification of the normal components of the retinal image will aid the future detection of diseases in these regions.

In diabetic retinopathy, for example, an image could be analyzed for retinopathy with reference to sight threatening complications such as disc neovascularisation, vascular changes, or fovea exudation.

The optic disc, blood vessels, and fovea were accurately detected using Principal Component Analysis (PCA).Computer based algorithms were used to detect the main regions of the images without any intervention from an operator. The accuracy of the detection was high for the optic disc, blood vessels.

The major disadvantage is WBC count. The process of white blood cell classification is not performed in the existing approach. So it can't able to get the accurate counting result. Similarly 2-D image processing is not providing the correct image view so, 2D processing is not valid process in the existing approach.

1.4.5 A FRAMEWORK FOR WHITE BLOOD CELL SEGMENTATION IN MICROSCOPIC BLOOD IMAGES USING DIGITAL IMAGE PROCESSING

The author (Farnoosh Sadeghian, Zainina Seman, and et al) proposed that a framework for WBC classification. Digital image processing techniques were used for analysis the disease occurrence. A disease like leukemia is detected based on the amount and condition of the WBC. The main objective of this method is to segment the WBC to its two dominant elements: nucleus and cytoplasm.

White blood cells (WBC) or leukocytes play a significant role in the diagnosis of different diseases, and therefore, extracting information about that is valuable for hematologists. In the traditional method, digital image processing techniques had helped to analyze the cells that lead to more accurate, standard, and remote disease diagnosis systems.

White blood cell segmentation, Active contours, Snake algorithm, Zack thresholding are the major processing in the WBC counting method. Snake algorithm is not related to size and color of nucleus because there are various shapes of nucleus in a different kind of white blood cells.

So, it gives high accuracy result in segmenting nucleus in any type of WBCs and in any capture illumination that cause different color space in images. In cytoplasm segmentation method uses the thresholding technique; background is completely segmented from other components.

There is also a drawback occurred that is, in blood image, there are similar color scales in WBC's with some other blood particles it cause a big error in thresholding method for cytoplasm segmentation (sub-image).

However, there are few complications in extracting WBC from the given input due to wide variation of shape in cells, size, edge, and position. This is the major drawback in the WBC counting approach.

1.4.6 IMPLEMENTATION OF FUZZY INFERENCE SYSTEM FOR WHITE BLOOD CELL CANCER DETECTION USING PATTERN MATCHING

The authors (Devesh D. Nawgaje1, Dr. Rajendra D.Kanphade2) proposed that a method for differential count of various types of white blood cells (WBC) in bone marrow smears. It is used to help detect infection, anemia, and leukemia. Fuzzy Inference System (FIS) approach to detect edge within color bone marrow microscopic images.

A lack of red blood cells leads to less oxygen being delivered to the organs and tissues of human body. This is called anemia, and it can make feel tired and breathless. These are the main drawbacks in the fuzzy interface method to reduce the drawback by introducing fuzzy logic process. The same fuzzy rule base is then implemented on DSP TMS320C6711 which provides a real time implementation.

Fuzzy Inference System (FIS); Leukemia; Digital Signal processing (DSP); Edge detection is used for the fuzzy interface process. The advantage of using fuzzy

interface, the image edge detection using FIS algorithm has been successful in obtaining the edges of the cancer affected cells that are present in medical images.

This method provides accuracy of the algorithm and also helps to detect exact shape of cancer cells. But the Implementation is difficult in DSP TMS320C6711 and also the WBC segmentation or counting is tedious process in DSP.

1.4.7 MULTIOBJECTIVE OPTIMIZATION BASED-APPROACH FOR DISCOVERING NOVEL CANCER THERAPIES

The authors (Arthur W. Mahoney, Gregory J. Podgorski, and Nicholas S. Flann) proposed that a search based approach for detecting novel cancer and therapies. The traditional method is to use a microscope to select an area of interest in a bone marrow slide, detect a white blood cell, classify the cell based on his knowledge and experience, increase the count of the corresponding cell class, and repeat the cycle until all cells in the area of interest are counted.

To perform all these processes manually is a very tedious process. Manual processes like,

- Detecting white blood cell,
- Classify the cell,
- Increase the count of the corresponding cell class.

