

PROFESSIONAL TRAINING REPORT
at
SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Submitted in partial fulfillment of the requirements for the award of
Bachelor Engineering Degree in Computer Science and Engineering

By

KOSURI NAGA PREETHI (Roll. No. 17SCS0219)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING
SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119, TAMILNADU.

MARCH-2021



SATHYABAMA

**INSTITUTE OF SCIENCE AND
TECHNOLOGY (DEEMED TO BE
UNIVERSITY)**



**Accredited with Grade “A” by NAAC
(Established under section 3 of UGC act,1956)**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119**

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Professional Training Report is the bonafide work of **KOSURI NAGA PREETHI (Roll. No. 17SCS0219)** who underwent the professional training in “**BREAST CANCER DETECTION**” under our supervision from APRIL 2020 to JUNE 2020.

Internal Guide

Mrs.SELVI, B.E., M.E., (Ph.D.,)

Head of the Department

**Dr. S. VIGNESHWARI, M.E., (Ph.D.,)
Dr. L. Lakshmanan, M.E., (Ph.D.,)**

Submitted for Viva voce Examination held on _____

Internal Examiner
Examiner

External

DECLARATION

I, **KOSURI NAGA PREETHI (Roll. No. 17SCS0219)** hereby declare that the Professional training report entitled "**BREAST CANCER DETECTION**" done by me under the guidance **Mrs.SELVI B.E.,M.E., (Ph.D.,)** is submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.T.Sasikala M.E.,Ph.D., Dean, School of Computing, Dr.S.Vigneshwari M.E.,Ph.D., and Dr.L.Lakshmanan M.E., Ph.D., heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs.SELVI,B.E.,M.E.,(Ph.D.,)** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

BREAST CANCER DETECTION USING MACHINE LEARNING TECHNIQUES

ABSTRACT

Breast cancer is one of the most dangerous diseases and the second largest cause of female cancer death. Breast cancer starts when malignant, cancerous lumps start to grow from the breast cells. Self-tests and Periodic clinical checks help to early diagnosis and thereby improve the survival chances significantly. The breast cancer classification is a medical method that provides researchers and scientists with a great challenge. Neural networks have recently become a popular tool in cancer data classification. In this paper, Deep Learning assisted Efficient Ada boost Algorithm (DLA-EABA) for breast cancer detection has been mathematically proposed with advanced computational techniques. In addition to traditional computer vision approaches, tumor classification methods using transfers are being actively developed through the use of deep convolutional neural networks (CNNs). This study starts with examining the CNN based transfer learning to characterize breast masses for different diagnostic, predictive tasks or prognostic or in several imaging modalities, such as Magnetic Resonance Imaging (MRI), Ultrasound (US), digital breast tomosynthesis and mammography. The deep learning framework contains several convolutional layers, LSTM, Max-pooling layers. The classification and error estimation that has been included in a fully connected layer and a soft max layer. This paper focuses on combining these machine learning approaches with the methods of selecting features and extracting them through evaluating their output using classification and segmentation techniques to find the most appropriate approach. The experimental results show that the high accuracy level of 97.2%, Sensitivity 98.3%, and Specificity 96.5% has been compared to other existing systems

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
		V
	ABSTRACT	
		VI
	TABLE OF CONTENTS	
	LIST OF FIGURES	VIII
1.	INTRODUCTION	1
	1.1 INTRODUCTION	1
	1.2 BREAST CANCER SUBTYPES	2
	1.3 SIGNS AND SYMPTOMS	5
	1.4 BREAST CANCER CLASSIFICATION	7
	1.5 INTRODUCTION TO IMAGE PROCESSING	9
	1.6 TYPES OF DIGITAL IMAGE	10
	1.7 COLOR MODEL	11
	1.7.1 RGB COLOR MODEL	11
	1.7.2 CONVERTING COLOR TO GRAY SCALE	11
	1.7.3 GRAYSCALE AS SINGLE CHANNEL OF MULTI CHANNEL COLOR IMAGE	12
	1.8 DIGITAL IMAGE PROCESSING TECHNIQUES	13
	1.8.1 IMAGE REPRESENTATION	14
	1.8.2 IMAGE ENHANCEMENT	15
	1.8.3 IMAGE SEGMENTATION	15

	1.9 FUNDAMENTAL IN DIGITAL IMAGE PROCESSING	15
	1.9.1 IMAGE ACQUISITION	16
	1.9.2 IMAGE PREPROCESSING	17
	1.9.3 IMAGE SEGMENTATION	17
	1.9.4 IMAGE REPRESENTATION AND DESCRIPTION	18
	1.9.5 IMAGE RECOGNITION AND INTERPRETATION	18
2.	LITERATURE SURVEY	19
3.	METHODOLOGY	27
	3.1 EXISTING SYSTEM	27
	3.2 PROPOSED SYSTEM	27
	3.3 MODULE DESCRIPTION	27
	3.4 ARCHITECTURE	28
	3.5 WATERSHED SEGMENTATION	29
	3.5.1 DISTANCE TRANSFORM APPROACH	29
	3.5.2 GRADIENT METHOD	30
	3.5.3 MARKER CONTROLLER METHODS	30
	3.6 GRAY LEVEL CO-OCCURRENCE MATRICES	31
	3.7 SUPPORT VECTOR MACHINE	32
4.	SYSTEM DESIGN	35
	4.1 ACTIVITY DIAGRAM	35
	4.2 SEQUENCE DIAGRAM	35
	4.3 USE CASE DIAGRAM	36
5.	SYSTEM REQUIREMENTS	38
	5.1 HARDWARE REQUIREMENTS	38

