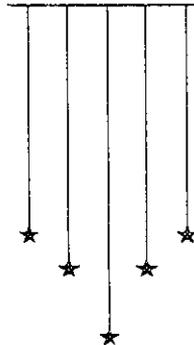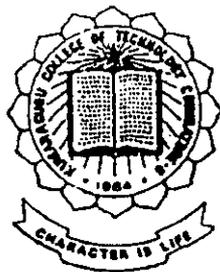# CANCER DIAGNOSIS – IMAGE MANIPULATION

PROJECT REPORT

Submitted in partial fulfillment of the requirements for award of degree

M.Sc., [Applied Science] Software Engineering

Submitted By

S. Prabhu Sethupathy

9837S0056

UNDER THE GUIDANCE OF,

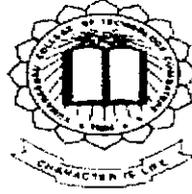| External Guide	Internal Guide
Dr Sudarsh Kailash	**Mr. M. Manikantan**
General Manager,	Lecturer,
Siemens Information Systems	CSE Department

Department Of Computer Science and Engineering
KUMARAGURU COLLEGE OF TECHNOLOGY
(Affiliated to Bharathiar University)
Coimbatore – 641006.

CERTIFICATE

Department of Computer Science and Engineering

Kumaraguru College of Technology

Coimbatore-641006.

This is to certify that the project work entitled "

Cancer Diagnosis-Image Manipulation"

Has been submitted by

Mr.S.Prabhu Sethupathy

In partial fulfillment of the award of the degree of

Master of Science in Applied Science- Software Engineering of

Bharathiar University, Coimbatore

during the academic year 2002-2003.

_____   _____
· Guide   Head of the Department

Certified that the candidate was examined by us in the Project Work Viva Voce

Examination held on ___4-04-2003___ and the University Register Number

was 9837S0056.

_____   _____
Internal Examiner   External Examiner

SIEMENS

Course Co-ordinator
Dept. of Computer Science & Engineering
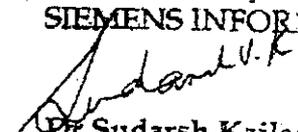Kumaraguru College of Technology
Chinnavedampatti
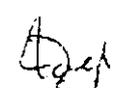Coimbatore – 641006.

Dear Sir,

We are pleased to inform that S. Prabhu Sethupatny is undergoing a project in our organization from December 2002.

External Guide:

Mr. Sai Kalyan
SISL/ Bangalore.

Yours faithfully,
SIEMENS INFORMATION SYSTEMS LTD.

Dr.Sudarsh Kailas
General Manager - Technology

S.Patanjali
Cheif Manager - HR

ACKNOWLEDGEMENT

"Gratitude is the mother of all virtues"

Any accomplishment requires the effort of many people and this work is no different. I am very much obliged to **Dr.K.K.Padmanaban B.Sc.(Engg), M.Tech, Ph.D.,** Principal, Kumaraguru College of Technology for giving me the opportunity to undertake the project work.

I express my sincere thanks to **Dr.S.Thangasamy Ph.D.** Prof. & Head of the Department, Kumaraguru College of technology for giving me this golden opportunity to carry out my project work successfully.

I wish to express my deep sense of gratitude to **Mrs.S.Devaki B.E, M.S, Asst.Prof,** for the permission she has granted and for the encouragement, she has given for the successful completion of the project.

Words fail to express my feelings for the encouragement and support bestowed on me by my Project Guide **Mr.M.Manikandan MCA,** Lecturer, and to **MR.Ramasubramaniam MCA,** Lecture Kumaraguru College of Technology.

I wish to record my sincere thanks to Dr.Subramaniam, Dr.Srivatsav and Dr.Umesh for giving me valuable suggestions and constant encouragement for the successful completion of the project work.

I am deeply indebted professionally to **Mr.Sudarsh Kailas,** General Manager -Technology, Siemens Information Systems, for providing me a chance to work under him and for all his encouragement's, advice and directions all through the course of the project.

I thank all my faculties whose diligent effort has led me to complete the project successfully

Motivation is the driving force in our lives and to me it is my parents and friends, words are inadequate to express my thanks to them.

Thanks to the Almighty for being with me all through the course of the project.

PREFACE

This document caters to the need of the readers to have a clear knowledge about the project. The document is put up in an ordered fashion to improve the understandability of the problem and the way of approach involved in a software development phase. At the end of every chapter Glossary of terms is given for further reader friendliness.

Chapter 1: this chapter discusses about the problem statement. The system constraints involved in the development of the project is discussed in detail. The processing environment of the project and the user characteristics are also dealt in this chapter. The last section in this chapter gives you the sample form of the input where the processing has to be done.

Chapter 2:Project plan, gives the reader the plan schedule of the project. It includes the development phase, the development schedule, monthly schedule, the programming tools and the manner of demonstration. The development phase deals with the various phases of the development of the project. The development schedule and the monthly schedule gives the details about the schedule of the project and the monthly activities involved. The programming tool tells you about the languages used for the development.

*Chapter 3:*Software requirements and specification gives the entire knowledge about the software specification and the hardware specification. Information on the process specification is also dealt. Further reading in the document gives the details on the performance specification and the system model template.

Chapter 4: Chapter 4 discusses on the design document of the system under development. It covers the sections on the data flow diagram, architectural diagram, data dictionary, pseudo code of the modules and the structured chart.

*Chapter 5:*Testing involves the details on the types of testing involved in the project. Reading on the types of testing involved gives the readers a clear view of its employment. Performance test is also involved with the details based on its functionality.

ABSTRACT

Cancer can originate in almost any part of the body. Different types of cancer respond to different forms of treatment. Thus, it is vital that cancer be diagnosed correctly. Cancer diagnosis's at an early stage is a task of paramount importance. Diagnosing cancer involves analyzing the cell at various stages.

The goal of this project is to generate image manipulation tools for the end user. These tools enhance the visibility of the image based on the user's perception.

The source image specimen is available in the digital format. The functions implemented provide image manipulation features to the user. The features provided are:

Pixel Replication: To zoom in and Zoom out the image.

ROI: The user can choose his region of interest and work on the portion of the image.

Shrink: The image can be shrinked to a finite extent as desired by the user. This tool is an added ambience apart from the zoom tool

High Pass and Low Pass Filtering: This provides the function of removing noise disturbances in the image.

Different users want different perceptions to the source image. Hence the User Interface is designed in such a way that it accommodates the interests of all the users of the system.

CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

About the organization

Siemens information system

Siemens in India, which comprises 11 separate legal entities, is active in the areas of Information and Communications, Automation and Control, Posiemensr, Transportation, Medical and Lighting. The flagship, Siemens Ltd., is a publicly listed company in which Siemens AG holds a 54.6% stake. The remaining equity is held by some 43,790 Indian shareholders.

In fiscal 2002, Siemens in India posted total sales of 812 million euros. New orders totaled 677 million euros. The Siemens Group in India currently provides direct employment to over 9,000 people.

Siemens Principles

Strengthen CUSTOMERS - to keep them competitive
Siemens success depends on the success of Siemens customers. Siemens provide Siemens customers with Siemens comprehensive experience and solutions so they can achieve their objectives fast and effectively.