These are very tedious process in the traditional method of using a microscope. This work also has some disadvantages. They are,

- Less Accuracy.
- Take more time to test the sample.
- High cost.

CHAPTER 2

IMPLEMENTATION

2.1 EXISTING SYSTEM

In traditional method an expert is to use a microscope to select an area of interest in a bone marrow slide, detect a white blood cell, classify the cell based on his knowledge and experience, increase the count of the corresponding cell class, and repeat the cycle until all cells in the area of interest are counted. To perform all these processes manually is a very tedious process for trained expert and, thus, an automated differential counting system that helps in saving time and money is highly desirable. Bone marrow cells are normally diagnosed by light microscopy. Flow cytometry, which is normally used for differential blood cell counting of peripheral blood, is not suitable for bone marrow cells. This is because, in addition to its high price and complicated structure, markedly hyper cells are packed bone marrow and sclerotic bone marrow may yield too few cells for adequate analysis by flow cytometry. Data from flow cytometry should always be correlated with that from light microscopy.

2.1.1 Draw Backs

- Disease finding in traditional method is tedious process.

- Accuracy of finding presence of tissues WBC RBC counting is low.

2.2 PROPOSED WORK

This work proposes a method for the classification of white blood cells using only their nucleus information. This idea is very useful in practice because one of the difficulties in the differential counting in bone marrow is how to deal with the cells that touch each other. This problem occurs frequently in cells of the bone marrow because the WBC is very dense. If the cell classification is based only on the information about the nucleus, then no need to segment the entire cell, and only nucleus segmentation is adequate.

Although many techniques have been applied to cell segmentation, this problem is not solved, especially in touching cells. To decouple the effects of segmentation errors, first extract the features from manually segmented nucleus of a white blood cell based on morphological granulometries. Then apply Baye's classifiers and artificial neural networks to the problem of white blood cell classification of single-cell images and compare their results to those obtained by an expert and also compare the results to other classifiers and other previously proposed features.

WBC Counting:

- Using X & Y pixel co-ordinates.
- Counting every grain in the image.
- Segmenting the WBC.

Morphological Granulomertrics:

The parallelism technique is done with the help of mathematical form called morphological GranulomErtrics.

Baye's Classifier:

To avoid the problem of white blood cell classification of single-cell images by applies the Baye's classifiers. It also used to compare their results with the trained data set. A theorem described how the conditional probability of set possible causes for a given observed event it can computed from knowledge of the probability of each cause and the conditional probability of the outcome of each cause.

In proposed system the occurrence of disease can find using the pattern matching and introducing high fidelity simulation of angiogenesis and uses this as the basis for a parallel search-based approach for the discovery of cancer and occurrence of disease that inhibit white blood cell growth.

Initially train the dataset for matching the pattern. The comparison processes consist of edge detection and WBC counting. The edge detection and WBC counting operations are process in parallel manner.

2.3 PROBLEM DEFINITION

Leukaemia is a type of cancer that affects the white blood cells. In leukemia, white blood cells become abnormal, and divide and grow in an uncontrolled way. They stay in the bone marrow and keep reproducing in an uncontrolled way. These abnormal white blood cells fill up the bone marrow and prevent it from making healthy white blood cells. This means the body is less able to fight off infections. A lack of platelets can lead to problems with the blood-clotting system, and results in bleeding and bruising much more easily than usual.

There are various types of WBCs (also called leukocytes) present in blood smear images. A differential count of cells determines the relative percentages of each one of them in the blood, note that any abnormal appearance of the cells and the presence of any abnormal immature cells. The differential count can used to help detect infection, anemia, and leukemia or to follow the progress of treatment. Any infection or acute stress results in increased production of WBCs. This usually entails an increased number of cells and an increase in the percent of immature cells (mainly band cells) in the blood. High WBC counts may indicate the presence of an inflammatory and immune response or it may result from other conditions such as leukemia.

It is important to realize that an abnormal increase in one type of leukocyte can produce an apparent decrease in the percentage of other types. Because of locating, identifying, and counting the different classes of WBC manually are tedious tasks, mechanisms for automating this activity have been investigated. Besides relieving the already time intensive job from the technologist, an advantage of automation is that

many more cells can inspected, giving rise to better statistical information in the differential counts.