	5.2 SOFTWARE REQUIREMENTS	38
	5.3 INTRODUCTION TO MATLAB R2018a	39
6.	RESULTS AND DISCUSSIONS	43
7.	CONCLUSION	45
	REFERENCE	46

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
1.1	NORMAL BREAST (LEFT) AND BREAST WITH CANCER (RIGHT)	4
1.2	BREAST CANCER STAGE	8
1.3	RGB COLOR MODEL	11
1.4	COMPOSITION OF RGB FROM GRAYSCALE IMAGE HISTOGRAM EQUALIZATION	12
1.5	HISTOGRAM EQUALIZATION OF A COLOR IMAGE	13
1.6	IMAGE REPRESENTATION	14
1.7	FUNDAMENTAL STEPS IN DIGITAL IMAGE PROCESSING	16
4.1	ACTIVITY DIAGRAM	35
4.2	SEQUENCE DIAGRAM	36
4.3	USE CASE DIAGRAM	37

CHAPTER 1

INTRODUCTION

1.1 Introduction:

The breast is made up of different tissue, ranging from very fatty tissue to very dense tissue. Within this tissue is a network of lobes. Each lobe is made up of tiny, tube-like structures called lobules that contain milk glands. Tiny ducts connect the glands, lobules, and lobes, carrying milk from the lobes to the nipple. The nipple is located in the middle of the areola, which is the darker area that surrounds the nipple. Blood and lymph vessels also run throughout the breast. Blood nourishes the cells. The lymph system drains bodily waste products. The lymph vessels connect to lymph nodes, the tiny, bean-shaped organs that help fight infection.

- ✓ Cancer begins when healthy cells in the breast change and grow out of control, forming a mass or sheet of cells called a tumor. A tumor can be cancerous or benign. A cancerous tumor is malignant, meaning it can grow and spread to other parts of the body. A benign tumor means the tumor can grow but will not spread.
- ✓ Breast cancer spreads when the cancer grows into other parts of the body or when breast cancer cells move to other parts of the body through the blood vessels and/or lymph vessels. This is called a metastasis.
- ✓ This guide covers early-stage and locally advanced breast cancer, which includes stages I, II, and III. The stage of breast cancer describes where the cancer is located, how much the cancer has grown, and if or where it has spread.
- ✓ Although breast cancer most commonly spreads to nearby

lymph nodes, it can also spread further through the body to areas such as the bones,

Learn more about metastatic breast cancer in a separate section on this website.

- ✓ If breast cancer comes back after initial treatment, it can recur locally, meaning in the breast and/or regional lymph nodes. The regional lymph nodes are those nearby the breast, such as the lymph nodes under the arm. It can also recur elsewhere in the body, called a distant recurrence or metastatic recurrence.

1.2 Breast cancer subtypes:

There are 3 main subtypes of breast cancer that are determined by doing specific tests on a sample of the tumor. These tests will help your doctor learn more about your cancer and recommend the most effective treatment plan.

Testing the tumor sample can find out if the cancer is:

-) Hormone receptor positive. Breast cancers expressing estrogen receptors (ER) and/or progesterone receptors (PR) are called “hormone receptor positive.” These receptors are proteins found in cells. Tumors that have estrogen receptors are called “ER positive.” Tumors that have progesterone receptors are called “PR positive.” Only 1 of these receptors needs to be positive for a cancer to be called hormone receptor positive. This type of cancer may depend on the hormones estrogen and/or progesterone to grow. Hormone receptor-positive cancers can occur at any age, but are more common in women who have gone through menopause. About 60% to 75% of breast cancers have estrogen and/or progesterone receptors. Cancers without these receptors are called “hormone receptor negative.”
-) HER2 positive. About 10% to 20% of breast cancers depend on the gene called *human epidermal growth factor receptor 2*

(*HER2*) to grow. These cancers are called “HER2 positive” and have many copies of the *HER2* gene or high levels of the HER2 protein. These proteins are also called “receptors.” The *HER2* gene makes the HER2 protein, which is found on the cancer cells and is important for tumor cell growth. HER2-positive breast cancers grow more quickly. They can also be either hormone receptor positive or hormone receptor negative. Cancers that have no or low levels of the HER2 protein and/or few copies of the *HER2* gene are called “HER2 negative.”

- J Triple negative. If a tumor does not express ER, PR, or HER2, the tumor is called “triple negative.” Triple-negative breast cancer makes up about 15% to 20% of invasive breast cancers. Triple-negative breast cancer seems to be more common among younger women, particularly younger black and Hispanic women. Triple-negative cancer is also more common in women with a mutation in the *BRCA1* or *BRCA2* genes. Experts recommend that all people with triple-negative breast cancer younger than 60 be tested for *BRCA* gene mutations. See the Risk Factors and Prevention section for more information on these genetic mutations.

In addition to invasive breast cancer, 58,590 new cases of in situ breast cancer are expected to occur among women during 2005. Of these, approximately 88% will be classified as ductal carcinoma in situ (DCIS). The detection of DCIS cases is a direct result of the increased use of mammography screening. This screening method is also responsible for detection of invasive cancers at a less advanced stage than might have occurred otherwise. An estimated 40,870 deaths (40,410 women, 460 men) were anticipated from breast cancer in 2005. Breast cancer ranks second among cancer deaths in women. According to the most recent data, mortality rates declined significantly during 1992-1998, with the largest decreases in younger women, both white and black.