Push INNOVATION - to shape the future
Innovation is Siemens lifeblood, around the globe and around the clock. Siemens turn Siemens people's imagination and best practices into successful technologies and products. Creativity and experience keep us at the cutting edge.

Enhance company VALUE - to open up new opportunities

Siemens generate profitable growth to ensure sustainable success. Siemens leverage Siemens balanced business portfolio, Siemens business excellence and synergies across all segments and regions. This makes us a premium investment for Siemens shareholders.

Emposiemensr Siemens PEOPLE – to achieve world-class performance

Siemens employees are the keys to success. Siemens work together as a global network of knowledge and learning. Siemens corporate culture is defined by diversity, by open dialogue and mutual respect, and by clear goals and decisive leadership.

Company Structure

Information and Communications

Information and Communication Networks (ICN)
Thomas Ganswindt; Michael Kutschenreuter, Andy W. Mattes, Anton Hendrik Schaaf
Information and Communication Mobile (ICM)
Rudi Lamprecht; Joe Kaeser, Lothar Pauly
Siemens Business Services GmbH & Co. OHG (SBS)
Paul A. Stodden; Jürgen Frischmuth, Bernd Regendantz

Automation and Control

Automation and Drives (A&D)

Helmut Gierse; Anton S. Huber, Alfred Ötsch

Industrial Solutions and Services (I&S)

Joergen Ole Haslestad; Bernd Euler, Joachim Möller, Dr. rer. oec. Udo N. Wagner

Siemens Dematic AG (SD)

Dr. Dietmar Straub; Dr. Peter Drexel, Alfred Frank, Pete Metros

Siemens Building Technologies AG (SBT)

Oskar K. Ronner; Hubert Ovenhausen, Rolf Renz

Transportation

Transportation Systems (TS)

Hans M. Schabert; Hans-Dieter Bott, Friedrich Smaxwil

Siemens VDO Automotive AG (SV)

Wolfgang Dehen; Dr.-Ing. Klaus Egger, Günter Hauptmann, Johann Löttner

Medical

Medical Solutions (Med)

Prof. Dr.-Ing. Erich R. Reinhardt; Dr. phil. nat. Hermann Requardt, Götz Steinhardt

Siemens Financial Services GmbH (SFS)

Dr. rer. Nat. Herbert Lohneiß

Corporate Finance (CF)

Heinz-Joachim Neubürger; Charles Herlinger, Karl Heinz

Midunsky, Dr. jur. Albrecht Schäfer

Prof. Peter Pribilla; Günther G. Goth

Prof. Dr. phil. Dr.-Ing. E.h. Claus Siemensyrich; Prof. Dr.-Ing. Winfried Büttner

Dr. jur. Dr.-Ing. E.h. Heinrich v. Pierer; Johannes Feldmayer, Pamela Knapp, Volkart P. Matthäus

Technologies for Tomorrow

As Siemens President and CEO Dr. Heinrich v. Pierer once said, "The surest way to predict the future is to create and shape it your self."

For Siemens, shaping the future is about developing technologies with a high-growth potential, recognizing technological breakthroughs, identifying customer needs, and creating new business opportunities, products and solutions. Planning Working in tandem with siemens operating units, Corporate Technology has developed different methods for preparing for and modeling the future. They are based on two opposing yet complementary perspectives: The one is obtained using extrapolations into the future based on the world of today; the other is obtained through retropolation back to the present, starting from the world of tomorrow.

Medical

Siemens' Medical area is renowned for its innovative products, services and complete solutions for the healthcare community. Siemens supply diagnostic and therapeutic systems as well as IT solutions for optimizing clinical workflow. Siemens offerings range from data management for healthcare providers to process management across the entire healthcare continuum.

CHAPTER 1

PROBLEM

STATEMENT

1.Problem Statement- A Brief Introduction

Cancer is a serious disease. *WHAT IS CANCER?* The word "cancer" actually refers to a group of over a hundred diseases. All forms are caused by the growth of abnormal cells within the body. Cancer cells are called "malignant" because they continue to grow and spread in a way that destroys normal parts of the body. Destructive masses called malignant tumors are formed. These tumors damage organs and may eventually spread (metastasize) to distant parts of the body. Cancers can originate in almost any cell of the body. Different types of cancer respond to different forms of treatment. Thus, it is crucial that cancer be diagnosed correctly. It is also important to determine the "grade" of a malignant tumor. This predicts the behavior or aggressiveness of the cancer, which is used to select the most effective form of treatment. Low-grade tumors grow and metastasize (spread) relatively slowly. High-grade tumors, in contrast, grow rapidly-- causing more destruction of normal tissue and tend to metastasize early.

Cancer diagnosis's is a task of paramount importance. Diagnosing cancer involves analyzing the cell at various stages. Cells undergo various stages of growth before maturing. A stage called the slight dysplasia held up the oncologist and the physicians in a dilemma situation. At this stage there is a probable situation, which might leave the cell cancerous or might resolve into the normal cell. Research has been into this stage and yet the results are not convincing. The problem is to find the cancer cells from the digital image. The objective of the project is to find the future development of the cell, whether it is going to be maligne or revert back to a normal cell. The cell is to be

analyzed at the intermediate level of its growth and conclusion has to be made. It is believed that processing the digital image of the cell yields a solution that effectively helps the cytologist and the pathologist to diagnose the disease with more accuracy.

1.1.1 System Justification

Today computerization covers most of the existing domain. Automating the system has seen a vital growth in the field of Medical science. Image processing is one such area where processing the digital image helps in one of the important criteria where a solution is sought, the *decision making*. The same applies here, where a set of digital image of the cell are analyzed for making a conclusion on the growth of the cell.

1.1.2 Goals of the System

A typical Client application system has to do the following

- Minimize maximum human intervention with the system
- Effective GUI

Minimizing the human intervention could be seen in all quality guaranteed software products. Software is developed to lessen the manual work that is been adapted for several years. GUI (graphical user interface) provides the user with options to be performed on the task. In-fact this feature makes the product efficient and effective.

1.2 System Constraints

Every system developed has a constraint in its perspective. The constraints encountered in first copy of the developed system might be resolved in the development of the second version of the system with more options to spare. In that way the constraints in the developing system are few to be quoted as follows

- Loading of files other than GIF and JPEG/JPG
- MDI not supported
- Processing of certain cell types

The developing system supports the loading of GIF and JPG image files. This is because the sources of the cell image are from a microscope with a limited resolution. The images are directly converted into a .gif or .jpeg file type. Conversion of the image to a BMP type might loose some characteristics of the image.

With a view that the user tends to process only one cell image, the system is aimed at proving a SDI (single document interface). Hence opening of multiple images is restricted.

Cells differ from region of extraction. All cells are not the same they differ in their size and shape. Developing of a system that could process all type of cells and predict possible traces of cancer is a time consuming process in the field of development of the system. Hence the system is developed with a limited scope of processing.

The *Programmers constraints* include:

As it is an on-going research activity where a desired solution is yet to be experienced, thorough understanding of the problem and the growth of the cell have to be understood. Clear knowledge of the development of the cell and its activities has to be studied before making a step of progress.

1.3 Processing Environment

The processing environment for the proposed system is listed on the basis of existing system.

1.3.1 Hardware environment

The detailed hardware specifications listed below is the environment of the developing system.