Edge detection of blood cell images is a crucial step for automatic cell analysis, because the success of the final classification depends mainly on the correct edge detection. This is the most difficult part of an automatic cell analysis, because there is considerable uncertainty in the microscopic images. The maturity classes of the white blood cells actually represent a continuum, cells frequently overlap each other, there are staining and illumination inconsistencies, and there is fairly wide variation of size and shape of nuclear and cytoplasm regions within given cell classes. This uncertainty makes bone marrow image edge detection is a difficult and challenging problem.

The earliest practical-oriented effort in WBC recognition on smear images was carried out. Later efforts have been focused on different aspects of segmentation of smear images and WBC differentiation. To introduced a hierarchical procedure using priority information regarding chromatic properties of background and cell properties and investigated the multi-step segmentation scheme. These implement the automatic thresholding method. Besides histogram thresholding, clustering, edge detection, and region growing methods are often used. As these methods are one-shot decision, they violate the Marr's Principle, because once a wrong decision has been made, it is difficult and it's impossible, to correct it. A fuzzy-based cell segmentation approach consistent with the Principle of Least Commitment was developed. The proposed system provided very accurate segmentation results working on monochrome bone marrow images.

On the other hand, in the case of color images, problems introduced by colors' low saturation and color illumination have to be considered. Even though human experts assert that white blood cells can differentiate with grayscale images, they use color bone marrow microscopic images to carry out this analysis. Although grayscale images need less storage resource and processing time than color bone marrow microscopic images.

2.4 MODULE DESCRIPTION

The Modules in the proposed system are,

Module 1: Pre Processing
Module 2: Training
Module 3: Classification
Module 4: Results

2.4.1 PRE PROCESSING

This module is used to give input for comparison process to find the disease. The input should be a blood cell image with the report. This will be collected from hospitals and bio-tech lab. These are called datasets this dataset consist of blood cell image, counting report of WBC, RBC, plasma, and myeloblast, promyelocyte, myelocyte etc., this process initially takes the blood samples and scans those cells as images. Every cell is spitted as images.

In preprocessing the original image will converted into gray scale image by using the gray scale value and also perform the segmentation process. This process consists of two parts they are Morphology Granulometry and Edge Detection process.

In morphology granulometry the process of noise removal and erosion process are takes place. This is very useful process in image processing because it produces accuracy and clarity of an image. Edge detection is an importation process in image processing method. Before going to processing with an image the edge detection is important part of image processing.

In proposed method "CANNY'S EDGE DETECTION" algorithm is used for detecting the edges of an image. This algorithm is an effective algorithm for detecting an edge of a blood cell image.

2.4.2 TRAINING

This learning process involved for the scanning process. Then it splits those cells as separately according to the cell's type. Then, this cell types are learnt in the databases as images according to that diseases type. Learning process is nothing but collecting all dataset for comparison process, dataset consist of an image counting report of WBC, RBC, plasma, and myeloblast, promyelocyte, myelocyte counting.

There are lot of image dataset are collected and these dataset are stored in data base this is a tedious process of storing the dataset in database. Because there are different types of disease are there according to the cell's type. Each and every type of disease is classified and stored in data base are in correct manner.

2.4.3 CLASSIFICATION

This module is to compare the input blood sample with the stored blood sample data set. These processes separate the scanned images then check this with the input blood cells. Then it will compare with existing images and separate those diseases by images. This module also helps us to classify the white blood cells.

Input blood cell image is compared with these dataset and produced an output. This comparison will provide the accurate resulting of occurrence of disease and also finding the types of disease.

The BAYE'S CLASSIFICATION method is used for classifying the WBC count from the given input blood cell image. By using this theorem the counting of WBC is an easiest way and also the classification of blood cell image is achieved.

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}.$$

2.4.4 RESULTS

This module is to detect the disease from segmentation and classification process. This will make sure the blood sample has any disease or not. Those diseases

may be a tumor, anemia, AIDS, leukemia, and jaundice etc. The parallel computing process is done for simultaneous operation. The classification and learning process uses parallel search based approach for performing the pattern matching process.

Two or more users sending request to the server at the same time for getting result. Without parallel computing process result cannot be provided at a particular time to the user. By using parallel computing easily get the result and no need to wait for any other process. Each and every process can run in parallel manner so times to performing more processes are reduced.