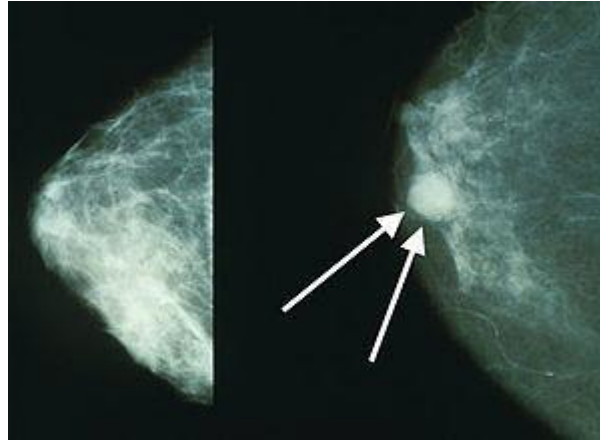


Fig 1.1 normal breast (left) and a breast with cancer (right)

Breast cancer is cancer that develops from breast tissue. Signs of breast cancer may include a lump in the breast, a change in breast shape, dimpling of the skin, and fluid coming from the nipple, a newly-inverted nipple, or a red or scaly patch of skin. In those with distant spread of the disease, there may be bone pain, swollen lymph nodes, shortness of breath, or yellow skin.

therapy during menopause, ionizing radiation, an early age at first menstruation, having children late in life or not at all, older age, having a prior history of breast cancer, and a family history of breast cancer. About 5–10% of cases are the result of a genetic predisposition inherited from a person's parents, including BRCA1 and BRCA2 among others. Breast cancer most commonly develops in cells from the lining of milk ducts and the lobules that supply these ducts with milk. Cancers developing from the ducts are known as ductal carcinomas, while those developing from lobules are known as lobular carcinomas. There are more than 18 other sub-types of breast cancer some, such as ductal carcinoma in situ, develop from pre-invasive lesions. The diagnosis of breast cancer is confirmed by taking a biopsy of the concerning tissue. Once the diagnosis is made, further tests are done to determine if the cancer has spread beyond the breast and which treatments are most likely to be effective.

The balance of benefits versus harms of breast cancer screening is controversial. A 2013 Cochrane review found that it was unclear if mammographic screening does more harm than good, in that a large proportion of women who test positive turn out not to have the disease.[9] A 2009 review for the US Preventive Services Task Force found evidence of benefit in those 40 to 70 years of age, and the organization recommends screening every two years in women 50 to 74 years of age. The medications tamoxifen or raloxifene may be used in an effort to prevent breast cancer in those who are at high risk of developing it. Surgical removal of both breasts is another preventative measure in some high risk women. In those who have been diagnosed with cancer, a number of treatments may be used, including surgery, radiation therapy, chemotherapy, hormonal therapy, and targeted therapy. Types of surgery vary from breast-conserving surgery to mastectomy. Breast reconstruction may take place at the time of surgery or at a later date. In those in whom the cancer has spread to other parts of the body, treatments are mostly aimed at improving quality of life and comfort.

Outcomes for breast cancer vary depending on the cancer type, the extent of disease, and the person's age. The five-year survival rates in England and the United States are between 80 and 90%. In developing countries, five-year survival rates are lower. Worldwide, breast cancer is the leading type of cancer in women, accounting for 25% of all cases. In 2018 it resulted in 2 million new cases and 627,000 deaths.[6] It is more common in developed countries and is more than 100 times more common in women than in men.

1.3 Signs and symptoms:

Breast cancer most commonly presents as a lump that feels different from the rest of the breast tissue. More than 80% of cases are discovered when a person detects such a lump with the fingertips. The earliest breast cancers, however, are detected by a mammogram. Lumps found in lymph nodes located in the armpits may also indicate breast cancer.

Indications of breast cancer other than a lump may include thickening different from the other breast tissue, one breast becoming larger or lower, a nipple

changing position or shape or becoming inverted, skin puckering or dimpling, a rash on or around a nipple, discharge from nipple/s, constant pain in part of the breast or armpit and swelling beneath the armpit or around the collarbone. Pain ("mastodynia") is an unreliable tool in determining the presence or absence of breast cancer, but may be indicative of other breast health issues.

Another symptom complex of breast cancer is Paget's disease of the breast. This syndrome presents as skin changes resembling eczema; such as redness, discoloration or mild flaking of the nipple skin. As Paget's disease of the breast advances, symptoms may include tingling, itching, increased sensitivity, burning, and pain. There may also be discharge from the nipple. Approximately half the women diagnosed with Paget's disease of the breast also have a lump in the breast.

Inflammatory Breast Cancer presents with similar effects. Inflammatory Breast Cancer is a rare (only seen in less than 5% of breast cancer diagnosis) yet aggressive form of breast cancer characterized by the swollen, red areas formed on the top of the Breast. The visual effects of Inflammatory Breast Cancer are a result of a blockage of lymph vessels by cancer cells. This type of breast cancer is seen in more commonly diagnosed in younger ages, obese women and African American women. As inflammatory breast cancer does not present as a lump there can sometimes be a delay in diagnosis.

In rare cases, what initially appears as a fibro adenoma (hard, movable non- cancerous lump) could in fact be a phyllodes tumor. Phyllodes tumors are formed within the stroma (connective tissue) of the breast and contain glandular as well as stromal tissue. Phyllodes tumors are not staged in the usual sense; they are classified on the basis of their appearance under the microscope as benign,

borderline or malignant.

Malignant tumors can result in metastatic tumors— secondary tumors (originating from the primary tumor) that spread beyond the site of origination. The symptoms caused by metastatic breast cancer will depend on the location of metastasis. Common sites of metastasis include bone, liver, lung, and brain. When cancer has reached such an invasive state, it is categorized as a

stage 4 cancer; cancers of this state are oftentimes fatal. Common symptoms of stage 4 cancers include unexplained weight loss, bone and joint pain, jaundice and neurological symptoms. These symptoms are called non-specific symptoms because they could be manifestations of many other illnesses.