Processor	INTEL Pentium II
RAM	128 MB
HD Capacity	6.4 GB
Disk Drives	GENERIC IDE DISK TYPE 46
	GENERIC NEC FLOOPY DISK
Display adapters	Cirrus Logic 5446PCI
	S3 Graphics PRO SavageDDR

1.3.2 Software environment

To make the system portable and give the user the best of the kind GUI, the system is developed using

Language   Java (Swing)

C

Java Swing comes with more upgraded user interface components. The use of "C" might be needed in certain areas of development. It does not raise the issue of non-portability because C is compatible with most of the well-known operating system.

The use of Java also minimizes the development cost because it is Free-Ware version, which could be downloaded from the net.

1.4 User Characteristics

The end users intended to use the system are

- o Pathologist
- o Cytologist
- o Physicians

The pathologist and the cytologist can make a conclusion from the result after the system diagnosis the image would be highly useful for them in reporting. The Physician who is an expert in Cancer could also make use of the system. Since this is a typical client side application the user needs to posses the knowledge of operating the system. The interface provided helps in the easy navigation of the process.

The user is restricted from the option of processing on the images that are not supported by the system.

1.5 Acceptance Criteria

The main criterion is speed. The next is accuracy and all other quality factors. The system needs to be scalable enough to do the analysis on the cell image of different size. The performance must meet up the expected through put rate. The required characteristics for acceptance criteria are

- Portability
- Reliability
- Efficiency
- Accuracy/correctness

The *portability* is a factor to be considered for the system to be compatible enough to work. The portability factor accounts for ·all up coming products. It increases the system reliability. In a way reliability of the system in someway is concerned with portability factor. *Reliability,* the word that holds up the quality of the software, in-fact a **de-facto** for all system should give the user a desired output and not an unlikely one. System reliability depends on the algorithms used. Complex algorithms make the system less reliable.

Reliability and portability gives an *efficient* system. If the system is efficient then the user can be more reliable on the system. *Accuracy/Correctness* is the main issue to be considered in the place of the quality of the system. Quality systems produce accurate result.

1.6 Sample Input

Few images of the cells are shown:

Superficial sq cells
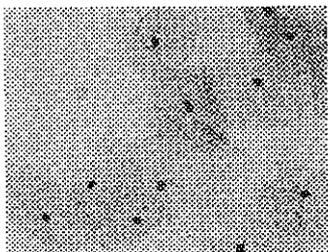
Intermediate squamous cells

Parabasal cells
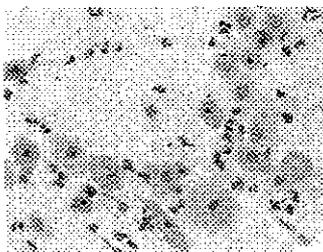
Slight dysplasia
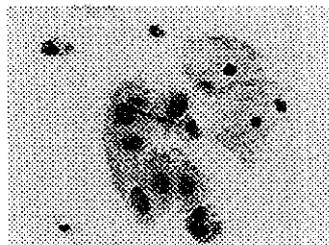
Severe Dysplasia

1.7 Sources of Information

- Problem details   *Client*   *(name undisclosed)*
- Data source   *Client*   *(name undisclosed)*
- Details on Cancer growth *Client*   *(name undisclosed)*

                          www.theprostatecenter.com

                          Cancer Institute

1.8 Glossary of Terms

Cytologist

Person who deals with the study on cells

Maligne

Cancerous cells are termed maligne because they affect other cells also

Moderate Dysplasia

A state where the cells are advanced to a cancerous stage

Pathologist

Person who studies on the behavior of the specified disease.

Severe Dysplasia

At this stage the cells are thoroughly affected by cancer

Slight Dysplasia

This is the state where the research is on process. At this stage the cells have an equal probability to advance to cancer stage or revert back to normal one.

CHAPTER 2

PROJECT PLAN

2.1 Development Phase

The development stage for this project includes the following

- Analysis phase
- Learning Image Processing concepts
- Design phase
- Review
- Implementation phase
- Testing phase

At the initial stage before discussion of the project with the clients a different dimension of the problem was assumed. After the meeting with the clients the problem known is to give a solution on the Slight Dysplasia. It was suggested as a requirement that the system to be developed needed to be portable and give the user the best of the GUI. Since the developers are much known to Java Environment and the upgraded components in Swing Java was taken up for the development of the product.

Analysis stage brought light the need of the minimizing maximum human intervention. After the analysis phase learning the concepts on image processing is an essential one as the project is purely based on the processing of the digital images of the cell, which are the data source. Design phase is the core of any successful system being developed. With good design the product increases in quality.

After every phase is completed it should be accounted for Review. It is a cautious measure to avoid any problem if encountered in the later stage of the development. Reviewing

the phases after completion brings the minute errors if made any. In the same way the system after being designed was subjected to design review where an error in the design was visualized and corrected.

Implementation phase brings the design into life. Here the implementation of the system is on process. After implementation the system has to be tested for portability issues, correctness of the process and other factors relevant to quality factor.

2.2 Development schedule

Analysis Phase	I.P Concepts	Design Phase	Implement	Testing
Req& Planning		Verify		
Verify			Code Debugging	

PR   SRR   PDR   CDR   SCR   PTR
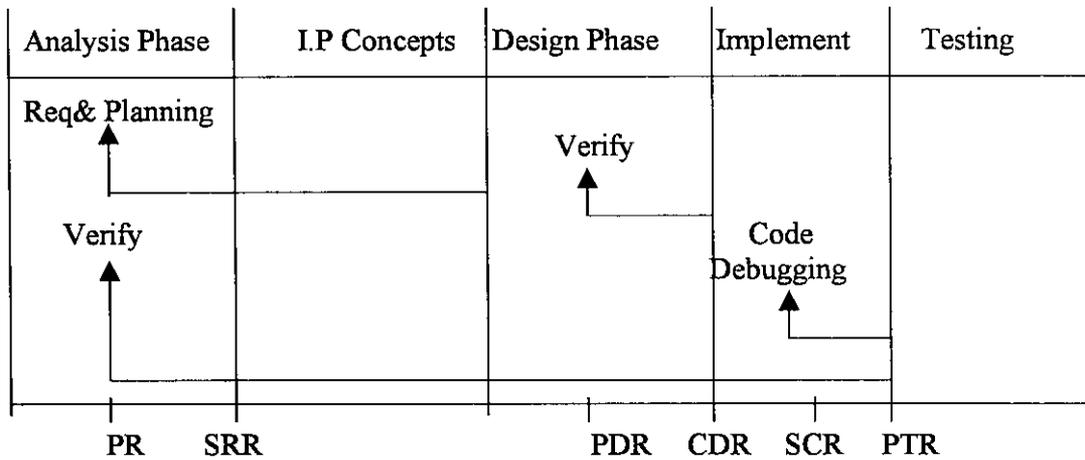
REVIEW	
PR	Problem review
SRR	Software requirements review
PDR	Preliminary design review
CDR	Critical design review
SCR	Source code review