CHAPTER 3

RESULT

3.1 HARDWARE REQUIREMENTS

Processor	: Intel i3
Speed	: Above 500 MHz
RAM capacity	: 2 GB
Hard disk drive	: 80 GB
Key Board	: Del108 keys
Mouse	: Logitech Optical Mouse
Motherboard	: Intel
Monitor	: 17" Del

3.2 SOFTWARE REQUIREMENTS:

Operating System	: Windows xp
Front end used	: C#.Net
Back End	: SQL Server 2000

3.3 DATABASE DESIGN

The Table 3.1 shows the dataset consist of different type of tissues and there data type value. Tissues like myeloblast, promyelocyte, myelocyte, metamyelocyte are in the dataset.

Table 3.1.Database for Tissues value

Column Name	Data Type	Length	Allow Nulls
id	int	4	
myeloblast	int	4	✓
promyelocyte	int	4	✓
myelocyte	int	4	✓
metamyelocyte	int	4	✓
band	int	4	✓
pmnnv	int	4	✓
disease	varchar	50	✓

In Table 3.2 shows the tissue count and their primary key value. Threshold value is set for calculating the tissue count and also compared with the input blood cell image dataset.

Table 3.2.Tissue counting Result

Column Name	Data Type	Length	Allow Nulls
Sid	numeric	9	
myeloblast	numeric	9	✓
promyelocyte	numeric	9	✓
myelocyte	numeric	9	✓
metamyelocyte	numeric	9	✓
band	numeric	9	✓
PMNnv	numeric	9	✓

3.4 CONCLUSION

In existing system to detect the cells in a bone marrow slide, detect a white blood cell, and classify the cell based on human's knowledge and experience are huge processes. To perform all these processes manually is a very tedious process for trained experts. The classification method is proposed here for classifying white blood cells using only their nucleus information. This idea is very useful in practice because the differential counting process in bone marrow is achieved . First extract features from manually segmented nucleus of a white blood cell based on morphological granulometries.

This approach applies Baye's classifiers and artificial neural networks for performing the white blood cell classification of a single-cell image and compares these results with the dataset.

This approach has demonstrated a proposed framework for segmenting white blood cells using integration of concepts in image processing. The proposed scheme has two parts: The nucleus segmentation part is based on morphological analysis, and the plasma segmentation is based on pixel-intensity thresholding.

3.5 FUTURE WORK

The identification of disease and finding the disease occurrence are processed by using the method of pattern matching and parallel computing manner. The parallel computing process is used for reducing the processing time for providing result in faster manner. In feature, perform the same process with the help of large computer network and also reduce the processing time of disease identification that, the data send from different area.