Most symptoms of breast disorders, including most lumps, do not turn out to represent underlying breast cancer. Less than 20% of lumps, for example, are cancerous, and benign breast diseases such as mastitis and fibro adenoma of the breast are more common causes of breast disorder symptoms.

1.4 Breast cancer classification:

Breast cancers are classified by several grading systems. Each of these influences the prognosis and can affect treatment response. Description of a breast cancer optimally includes all of these factors.

) **Histopathology.** Breast cancer is usually classified primarily by its histological appearance. Most breast cancers are derived from the epithelium lining the ducts or lobules, and these cancers are classified as ductal or lobular carcinoma. *Carcinoma in situ* is growth of low-grade cancerous or precancerous cells within a particular tissue compartment such as the mammary duct without invasion of the surrounding tissue. In contrast, *invasive carcinoma* does not confine itself to the initial tissue compartment.

) **Grade.** Grading compares the appearance of the breast cancer

cells to the appearance of normal breast tissue. Normal cells in an organ like the breast become differentiated, meaning that they take on specific shapes and forms that reflect their function as part of that organ. Cancerous cells lose that differentiation. In cancer, the cells that would normally line up in an orderly way to make up the milk ducts become disorganized. Cell division becomes uncontrolled. Cell nuclei become less uniform. Pathologists describe cells as well differentiated (low grade), moderately differentiated (intermediate grade), and poorly differentiated (high grade) as the cells progressively lose the features seen in normal breast cells.

Poorly differentiated cancers (the ones whose tissue is least like normal breast tissue) have a worse prognosis.

-) **Stage.** Breast cancer staging using the TNM system is based on the size of the tumor (T), whether or not the tumor has spread to the lymph nodes (N) in the armpits, and whether the tumor has metastasized (M) (i.e. spread to a more distant part of the body).
-) Stage 0 is a pre-cancerous or marker condition, either ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS).
-) Stages 1–3 are within the breast or regional lymph nodes.
-) Stage 4 is 'metastatic' cancer that has a less favorable prognosis since it has spread beyond the breast and regional lymph nodes.



Fig 1.2 Breast cancer stage

1.5 INTRODUCTION TO IMAGE PROCESSING:

Image Processing is a method to enhance raw images received from cameras/sensors placed on satellites, space probes and aircrafts or pictures taken in normal day-to-day life for various applications. Various techniques have been developed in Image Processing during the last four to five decades. Most of the techniques are developed for enhancing images obtained from unmanned spacecrafts, space probes and military reconnaissance flights. Image Processing systems are becoming popular due to easy availability of powerful personal computers, large size memory devices, graphics software's etc. Image processing is a physical process used to convert an image signal into a physical image. The

image signal can be either digital or analog. The actual output itself can be an actual physical image or the characteristics of an image.

Image processing is photography. In this process, an image is captured or scans using a camera to create a digital or analog image. In order to produce a physical picture, the image is processed using the appropriate technology based on the input source type. In digital photography, the image is stored as a computer file. This file is translated using photographic software to generate an actual image. The colors, shading, and nuances are all captured at the time the photograph is taken the software translates this information into an image. When creating images using analog photography, the image is burned into a film using a chemical reaction triggered by controlled exposure to light. The image is processed in a darkroom, using special chemicals to create the actual image. This process is decreasing in popularity due to the opening of digital photography, which requires less effort and special training to produce images. The field of digital imaging has created a whole range of new applications and tools that were previously impossible. Face recognition software, medical image

processing and remote sensing are all possible due to the development of digital image processing. Specialized computer programs are used to enhance and correct images.

1.6 Types of Digital Image

For photographic purposes, there are two important types of digital images: color and grayscale. Color images are made up of colored pixels while grayscale images are made of pixels in different shades of gray.

- **Grayscale Images:** A grayscale image is made up of pixels, each of which holds a single number corresponding to the gray level of the image at a particular location. These gray levels span the full range from black to white in a series of very fine steps, normally 256 different grays. Assuming 256 gray levels, each black and white pixel can be stored in a single byte (8 bits) of memory.

- **Color Images:** A color image is made up of pixels, each of which holds three numbers corresponding to the red, green and blue levels of the image at a particular location. Assuming 256 levels, each color pixel can be stored in three bytes (24 bits) of memory. Note that for images of the same size, a black & white version will use three times less memory than a color version.

- **Binary Images:** Binary images use only a single bit to represent each pixel. Since a bit can only exist in two states- ON or OFF; every pixel in a binary image must be one of two colors, usually black or white. This inability to represent intermediate shades of gray is what limits their usefulness in dealing with photographic images

1.7 Color Model:

1.7.1 RGB Color Model:

A representation of additive color mixing. Projection of primary color lights on a screen shows secondary colors where two overlap; the combination of all three of red, green, and blue in appropriate intensities makes white.



Fig 1.3 RGB color model

The RGB color model is an additive color model in which red, green, and blue light is added together in various ways to reproduce a broad array of colors. The name of the model comes from the initials of the three additive primary colors, red, green, and blue. The main purpose of the RGB color model is for the sensing, representation, and display of images in electronic systems, such as televisions and computers, though it has also been used in conventional photography.

1.7.2 CONVERTING COLOR TO GRAYSCALE

Conversion of a color image to grayscale is not unique; different weighting of the color channels effectively represents the effect of shooting black- and-white film _ with different-colored photographic filters on the cameras.

To convert any color to a grayscale representation of its

luminance, first one must obtain the values of its red, green, and blue (RGB) primaries in linear intensity encoding, by gamma expansion. Then, add together 30% of the red value, 59% of the green value, and 11% of the blue value (these weights depend on the exact choice of the RGB primaries, but are typical). Regardless of the scale employed (0.0 to 1.0, 0 to 255, 0% to 100%, etc.) the resultant number is the desired linear luminance value; it typically needs to be gamma_compressed to get back to a conventional grayscale representation.