2.3 Monthly Schedule

ACTIVITES	JAN	FEB	MARCH	APRIL
ANALYSIS/REQ				
PLANING				
LEARNING				
DESIGN				
*IMPLEMENT				
* TESTING				
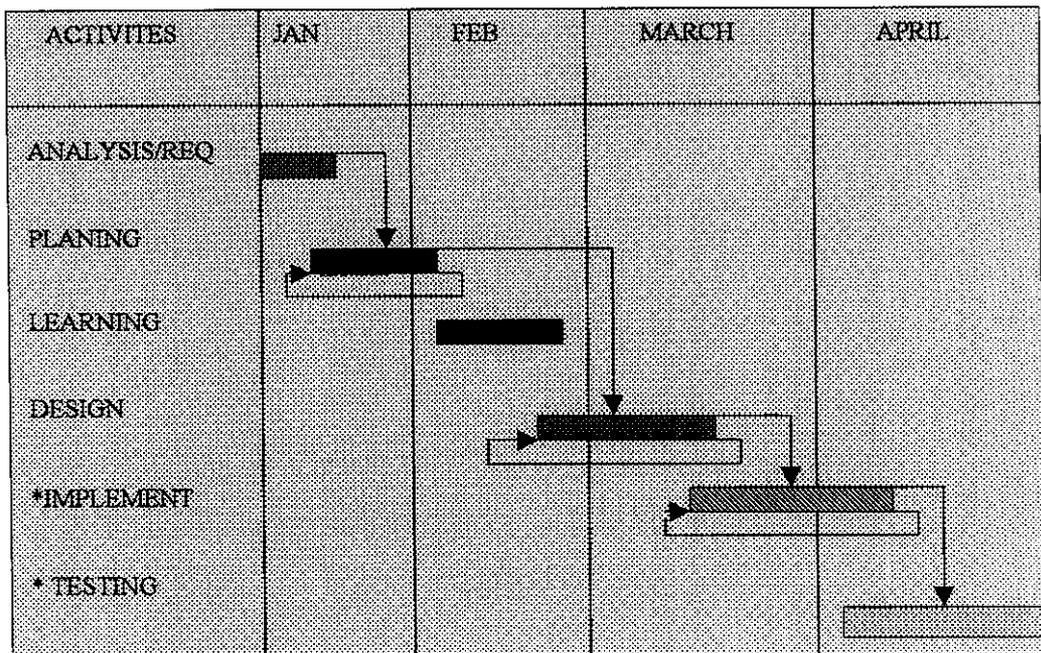
* Subjected to change on date

2.4 Programming Languages

The programming language used is Java and C. For the GUI components in Swing are used. The operating system used is Windows.

2.4.1 Java Swing

Java is fully object-oriented language with strong support for proper engineering technique. It is difficult to write so-called procedures in Java. One must create and manipulate objects. Java program consists of pieces called classes. Classes consists pieces called methods that performs tasks and return information when they complete their task. One can program each piece need to forma java program, but most java programmers takes an advantage of rich collections of existing classes in java class libraries.

The class libraries are also called as java API's (application programming interface). Java systems consist of several parts: the environment, the language, API and various class libraries. Java programs normally goes to five steps in executing a program: edit, compile, load, verify and execute.

Typical Java Environment
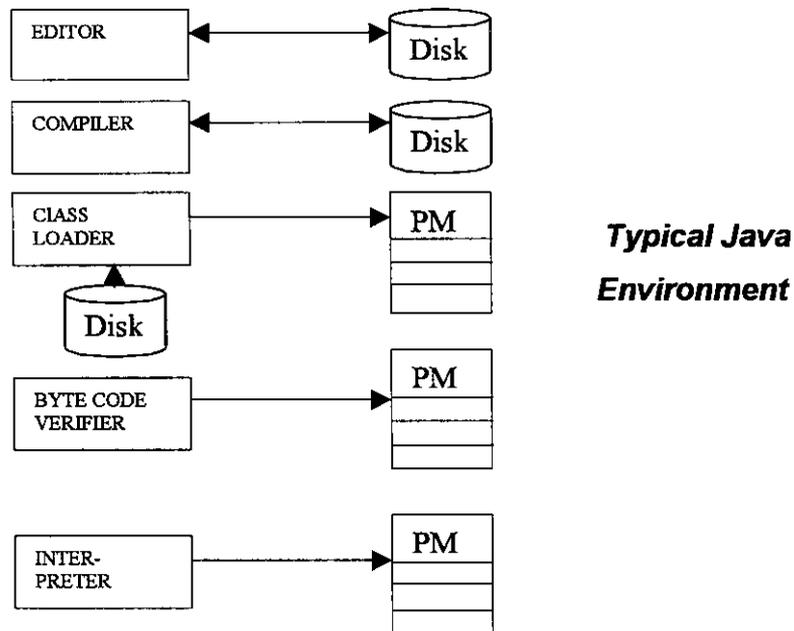
Java programs come in two main types: application and applets. Applets are java programs one can download and run in web browser. Besides downloadable applets java also supports application designed to be run on local machines. Java application work like other computer application.

2.4.2 Security

When it executes a program, JVM can strictly monitor what's going on, which makes it great for the Internet application. The JVM can watch all that a program does and if it does something questionable it then prevents that process. One can tailor the JVM security the way one like it.

2.4.3 SWING

The AWT is one of the powerful toolkit when introduced and it was original driving force behind java popularity. Now that SWING which has four times the number of user interface components ads the AWT, is a part of the standard java distribution, it's becomes fashionable to deride AWT. Swing is a set off packages build on the top of the AWT. In swing the UI components begins with J and few of them are listed

- JApplet
- JButton
- JCheckbox
- JComponent
- JEditorPane
- JFrame
- JDesktopPane
- JColorChooser... *Etc*

All swing components are derived from Jcomponent class and this class is, in turn, derived from the AWT container class, which has no heavy weight window (called a peer). Besides the large array components in swings and the fact that they are lightweight, swing introduces many other innovations. Some of the major one are

1. Borders-one can draw borders in different styles
2. Graphics debugging-one can use the method in it to set up graphic debugging that can watch each line as it draws.

3. Easy mouseless operations-to connects keystrokes to components.

4. Tool tips

5. Easy scrolling

6. Pluggable look and fell

7. New layout managers.

2.5 Manner of Demonstration

Presentation is nothing more than outlines. Outlines are nothing more than sentences. Once an outline is created, PowerPoint let's user generate slides shows, handouts, note pages etc from it. This is a great feature, for if an outline was created in a word processor, handouts and the like would have to either be cut or pasted from the outlines, or the formatting of the information would have to be change. Power point streams these tasks for the user once a presentation or outline is created. Power point is a complete presentation graphics package. It gives you everything you need to produce a professional-looking presentation. PowerPoint offers word processing, out lining drawing, graphing and presentation management tools- all designed to be easy to use and learn. With its easy tool for creating professional presentation PowerPoint long ago transformed the way people communicate ideas. The various presentations are done in the aid of power point and a multimedia projector was used for demonstration to a large gathering. Power point has the capability to create complete presentation that include

- Slides
- Outlines
- Speaker notes
- Audience handouts
    - Each one generated by creating any of the other. PowerPoint has nearly unlimited multi-media capabilities and the only limitation are you and the computer.