APPENDIX

SOURCE CODE

CANNY EDGE DETECTION

```
namespace ImageLib.Imaging.Filters
{
using System;
using System.Drawing;
using System.Drawing.Imaging;
public class CannyEdgeDetector : IFilter
{
private IFilter grayscaleFilter = new GrayscaleBT709();
private GaussianBlur          gaussianFilter = new GaussianBlur();
private byte low Threshold = 20;
private byte high Threshold = 100;
private static int[,]    xKernel = new int[,]
{
{-1,  0,  1},{-2,  0,  2},{-1,  0,  1}
};
private static int[,] yKernel = new int[,]
{
{ 1,  2,  1},{ 0,  0,  0},{-1, -2, -1}
};
public byte LowThreshold
```

```
{
get { return lowThreshold; }
set { lowThreshold = value; }
}
public byte HighThreshold
{
get { return highThreshold; }
set { highThreshold = value; }
}public double GaussianSigma
{
get { return gaussianFilter.Sigma; }
set { gaussianFilter.Sigma = value; }
}
public int GaussianSize
{
get { return gaussianFilter.Size; }
set { gaussianFilter.Size = value; }
}
public CannyEdgeDetector()
{
}
public CannyEdgeDetector(byte lowThreshold, byte highThreshold)
{
this.lowThreshold    = lowThreshold;
this.highThreshold   = highThreshold;}
```



```
public CannyEdgeDetector(byte lowThreshold, byte highThreshold, double sigma)
{
this.lowThreshold    = lowThreshold;
this.highThreshold   = highThreshold;
gaussianFilter.Sigma = sigma;
}
public Bitmap Apply(Bitmap srcImg)
{
Bitmap grayImage = (srcImg.PixelFormat == PixelFormat.Format8bppIndexed) ?
srcImg : grayscaleFilter.Apply(srcImg);
Bitmap blurredImage = gaussianFilter.Apply(grayImage);
int width = srcImg.Width;
int height = srcImg.Height;
BitmapData srcData = blurredImage.LockBits(
new Rectangle(0, 0, width, height),
ImageLockMode.ReadOnly, PixelFormat.Format8bppIndexed);
Bitmap dstImg = ImageLib.Imaging.Image.CreateGrayscaleImage(width, height);
BitmapData dstData = dstImg.LockBits(
new Rectangle(0, 0, width, height),
ImageLockMode.ReadWrite, PixelFormat.Format8bppIndexed);
int stride = srcData.Stride;
int offset = stride - width;
int      widthM1 = width - 1;
int heightM1 = height - 1;
int i, j, ir;
```



```
double v, gx, gy;
double orientation, toPI = 180.0 / System.Math.PI;
byte leftPixel = 0, rightPixel = 0;
byte[] orients = new byte[width * height];
unsafe
{
byte * src = (byte *) srcData.Scan0.ToPointer() + stride;
byte * dst = (byte *) dstData.Scan0.ToPointer() + stride;
int p = width;
for (int y = 1; y < heightM1; y ++)
{
src++;
dst++;
p++;
for (int x = 1; x < widthM1; x ++, src ++, dst ++, p ++)
{
gx = gy = 0;
for (i = 0; i < 3; i++)
{
ir = i - 1;
for (j = 0; j < 3; j++)
{
v = src[ir * stride + j - 1];
gx += v * xKernel[i, j];
gy += v * yKernel[i, j];
```



```
}}
dstImg.UnlockBits(dstData);
blurredImage.UnlockBits(srcData);
blurredImage.Dispose();
if (grayImage != srcImg)
grayImage.Dispose();
return dstImg;
}}}
```

NEURAL NETWORKS LEARNING ALGORITHM

```
using System.Collections;
using System.Drawing;
namespace NeuralNetwork
{
public abstract class LearningAlgorithm
{
#region PROTECTED FIELDS
protected NeuralNetwork nn;
protected float ERROR_THRESHOLD = 0.001f;
protected int MAX_ITER = 1000;
protected protected float[][] outs;
protected int iter = 0;
protected float error = -1;
protected LearningListener listener;
```

```
#endregion
#region CONSTRICTOR AND METHODS
public LearningAlgorithm(NeuralNetwork n)
{
nn = n;
}
public virtual void Learn(float[][] inputs, float[][] expected_outputs)
{
if (inputs.Length < 1)
throw new Exception("LearningAlgorithme : no input data : cannot learn from
nothing");
if (expected_outputs.Length < 1)
throw new Exception("LearningAlgorithme : no output data : cannot learn from
nothing");
if (inputs.Length != expected_outputs.Length)
throw new Exception("LearningAlgorithme : inputs and outputs size does not match :
learning aborded ");
ins = inputs;
outs = expected_outputs;
}
protected bool Continue()
{
if (listener != null) return listener.Continue();
else return true;
}
protected void Progress()
{
```



```
if (listener != null) listener.Progress(iter, error);
}
#endregion
}
public float Gamma
{
get { return gamma; }
set { gamma = (value>0)?value:gamma; }
}
#endregion
#endregion
#region LEARNING METHODS
public override void Learn(float[][] inputs, float[][] expected_outputs)
{
base.Learn(inputs, expected_outputs);
float[] nout;
float err;
iter = 0;
do
{
error = 0f;
e = new float[nn.N_Outputs];
for(int i=0; i<ins.Length; i++)
{
err = 0f;
```



```
nout = nn.Output(inputs[i]);
for(int j=0; j<nout.Length; j++)
{
e[j] = outs[i][j] - nout[j];
err += e[j] * e[j];
}
err /= 2f;
error += err;
ComputeA(i);
setWeight(i);
}
iter++;
Progress();
if (!Continue()) break;
}
while(iter < MAX_ITER && this.error > ERROR_THRESHOLD);
}protected void ComputeA(int i)
{
float sk;
int l = nn.N_Layers-1;
for (int j=0; j<nn[l].N_Neurons; j++)
nn[l][j].A = nn[l][j].OutputPrime * e[j];
for(l--; l>=0; l--)
{
for (int j=0; j<nn[l].N_Neurons; j++)
```



```
{
sk = 0f;
for (int k = 0; k<nn[l+1].N_Neurons; k++)
sk += nn[l+1][k].A * nn[l+1][k][j];
nn[l][j].A = nn[l][j].OutputPrime * sk;
}}}
protected void setWeight(int i)
{
float[] lin;
for(int j=0; j<nn.N_Layers; j++)
{
if (j==0) lin = ins[i];
else lin = nn[j-1].Last_Output;
for (int n = 0; n< nn[j].N_Neurons; n++)
{
for (int k = 0; k < nn[j][n].N_Inputs; k++)
nn[j][n][k] += alpha * lin[k] * nn[j][n].A + gamma * (nn[j][n][k] -
nn[j][n].Last_W[k]);
nn[j][n].Threshold -= alpha * nn[j][n].A + gamma * (nn[j][n].Threshold -
nn[j][n].Last_Threshold);
}}}
#endregion
}
#endregion

#region GeneticNeuralNetwork
```

```
protected class GeneticNeuralNetwork : IComparable
{
protected float[] genes;
protected float sq_err = -1f;
protected NeuralNetwork nn;
public float this[int index]
{
get { return genes[index]; }
set { genes[index] = value; }
}
public float Error
{
get { return sq_err; }
set {sq_err = value;}
}
public int CompareTo(Object other)
{
return sq_err.CompareTo(((GeneticNeuralNetwork)other).Error);
}}
#endregion
}
public interface LearningListener
{
bool Continue();
void Progress(int iteration, float error);}}
```