This is not the method used to obtain the luma in the YUV and related color models, used in standard color TV and video systems as PAL and NTSC, as well as in the L*a*b color model. These systems directly compute a gamma- compressed luma as a linear combination .

1.7.3 GRAYSCALE AS SINGLE CHANNELS OF MULTICHANNEL COLOR IMAGES

Color images are often built of several stacked color channels, each of them representing value levels of the given channel. For example, RGB images are composed of three independent channels for red, green and blue primary color components; CMYK images have four channels_for cyan, magenta, yellow and black ink plates, etc.

Here is an example of color channel splitting of a full RGB color image. The column at left shows the isolated color channels in natural colors, while at right there are their grayscale equivalences:

The reverse is also possible: to build a full color image from their separate grayscale channels. By mangling channels, using offsets, rotating and other manipulations, artistic effects can be achieved instead of accurately reproducing the original image.

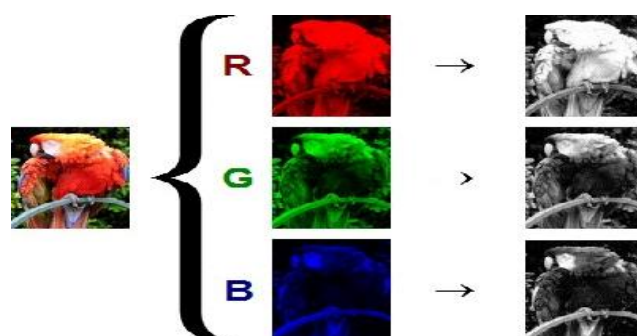
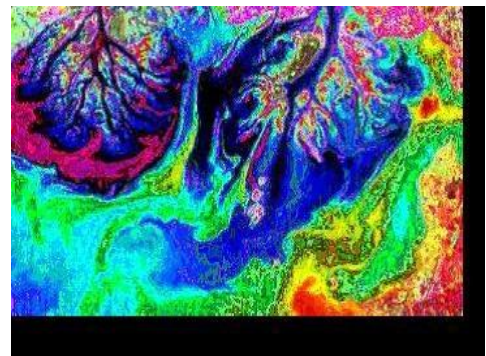
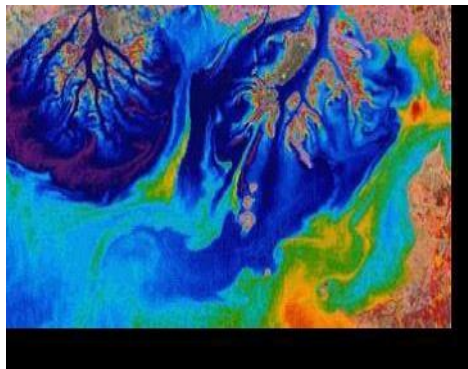


Fig 1.4 Composition of RGB from 3 Grayscale images
HISTOGRAM EQUALIZATION OF A COLOR IMAGE:

Figure shows a color image with 768 x 512 pixels and 16 levels for each band. As a result of the color histogram equalization, Figure 3 shows more diversity in color than. Due to lack of space, the three-dimensional histogram is not depicted here. In fact, it is equalized uniformly. It took 412.6 seconds to equalize the histogram with nonzero values at 1501 colors out of 4096. Consecutive probabilities in the histogram would be further investigated. The results of the histogram equalization for two color images are displayed.



(a)Original image

(b) Color-equalized image

Fig 1.5 Histogram equalization of a color image.

1.8 DIGITAL IMAGE PROCESSING TECHNIQUES

The various Image Processing techniques are:

-) Image representation
-) Image preprocessing
-) Image enhancement
-) Image restoration

-) Image analysis
-) Image reconstruction
-) Image data compression

1.8.1 IMAGE REPRESENTATION

An image defined in the "real world" is considered to be a function of two real variables, for example, $f(x,y)$ with f as the amplitude (e.g. brightness) of the image at the *real* coordinate position (x,y) . The effect of digitization is shown in Figure.

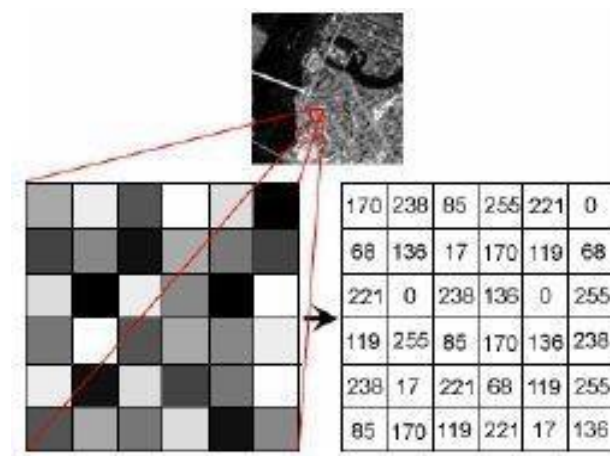


Fig.1.6 Image Representation

The 2D continuous image $f(x,y)$ is divided into N rows and M columns. The intersection of a row and a column is called as *pixel*. The value assigned to the integer coordinates $[m,n]$ with $\{m=0,1,2,\dots,M-1\}$ and $\{n=0,1,2,\dots,N-1\}$ is $f[m,n]$. In fact, in most cases $f(x, y)$ which we might consider to be the physical signal that impinges on the face of a sensor. Typically an image file such as BMP, JPEG, TIFF etc., has some header and picture information. A header usually includes details like format identifier (typically first information), resolution, number of bits/pixel, compression type, etc.