2.6 Glossary of terms

AWT
> *Abstract windowing toolkit. Provides the foundation for graphics working java.*

Compiler
> *Program that compiles the program for errors.*

Interpreter
> *Program that analyze and executes a program line by line.*

JVM
> *Java Virtual Machine, support in running java program in all platforms.*

PowerPoint
> *Power point is a complete presentation graphics package. It gives you everything you need to produce a professional-looking presentation.*

Swing
> *Toolkit Provided by Java for building UI.*

CHAPTER 3

SOFTWARE REQUIREMENTS

AND SPECIFICATION

3.1 Software Requirements and Specification-

An overview

The system is developed for a client in India, involved in Cancer Research for the last 16 years. The data provided is in the form of digital images. The task identified is to scan the image, which is a cell, and to give a report on the future enhancement of the cell, which is believed to help the cytologist and the pathologist in making a decision on the type of treatment to be preceded. The project is carried out in Siemens information system, Bangalore. Further reading of the document provided you the best of the knowledge in

- Software specification
- Hardware specification
- Performance specification
- System model template

3.2 Software specification

3.2.1 The Platform

The platform under which the system is developed is Windows version 98. Being one of the most popular Operating system available today, gives the user the touch and look making the application developed under its base graceful. Windows helps the user to navigate and explore throughout the system in friendly fashion.

3.2.2 Language

Java was discussed much in the section Programming Languages. To have continuity on the discussion on let have a brush up with the topic on Java with few points

- Java is fully object-oriented language.
- Classes consists pieces called methods that performs tasks and return information when they complete their task.
- The class libraries are also called as java API's.
- Java systems consist of several parts: the environment, the language, API and various class libraries.
- Java programs normally goes to five steps in executing a program: edit, compile, load, verify and execute.

Swing

All swing components are derived from JComponent class. Besides the large array components in swings and the fact that they are lightweight, swing introduces many other innovations. Some of the major ones are

- Lightweight component
- Borders Graphics debugging
- Easy mouseless operations
- Tool tips
- Easy scrolling
- Pluggable look and fell

C

C, one of the most powerful languages is a popular programming language among the programmers. C is a versatile language for handling large systems. It is virtually suitable for most of the programming tasks including

- Editors
- Compilers
- Databases
- Communication systems and
- Complex real life application systems.

Pointers concepts are the true essence of C programming language.

3.3 Hardware Specification

The hardware specification listed is on minimum basis for optimum performance.

Processor	Genuine Intel Pentium II
Frequency	833 MHZ
RAM	64 MB
HD Capacity	6 MB
Cache	512 KB
Video Memory	1 MB RAM
FD	3 $^{1/2}$ inch FD

3.4 Process Specification

The various methodologies handled are

3.4.1 Convolution Matrix

It is a matrix, which results in a convolution matrix for a set of 3 X 3 matrix. Common 3X3 convolution matrixes are in the form

$$
\begin{pmatrix} 1 & 1 & 1 \\ 1 & 8 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \& \quad \begin{pmatrix} -1 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & -1 \end{pmatrix}
$$

The values of the pixels are multiplied with this convolution matrix to form a resultant matrix, which will be the desired result after applying any filter option with the corresponding algorithm in action. In a convolution, the calculation performed is a weighed sum of gray levels from a neighborhood-surrounding pixel. The neighborhood includes the pixel under consideration. Grey levels from the neighborhood are weighted by coefficients that come from a matrix or convolution kernel. Usually the kernel is fairly small relative to the image-dimension of 3 X 3 are the most common. During convolution we take each kernel coefficients in turn and multiply it by a value from the neighborhood of the image lying under the kernel.

3.4.2 Edge Detection- by mathematical quantity

The simplest detectors perform minimal noise smoothing and fairly crude localization. They are based on the estimation of gray level gradient at a pixel. The gradient, may be approximated in the x and y directions by

$$G_x(x,y)=f(x+1,y)-f(x-1,y)$$
$$G_x(x,\ y)=f(x,y+1)-f(x,y-1)$$

Small amount of noise smoothing can be introduced if we compute averages of these gradients over a 3 X 3 neighborhood. This allows us to express gradient calculations as a pair of convolution operation where the kernels are

$$Hx=\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \qquad Hy=\begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

3.5 Performance specification

The required *qualitative* performance measures are

- High system efficiency
- Through put rate
- Less execution time
- Scalability
- Modularity

The required quantitative performance specifications are

- Relative efficiency and
- Relative speed

This is related by formulae

$$E_{relative} = \frac{T_1}{P}$$

Where T1 is the execution time on processor P

The related quality speedup is

$$S_{Relative} = PE_1$$

The accuracy of the result should reach the desired expectation because we are dealing with the processing of the images, manipulating with the pixel values. The overall execution time excluding used elapsed time is

$$T = T^J_{Comp} + T^J_{Comm} + T^J_{Idle}$$

Where

T_{Comp} **is computation time**

T_{Comm} **is communication time**

T_{Idle} **is waiting or ideal time**

3.6 System Model Template

USER INTERCACE PROCESSING (GUI)		
INPUT PROCESS (Image of the Cell)	PROCESS CONTROL AND FUNCTIONS MAINTENANCE AND TEST *(Yet to be covered)*	OUTPUT (Diagnosed report)
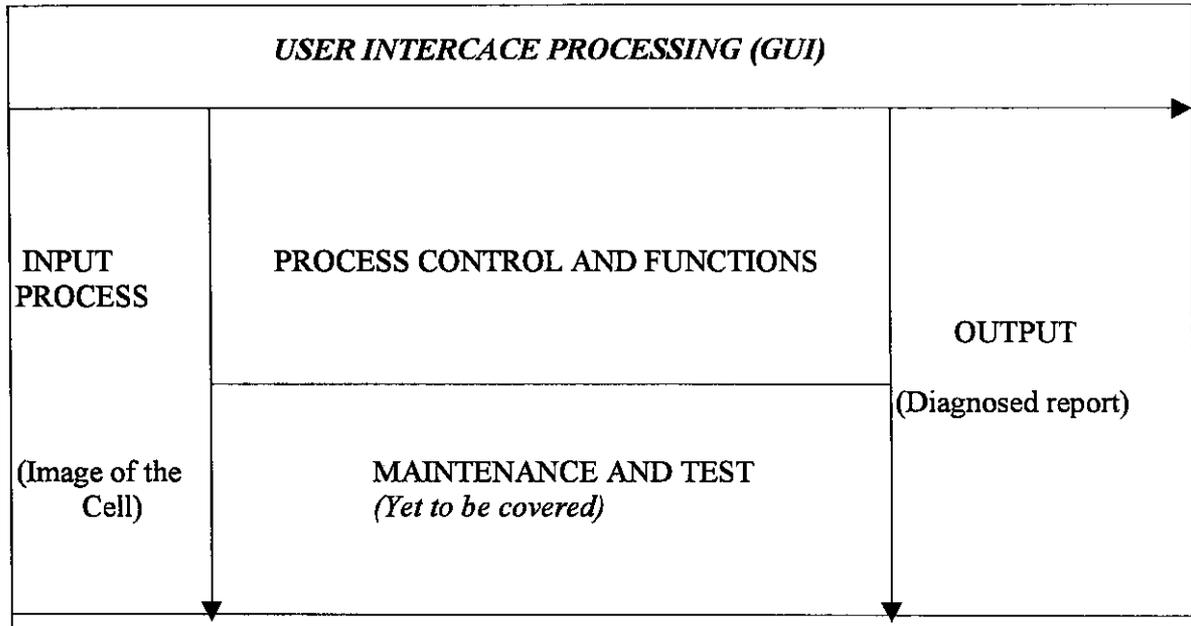
3.7 Glossary of terms