SCREEN SHOTS

The Figure A1 shows the main form of WBC count.

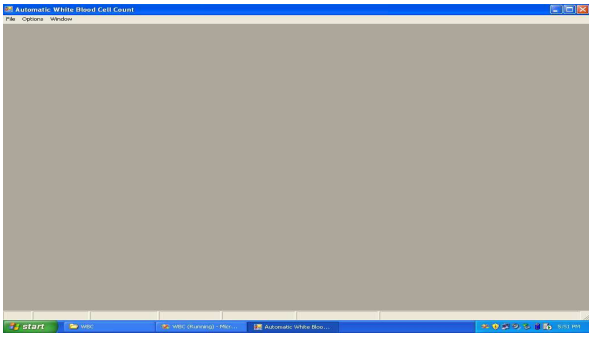

Figure A1 White Blood Cell Calculation

The Figure A2 shows the input image of WBC classifications. This is a normal blood cell image developed by using digital camera with the help of Bio-tech lab.

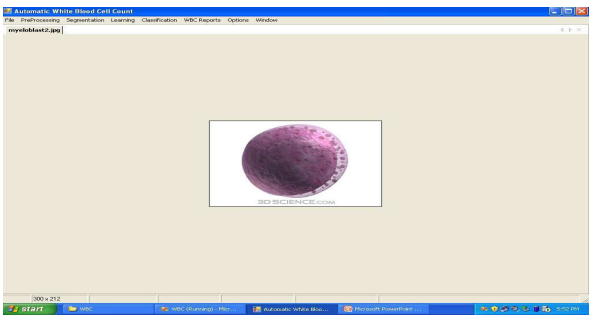

Figure A2 Blood Cell Image

This Figure A3 shows the gray scale image of input image. The WBC counting process is done by using the gray scale image because; it produces effective result in counting. So gray scale image is used for perform the pattern matching and classification process.