1.8.2 IMAGE ENHANCEMENT

Sometimes images obtained from satellites and conventional and digital cameras lack in contrast and brightness because of the limitations of imaging sub systems and illumination conditions while capturing image. Images may have different types of noise. In image enhancement, the goal is to accentuate

certain image features for subsequent analysis or for image display. Examples include contrast and edge enhancement, pseudo-coloring, noise filtering, sharpening, and magnifying. Image enhancement is useful in feature extraction, image analysis and an image display. The enhancement process itself does not increase the inherent information content in the data. It simply emphasizes certain specified image characteristics. Enhancement algorithms are generally interactive and application-dependent.

1.8.3 IMAGE SEGMENTATION:

Image segmentation is the process that subdivides an image into its constituent parts or objects. The level to which this subdivision is carried out depends on the problem being solved, i.e., the segmentation should stop when the objects of interest in an application have been isolated e.g., in autonomous air-to-ground target acquisition, suppose our interest lies in identifying vehicles on a road, the first step is to segment the road from the image and then to segment the contents of the road down to potential vehicles.

1.9 FUNDAMENTAL IN DIGITAL IMAGE PROCESSING

) Image Acquisition

-) Image preprocessing
-) Image segmentation
-) Image Representation and Description
-) Image Recognition and Interpretation and
-) Knowledge base.

1.9.1 IMAGE ACQUISITION

In the process, the first step in the process is image acquisition that is, to acquire a digital image. It highly requires an imaging sensor and the capability to digitize the signal produced by the sensor.

The sensor could be a monochrome or color TV camera that produces an entire image of the problem domain every 1/30 sec.

The imaging sensor could also a line-scan camera that produces a single image line at a time. In this case, the object's motion past the line scanner produces two-dimensional image. If the output of the camera or other imaging sensor is not already in digital form, an analog-to-digital converter digitizes it. The nature of the sensor and image it produces are determined by the application.

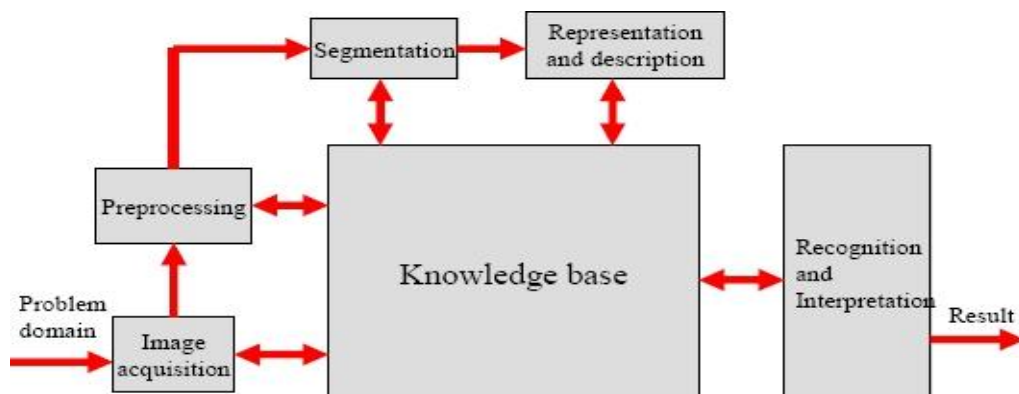


Figure 1.7 Fundamental steps in digital image processing

1.9.2 IMAGE PREPROCESSING

The key function of preprocessing is to improve the image in ways that increase the chances for success of the other processes. Preprocessing typically deals with techniques for enhancing contrast, removing noise, and isolating regions whose texture indicate a likelihood of alphanumeric information.

-) Image Enhancement.
-) More suitable than original image for a specific application.
-) Image Restoration.
-) A process that attempts to reconstruct or recover an image that has been degraded using prior knowledge of the degradation concept.

1.9.3 IMAGE SEGMENTATION

The next stage deals with segmentation. Segmentation partitions an input image into its small constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. On the one hand, a rugged segmentation procedure brings the process a long way towards the successful solution of an imaging problem. On the other hand, weak or erratic segmentation algorithms almost guarantee eventual failure.

In terms of character recognition, the key role of segmentation is to extract individual characters and words from the background.

1.9.4 IMAGE REPRESENTATION AND DESCRIPTION

The output of the segmentation stage usually is raw pixel data, constituting either the boundary of a region or all the points in the region itself. In either case converting the data to a form suitable for computer processing is necessary.

The first decision that must be made, whether the data should be represented as a boundary or as a complete region. Boundary representation is appropriate when the focus is on external shape characteristics, such as corners and inflections.

Regional representation is appropriate when the focus is on internal properties, such as texture or skeletal shape. In some applications, however, these representations coexist.

This situation occurs in character recognition applications, which often require algorithms based on boundary shape as well as skeletons and other internal properties.

1.9.5 IMAGE RECOGNITION AND INTERPRETATION

Recognition is the process that assigns a label to an object based on the information provided by its descriptors. Interpretation involves assigning meaning to an ensemble of recognized objects. For example, identifying a character as, say, *c* requires associating the descriptors for that character with the label *c*. Interpretation attempts to assign meaning to a set of labeled entities. For example, a string of five numbers are followed by a hyphen and four more numbers can be interpreted to be a ZIP code.

.CHAPTER 2

LITERATURE SURVEY

2.1 Project Title : Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis

Author Name : Alexander Rakhlin¹, Alexey Shvets², Vladimir Iglovikov³

Year of Publish : 2018

Abstract :

Breast cancer is one of the main causes of cancer death worldwide. Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed potential for improving the diagnostic accuracy. In this work, we develop the computational approach based on deep convolution neural networks for breast cancer histology image classification. Hematoxylin and eosin stained breast histology microscopy image dataset is provided as a part of the ICIAR __Grand Challenge on Breast Cancer Histology Images. Our approach utilizes several deep neural network architectures and gradient boosted trees classifier. For __-class classification task, we repor._% accuracy. For __-class classification task to detect

carcinomas we report_._% accuracy, AUC _._%,andsensitivity/specificity _._/____.%at the high-sensitivity operating point. To our knowledge, this approachoutperforms other common methods in automated histopathological imageclassification.