Execution time

> *The time taken by a processor to execute its code.*

Modularity

> *The decomposition of the problem into several subtasks*

Portability

> *The ability of the system to work on different platforms*

Scalability

> *The ability of the system to adapt to increase or decrease on size or memory*

System efficiency

> *The ability of the system to make best use of the available resource*

Throughput rate

> *The frequency of execution given in second base on performance hits and misses*

CHAPTER 4

DESIGN

DOCUMENT

4.1 Architectural design

```
                              ┌──────────────┐
                              │   FILTERS    │
                              └──────┬───────┘
            ┌────────────────┬──────┴───────┬──────────────────┐
     ┌──────┴──────┐  ┌──────┴──────┐  ┌─────┴───────┐  ┌───────┴──────┐
     │  FILTER1    │  │  FILTER2    │  │  FILTER3    │  │   FILTER4    │
     │   F.Ed      │  │   G.Sc      │  │   SHARP     │  │    E.E       │
     └──────┬──────┘  └─────────────┘  └─────────────┘  └──────────────┘

         ┌──────┐
    ───► │ CED  │
         └──────┘                ┌────────────┐        ┌──────────────┐
                                 │   CELL     │ ─────► │  DIAGONISED  │
         ┌──────┐                │   IMAGE    │        │   REPORT     │
    ───► │ SED  │                └────────────┘        └──────────────┘
         └──────┘

      ┌──────────────┐                    ┌──────────────┐
      │  IMG.OPTION  │                    │  DIAGNOSIS   │
      └──────────────┘                    └──────────────┘
```
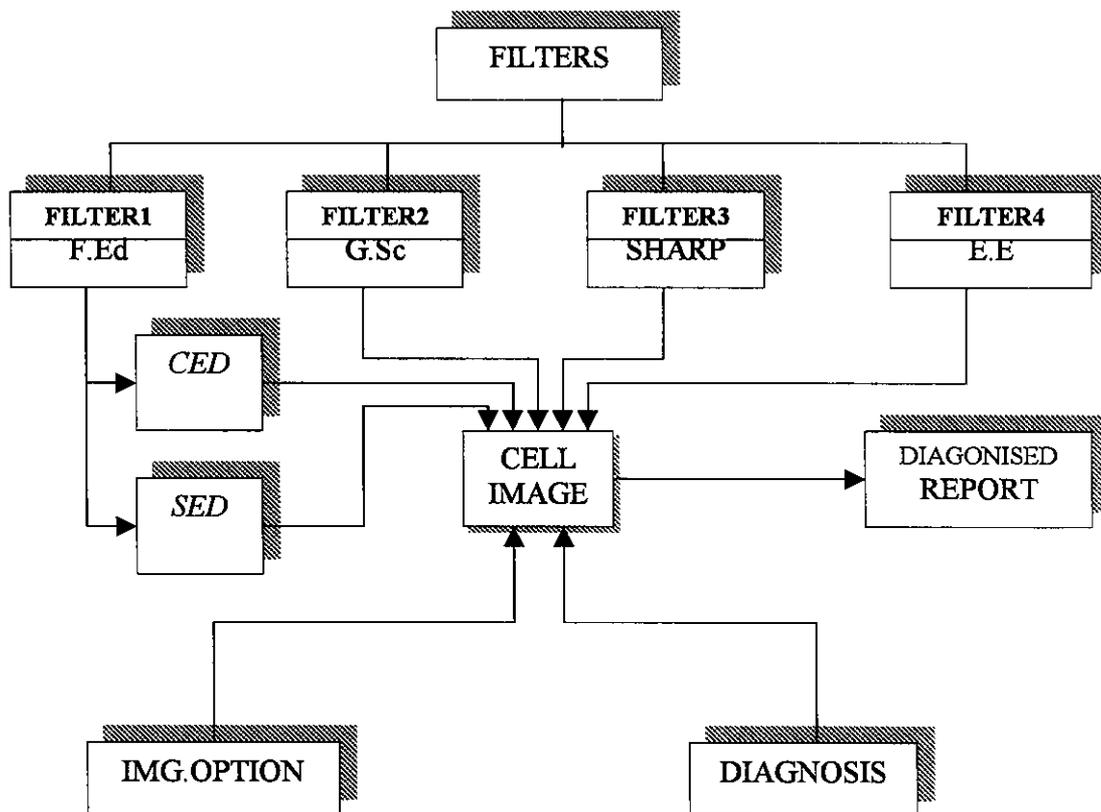
4.2 Data flow diagram
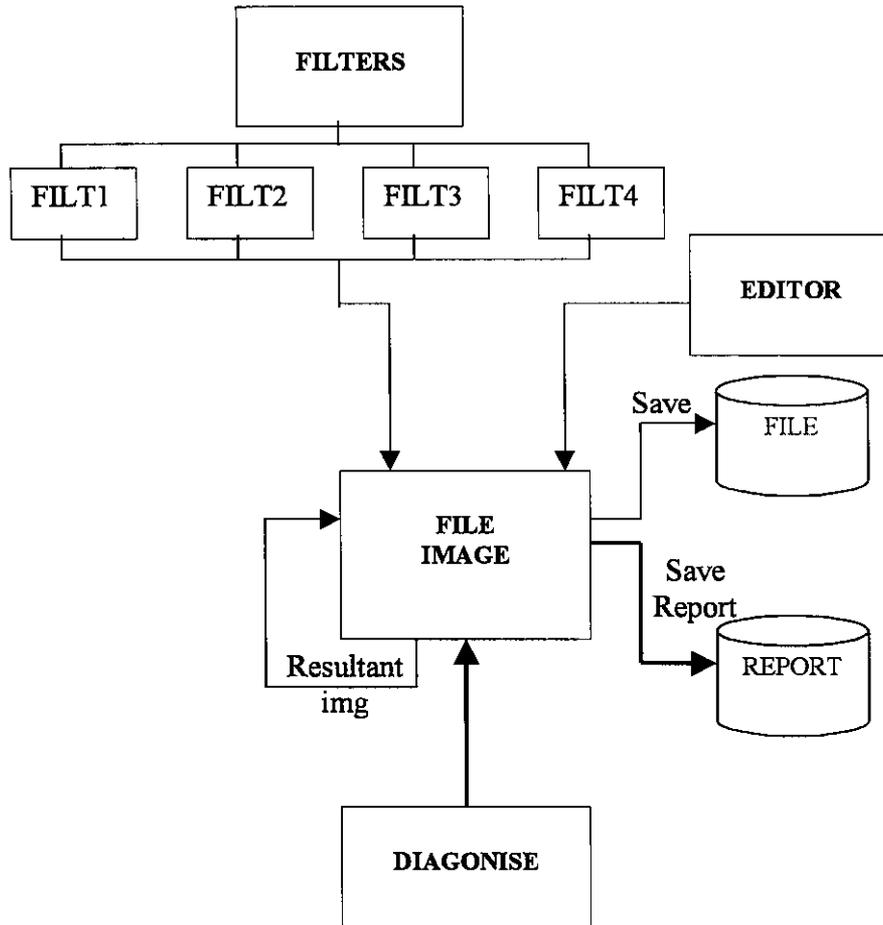
4.3 Structured chart

The structured chart gives you a typical representation of what's happening in the system. For the convenience of the readers and in the view of increasing the readability of the readers the structured chart is shrinked to the maximum without showing any complexity within the modules. Beside the structured chart labels are given to show the task the operations that shall be performed by the user. The labels are not prioritized and the sequence shall flow depending upon the operation the user wishes to perform on the cell image. The end task will be the final task that would be performed by the system and would yield the result of the filtered image or the report for the diagnosed image, which would enable to make a decision.