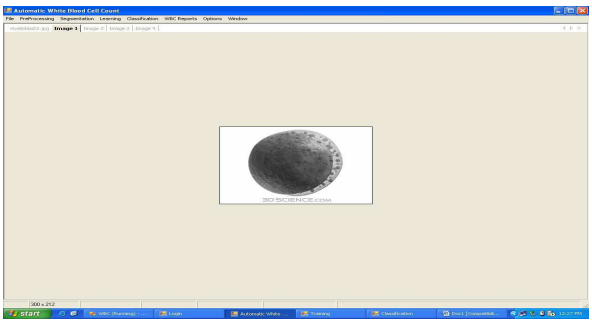

Figure A3 Gray Scale Image

The Figure A4 is used to darking the cells in the image for easy of counting the tissuses in blood cell image

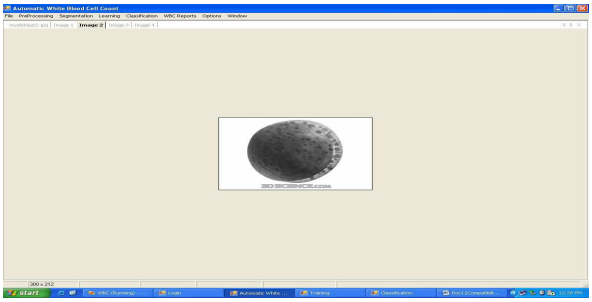

Figure A4 Erosion Gray Scale Image

This Figure A5 is used for removing the noise from the input image for providing good image quality.

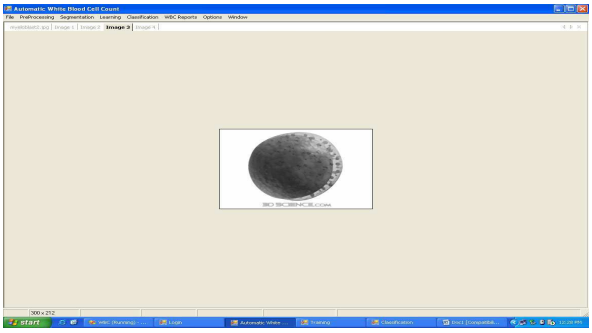

Figure A5 Noise Removed Gray Scale Image

The Figure A6 shows an edge detected blood cell image. Before go to the classification process initially detect the edges of an image to provide an accurate classification result.

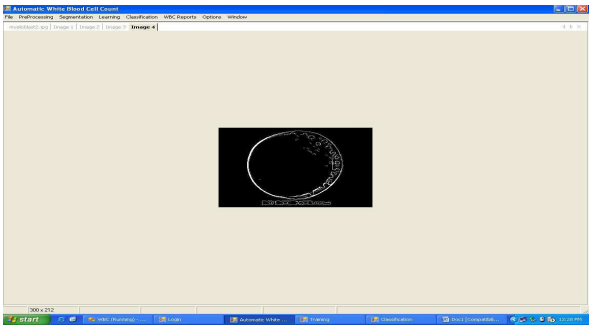

Figure A6 Edge Detection Image

The Figure A7 shows a dataset gathering process. Collect all the datasets like image tissues values (counting) WBC, RBC, PLASMA counting etc,.The WBC value of input image is compared with this dataset for finding accurence of disease.

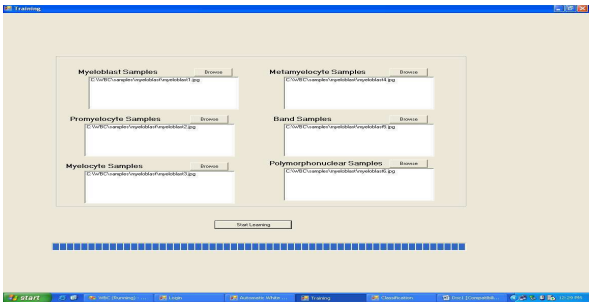

Figure A7 Dataset gathering

The Figure A8 shows the completion of data gathering process. After the starting of data gathering process first collect different type of data from different location. Then test the dataset and stored in to the database.

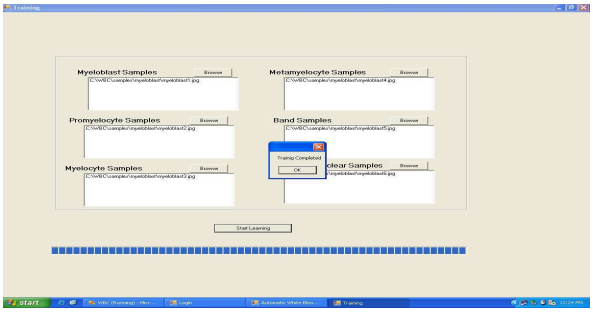

Figure A8 Completion of Data Gathering Process

The Figure A9 performs the classification process of given input blood cell image. The classification is done with the help of the above process like noise removal, edge detection, and dataset gathering.