2.2 Project Title : Breast Cancer Classification Using Deep Neural Networks

Author Name : S. Karthik, R. Srinivasa Perumal and P. V. S. S. R. Chandra Mouli

Year of Publish : 2018

Abstract :

Early diagnosis of any disease can be curable with a little amount of human effort. Most of the people fail to detect their disease before it becomes chronic. It leads to increase in death rate around the world. Breast cancer is one of the diseases that could be cured when the disease identified at earlier stages before it is spreading across all the parts of the body. The medical practitioner may diagnose the diseases mistakenly due to misinterpretation. The computer-aided diagnosis (CAD) is an automated assistance for practitioners that will produce accurate results to analyze the criticality of the diseases. This chapter presents a CAD system to perform automated diagnosis for breast cancer. This method employed deep neural network (DNN) as classifier model and recursive feature elimination (RFE) for feature selection. DNN with multiple layers of processing attained higher classification rate than SVM. So, the researchers used deep learning method for hyper-spectral data classification. This chapter used DNN to learn deep features of data. The DNN with multiple layers of processing is applied to classify the breast cancer data. The system was experimented on Wisconsin Breast Cancer Dataset (WBCD) from UCI repository. The dataset partitioned into different sets of train-test split. The performance of the system is measured based on accuracy, sensitivity, specificity, precision, and recall. From the results, the accuracy obtained 98.62%, which is better than other state-of-the-art methods. The results show that the system is comparatively outperformed than the existing system.

2.3 Project Title : Label-Efficient Breast Cancer
Histopathological Image Classification

Author Name : Yanlong Li, Jitian Wang, Han Zheng, Yue Huang,
Xinghao Ding,

Year of Publish : 2018

Abstract :

The automatic classification of breast cancer histopathological images has great significance in computer- aided diagnosis. Recently, deep learning via neural networks has enabled pattern detection and prediction using large, labeled datasets; whereas, collecting and annotating sufficient histological data using professional pathologists is time-consuming, tedious, and extremely expensive. In the proposed work, a deep active learning framework is designed and implemented for classifica- tion of breast cancer histopathological images, with the goal of maximizing the learning accuracy from very limited labeling. This method involves manual annotation of the most valuable unlabeled samples, which are then integrated into the training set. The model is then iteratively updated with an increasing training set. Here, two selection strategies are discussed for the proposed deep active learning framework: an entropy-based strategy and a confidence-boosting strategy. The proposed method has been val- idated using a publicly available breast cancer histopathological image dataset, wherein each image patch is binarily classified as benign or malignant. The experimental results demonstrate that, compared with a random selection, our proposed framework can reduce annotation costs up to 66.67%, with higher accuracy and less expensive annotation than standard query strategy

2.4 Project Title : Adaptive deep learning for head and neck cancer detection using hyperspectral imaging

Author Name : Ling Ma^{1,2}, Guolan Lu¹, Dongsheng Wang³, Xulei Qin¹, Zhuo Georgia

Year of Publish : 2019

Abstract :

It can be challenging to detect tumor margins during surgery for complete resection. The purpose of this work is to develop a novel learning method that learns the difference between the tumor and benign tissue adaptively for

cancer detection on hyperspectral images in an animal model. Specifically, an auto-encoder network is trained based on the wavelength bands on hyperspectral images to extract the deep information to create a pixel-wise prediction of cancerous and benign pixel. According to the output hypothesis of each pixel, the misclassified pixels would be reclassified in the right prediction direction based on their adaptive weights. The auto-encoder network is again trained based on these updated pixels. The learner can adaptively improve the ability to identify the cancer and benign tissue by focusing on the misclassified pixels, and thus can improve the detection performance. The adaptive deep learning method highlighting the tumor region proved to be accurate in detecting the tumor boundary on hyperspectral images and achieved a sensitivity of 92.32% and a specificity of 91.31% in our animal experiments. This adaptive learning method on hyperspectral imaging has the potential to provide a noninvasive tool for tumor detection, especially, for the tumor whose margin is indistinct and irregular

2.5 Project Title : Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection

Author Name : Richard Platania, Shayan Shams, Seungwon Yang

Year of Publish : 2018

Abstract :

Detection of suspicious regions in mammogram images and the subsequent diagnosis of these regions remains a challenging problem in the medical world.

There still exists an alarming rate of misdiagnosis of breast cancer. This results in both over treatment through incorrect positive diagnosis of cancer and under treatment through overlooked cancerous masses. Convolutional neural networks have shown strong applicability to various image datasets, enabling detailed features to be learned from the data and, as a result, the ability to classify these images at extremely low error rates. In order to overcome the difficulty in

diagnosing breast cancer from mammogram images, we propose our framework for automated breast cancer detection and diagnosis, called BC-DROID, which provides automated region of interest detection and diagnosis using convolutional neural networks. BC-DROID _rst pretrains based on physician- de_ned regions of interest in mammogram images. It then trains based on the full mammogram image. The resulting network is able to detect and classify regions of interest as cancerous or benign in one step. We demonstrate the accuracy of our framework's ability to both locate the regions of interest as well as diagnose them. Our framework achieves a detection accuracy of up to 90% and a classi_cation accuracy of 93.5% (AUC of 92.315%). To the best of our knowledge, this is the _rst work enabling both automated detection and diagnosis of these areas in one step from full mammogram images. Using our framework's website, a user can upload a single mammogram image, visualize suspicious regions, and receive the automated diagnoses of these regions