TASK A

A- Open file

TASK B		TASK C

B- Apply filters

C- Diagnose Image

END TASK

D- Result/Report
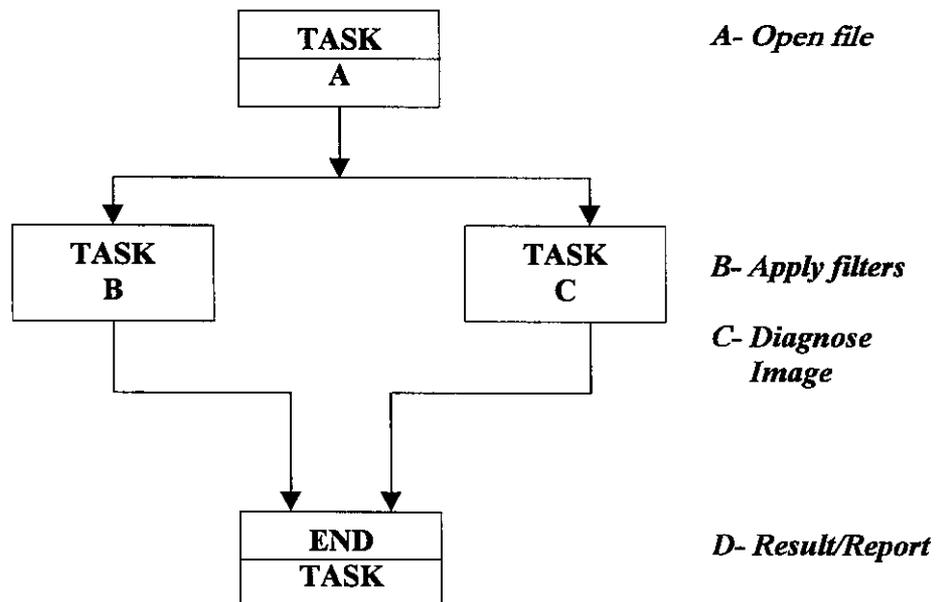
4.4 Detailed design

Detailed design is concerned with specifying the algorithmic details, concrete data representation, interconnection within the function and data structures and the packaging of the software products. Detailed design is strongly influenced by the implementation language; it is not the same as the implementation. Detailed design is more concerned with semantic issues and it isles concerned with the syntactic details than its implementation.

The starting point for the detailed design is an architectural structure or which algorithmic details and concrete data representation are to be provided. The section ahead of the detailed design gives the

- Pseudo code
- Data structure
- File representation

for this project on Cancer Diagnosis.

4.5 Pseudo Code

The pseudo codes for very few modules of the system are listed below.

4.5.1 Calculating the ROI

```
Double meanvalue (BufferedImage image)
{
 Raster raster=image.getRaster ();
 Double sum=0.0;
      For (int y=0;y<image.getHeight (), y++)
      {
              For (int x=0;x<image.getWidth (); x++)
              {
              Sum+=raster.getSample (x, y, 0);
      Return sum/(image.getWidth ()*image.getHeigth ();
      }
      }
}

Double meanvalues (BufferedImage image, Rectangle ROI)
{
Return meanValue (image.getSubimage(ROI.x,ROI.y,
                     ROI.width, ROI.height);
}
```

4.5.2 Pixel Replication

```
Double meanvalue (BufferedImage image, int n)
{
int w=n*image.getWidth();
int h=n*image.getHeigth();
BufferedImage enlargedimage=new BufferedImage(w,h,
                                image.getType());

For (int y=0;y<h, y++)
     {
             For (int x=0;x<w; x++)
             {
             enlargedimage.setRGB(x,y,image.getRGB(x/n,y/
                            n));
             Return enlargedimage;
```

```
                            }
                        }
                }
```

4.5.3 Shrink by Skipping Pixels

```
        Double meanvalue (BufferedImage image, int n)
        {
        int w=n*image.getWidth();
        int h=n*image.getHeigth();
        BufferedImage shimg=new BufferedImage(w,h,
                                        image.getType());

        For (int y=0;y<h, y++)
            {
                    For (int x=0;x<w; x++)
                    {
                    shimg.setRGB(x,y,image.getRGB(x*n,y* n));
                    Return shimg;
                    }
            }
}
```

4.5.4 Convolution for single pixel

```
Create an array h,indexed from0 to m-1 horizontally and 0 to
n-1 vertically
Fill h with kernel coefficients
M2=m/2
N2=n/2
Sum=0
        For k=-n2 to n2 do
                For j=-m2 to m2 do
                        Sum=sum+h (j+m2, j+m2) f (x-j, y-k)
                End for
        End for
G (x, y)=sum
```

4.5.5 Error Diffusion

```
Threshold=(black+white)/2
```

46

For all x and y do

    If f(x,y)<threshold then

        G(x,y)=black

        E(x,y)=f(x,Y)-balck

    Else

        G(x,y)=white

        E(x,y)=f(x,y)-white

    Endif

F(x+1,y)=f(x+1,y) +7E/16

F(x-1,y+1)=f(x-1,y+1) +3E/16

F(x,y+1)=f(x,y+1) +5E/16

F(x+1,y+1)=f(x+1,y+1)+E/16

4.5.6 Find edges

Algorithm *The idea is to look the surroundings of the pixel being processed:*

•if the surroundings are very different than the pixel, then we're at an edge, a bright pixel must be outputted.

•If the surroundings are as bright as the pixel, then a dark pixel is the wanted result.

    Pseudo-code:

    for each pixel in the destination image do

    if the source pixel differs much from its surroundings

        then the resulting pixel is bright

        else the resulting pixel is dark

4.6 Data Dictionary

Class	Extended/ Implements	Components	Methods	Constructor
Tmh	Jframe	JButton	Accept	Tmh
List		JMenuBar	getDescription	
Giff		JMenuItem		
Jpgf		JSplitPane		
Edt		JToolBar		
Filt		JComboBox		
Eedge		JFileChooser		
Ehan		JTextField		
Gscale		JScrollBar		
Ced				
Sed				
Hlp				
Roi				
Scale				
Shrks				
Elrge				
Ct				
Cpy				
Pste				

*The table contents might subject to change

4.7 Glossary

Data Structure

One of the basic structure to improve algorithm is to structure the data in such a way that the resulting operations can be efficiently carried out.

Pseudo Code

Pseudo code is simply a numbered list of instruction to perform some task. In this course we will enforce three standards for good pseudo code

- *Number each instruction*

- *Each instruction should be unambiguous.*

ROI

Region of interest - where the user wishes to process a particular area within the image.