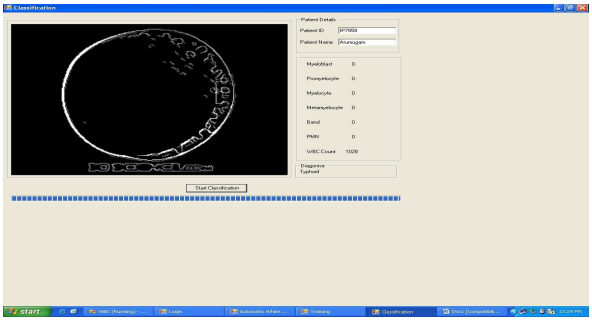

Figure A9 Classifications

The Figure A10 shows the final classification of disease identification. It shows the count value of every tissues and blood cells. The counting value shows only the binary value. The binary value is set based on the threshold value.

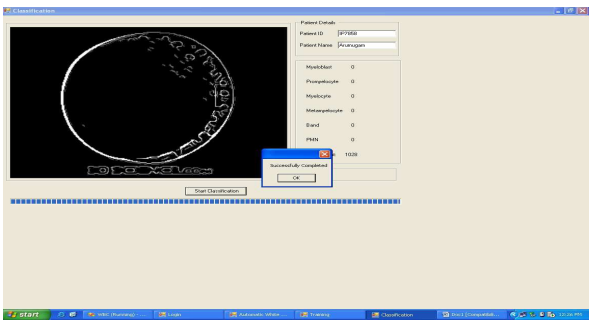

Figure A10 Classifications Result

REFERENCE

1. T. Alarcona, H.M. Byrneb, P.K. Maini "A cellular automaton model for tumour growth in inhomogeneous environment." Centre for Mathematical Biology, Mathematical Institute, University of Oxford, UK Received 16 October 2002; received in revised form 30 May 2008; accepted 11 June 2010.

2. Arthur W. Mahoney, Gregory J. Podgorski, and Nicholas S. Flann "Multiobjective Optimization Based-Approach for Discovering Novel Cancer Therapies." A.W. Mahoney and N.S. Flann are with Computer Science Department, Utah State University, Logan G.J.

3. Chanjira Sinthanayothin, James F Boyce, Thomas H Williamson "Automated localization of the optic disc, fovea, and retinal blood vessels from digital colour fundus images". Image Processing Group, Department of Physics, King's College, London C Sinthanayothin J F Boyce Department of Ophthalmology, St Thomas's Hospital, London

4. Devesh D. Nawgaje1, Dr. Rajendra D.Kanphade2 "Implementation of Fuzzy Inference System For White Blood Cell Cancer Detection Using DSP TMS320C6711" 1Faculty, EXTC, SSGMCE Shegaon, Shegaon-444203 India 2Principal, Dhole Patil College Of Engineering, Wagholi, Pune-412207 India,

5. Farnoosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, and M-Iqbal Saripan "A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing. "To the author(s) 2009

6. Thomas *P* Conrads, Ming Zhou, Emmanuel *F* Petricoin III, Lance Liotta and Timothy *D* Veenstra"Cancer diagnosis using proteomic patterns" Author for correspondence Biomedical Proteomics Program, SAIC-Frederick, Inc., National Cancer Institute

7. Yong-Gyu Lee and Gilwon Yoon "Improvement of Blood Detection Accuracy using Image Processing Techniques suitable for Capsule Endoscopy. "World Academy of Science, Engineering and Technology 65 201

LIST OF PUBLICATIONS

1. G.A.VINOTHINI, and Mrs. P.DEVAKI, "Multiobjective optimization based pattern matching approach for discovering cancer" National conference on recent trends in IT SURYARISE 2013, Surya Engineering College, Erode.

2. G.A.VINOTHINI, and Mrs. P.DEVAKI, "Optimization base sample toning approach for discovering tumour" International Conference on Advanced Computing Machines and Embedded Technologies, J.K.K.N College of Engineering and Technology, Komarapalayam, 2013.