CHAPTER 5

TESTING

5.1 Testing

Testing is a vital process to the success of any system. The system as a whole will be tested for the following

- Consistency with the application
- Displaying the Resultant image
- Integrity of the modules
- Referential integrity test

System testing makes a logical assumption that if all the parts of the system are correct, the system will be successfully achieved. The objective of testing is to discover errors. To fulfill these objectives a series of test is planned to execute. Software testing can be looked upon as one among many processes. This is the last opportunity to correct any possible flaws in the developed system. Software testing includes selection test data that have more probability of finding errors.

Systems are the stage of implementation that is aimed at ensuring that the system works accurately and efficiently before live operation commences. In principle, system proving is an on-going activity throughout the project.

The first step-in system testing is to develop a plan that tests all the aspects of the system. Completeness, correctness, reliability and maintainability of the software are to be tested for the best quality assurance that the system meets the specification and the requirements for its intended users and performance. System testing is most useful practical process of executing a program with explicit

intention of finding errors that make the program fail. The following phases are developed for the purpose of testing the system.

5.1.1 Module Testing

Each individual program module will test for any possible errors. They will be tested for specification i.e., to see whether they work as per what the program should do and how it should perform under various conditions.

5.1.2 Event testing

Each item present in the menu will be tested to check whether the appropriate actions are triggered. This in a way helps in checking the functionality of menu items.

5.1.3 Integration testing

This is done to check whether the modules present in the developing system are interconnected as per the requirements. This sort of checking helps in analyzing the behavior of the program when subjected to perform certain functions.

5.1.4 Display Testing

In fact this test in the effective test to be carried out. The display procedures are tested since the data displayed is of much importance. The consistency of the display and the attractiveness of the display were also tested.

A testing is an examination with the intent of finding the errors. Concentration was more on errors than resting on the glory of apparently perfect outputs. In the case of *Cancer Diagnosis,* two types of test were mainly conducted namely

- Unit testing and

- Integrated testing

Integrated Program Diagram
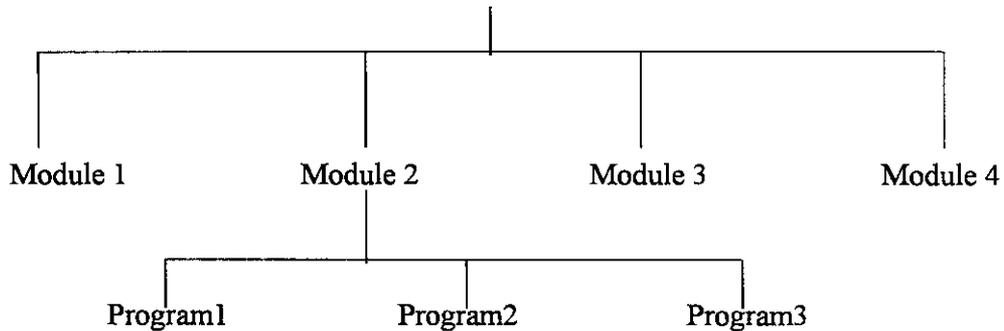
In unit testing each individual program is tested and would be checked whether it satisfies with the require output. As an illustration in the project, if grayscale filter is applied then the resultant image should be of grayscale image filtering the image from the original RGB specifications values. The individual programs are the combined to form modules. Integrity test will be performed on each of the modules and again the validity will be checked. After all the modules are brought under single module integrity test is performed and the result should yield success.

5.2 Performance Test- Performance Analysis

What is it?

The given design or code must make efficient use of the resources and give optimum output with effective throughput. This is one of the prerequisites of any good code. The methods employed to analyze and study the performance of a good code is called performance analysis.

The performance analysis is primarily dealt with two factors. They are

- Time complexity
- Space complexity

Performance analysis is done through data collection

Data collection is the process by which data about program performance are obtained from an executing program. Data are normally collected in a file either during or after execution, although in some situations it may be presented to the user in real-time. Three basic data collection techniques can be distinguished, namely

Profiles, record the amount of time spent in different parts of a program. This information, though minimal, is often invaluable for highlighting performance problems. Profiles typically are gathered automatically.

Counters, record either frequencies of events or cumulative times. The insertion of counters may require some programmer intervention.

Event traces, records each occurrence of various specified events, thus typically producing a large amount of data. Traces can be produced either automatically or with programmer intervention.

The raw data produced by profiles, counters or traces are rarely in the form required to answer performance questions. Hence, data transformations are applied, often with the goal of reducing total data volume. Transformations can be used to determine mean values or higher-order statistics or to extract profile and counter data from traces.

The concept of profiles should be familiar from sequential computing. The information is often obtained by sampling techniques, which are simple but are nit necessarily accurate. The values of the program counter is determined at fixed intervals and used to construct a histogram of execution frequencies. These are then combined with the complier symbol table information to estimate the amount of time spent in different parts of the program. Profiles have two important advantages. They can be obtained automatically, at relatively low cost and they can provide a high level view of program behavior that allows the programmer to identify the problematic program components without generating huge amount of data. Profiles also have some limitations. In particular, they don't incorporate temporal aspects of program execution.

Counters can be used to record the number of procedure calls and other relevant details pertaining to the problem. Counts

might be generated by Compliers generated code by code incorporated by user-inserted calls to counter routines. Counters complements profilers by providing information that's is not easily obtainable using sampling techniques.

A useful variant of the counter is an interval timer; a timer used to determine the length of time spent executing a particular piece of code. This information can be accumulated in a counter to provide an accurate determination of the total spent executing that program components. A disadvantage of interval timers is that logic required to obtain a timer values can be expensive.

With regard to the developing system this profile and counter will do a good work for performance analysis.

5.3 Glossary of terms

Integrity test

Tests the integration of the modules or programs

Modules

Sub programs or sub divided tasks

Performance Test

Test to check the performance of the system

Space complexity

Complexities arising due to memory space available

Time complexity

Complexities arising due to time management of the process

Unit test

Test to check the proper functioning of the programs present in the modules.

6 Future Enhancements

There is no barrier for innovation, requirements changes on constant basis, new era of technology comes to focus, these factors influences on the enhancement of any software project.

The developing system focuses on service to mankind. The research on cancer has been on the way thru 16 years and still on its way. On a broader scale to list about the future enhancements, the system should be able to detect the cells associated with the cancer cells, its rate of infection and its growth properties. The present system accounts for a minimum scale of diagnosing. Future product should diagnose for all cells.

An executive breed software should be developed that could trace the cell activity in future when viewed through the microscope. The system should also focus on accessing all sorts of image files and should facilitate multi-user operation.

7 Conclusion

The requirements gathered for the development of this project was done with the guidance of many cytologist, pathologist and Cancer specialist. Meeting with the client helped greatly in knowing the problem than getting the information about the problem from the second source. Any minor flutter in the information collected would account for unlikely results. Study on the basics of cell development was carried out with care.

The concepts in the software engineering book were followed to the point during the design phase. Requirement gathering, requirement analysis were done on first hand basis. Design, a De-facto for any software products was carried out after gaining much experience thru stipulated period of learning. The phrase "Learn to Learn" made us realize the things that could be done by our own.

The document was aligned in a way that the readers would find it much comfortable to go through. Information about the project is given to the max. I had put the best of my effort to bring out the document in its best way as it is, hopefully.

The requirements of the project were satisfied. The scope for further development is innumerable and the possibilities of renovation are infinite.