2.6 Project Title : Breast Cancer Histopathological Image Classification using Convolutional Neural Networks

Author Name : Fabio Alexandre Spanhol¹, Luiz S. Oliveira², Caroline Petitjean³

Year of Publish : 2018

Abstract :

The performance of most conventional classification systems relies on appropriate data representation and much of the efforts are dedicated to feature engineering, a difficult and time-consuming process that uses prior expert domain knowledge of the data to create useful features. On the other hand, deep

learning can extract and organize the discriminative information from the data, not requiring the design of feature extractors by a domain expert. Convolutional Neural Networks (CNNs) are a particular type of deep, feedforward network that have gained attention from research community and industry, achieving empirical successes in tasks

such as speech recognition, signal processing, object recognition, natural language processing and transfer learning. In this paper, we conduct some preliminary experiments using the deep learning approach to classify breast cancer histopathological images from BreakeHis, a publicly dataset available at <http://web.inf.ufpr.br/vri/breast-cancer-database>. We propose a method based on the extraction of image patches for training the CNN and the combination of these patches for final classification. This method aims to allow using the highresolution histopathological images from BreakeHis as input to existing CNN, avoiding adaptations of the model that can lead to a more complex and computationally costly architecture. The CNN performance is better when compared to previously reported results obtained by other machine learning models trained with hand-crafted textural descriptors. Finally, we also investigate the combination of different CNNs using simple fusion rules, achieving some improvement in recognition rates

2.7 Project Title : Mitosis Detection Using Convolutional Neural Network Based Features

Author Name : Abdulkadir Albayrak, Gokhan Bilgin

Year of Publish : 2016

Abstract :

Breast cancer is the second leading cause of cancer death in women according to World Health Organization (WHO). Development of computer aided diagnostic (CAD) systems has great importance as a secondary reader systems for a correct diagnosis and treatment process. In this paper, a deep learning based feature extraction method by convolutional neural network (CNN) is proposed for automated mitosis detection for cancer diagnosis and grading by histopathological images. The proposed framework is tested on the MITOS data set provided for a contest on mitosis detection in breast cancer histological images released for research purposes in International

Conference on Pattern Recognition (ICPR'2014). By using provided histopathological images, cellular structures are initially found by combined clustering based segmentation and blob analysis after preprocessing step. Then, obtained cellular image patches are cropped automatically from the histopathological images for feature extraction stage. CNN, which is a prominent deep learning method on image processing tasks, is utilized for extracting discriminative features. Due to the high dimensional output of the CNN, combination of PCA and LDA dimension reduction methods are performed respectively for regularization and dimension reduction process. Afterwards, a robust kernel based classifier, support vector machine (SVM), is used for final classification of mitotic and non-mitotic cells. The test results on MITOS data set prove that the proposed framework achieved promising results for mitosis detection on histopathological images.

2.8 Project Title : Deep learning beyond cats and dogs: Recent advances in diagnosing breast cancer with deep neural networks

Author Name : 1,2Jeremy R Burt, MD, 2Neslisah Torosdagli

Year of Publish : 2018

Abstract :

Deep learning has demonstrated tremendous revolutionary changes in the computing industry and its effects in radiology and imaging sciences have begun to dramatically change screening paradigms. Specifically, these advances

have influenced the development of computer-aided detection and diagnosis (CAD) systems. These technologies have long been thought of as “second- opinion” tools for radiologists and clinicians. However, with significant improvements in deep neural networks, the diagnostic capabilities of learning algorithms are approaching levels of human expertise (radiologists, clinicians etc.), shifting the CAD paradigm from a “second opinion” tool to a more collaborative utility. This paper reviews recently developed CAD systems based on deep learning technologies for breast cancer diagnosis, explains their superiorities with respect to previously established systems, defines

the

methodologies behind the improved achievements including algorithmic developments, and describes remaining challenges in breast cancer screening and diagnosis. We also discuss possible future directions for new CAD models that continue to change as artificial intelligence algorithms evolve

2.9 Project Title : Automatic mass detection in mammograms using deep convolutional neural networks

Author Name : Richa Agarwal, Oliver Diaz, Xavier Lladó

Year of Publish : 2019

Abstract :

With recent advances in the field of deep learning, the use of convolutional neural networks (CNNs) in medical imaging has become very encouraging. The aim of our paper is to propose a patch-based CNN method for automated mass detection in full-field digital mammograms (FFDM). In addition to evaluating CNNs pretrained with the ImageNet dataset, we investigate the use of transfer learning for a particular domain adaptation. First, the CNN is trained using a large public database of digitized mammograms (CBIS-DDSM dataset), and then the model is transferred and tested onto the smaller database of digital mammograms (INbreast dataset). We evaluate three widely used CNNs (VGG16, ResNet50, InceptionV3) and show that the InceptionV3 obtains the best performance for classifying the mass and nonmass breast region for CBIS- DDSM. We further show the benefit of domain adaptation between the CBIS- DDSM (digitized) and INbreast (digital) datasets using the InceptionV3 CNN. Mass detection evaluation follows a fivefold cross-validation strategy using free- response operating characteristic curves. Results show that the transfer learning from CBIS-DDSM obtains a substantially higher performance with the best true positive rate (TPR) of 0.98 ± 0.02 at 1.67 false positives per image (FPI), compared with transfer learning from ImageNet with TPR of 0.91 ± 0.07 at 2.1 FPI. In addition, the proposed framework improves upon mass detection results described in it.

CHAPTER 3

METHODOLOGY

3.1 Existing System:

The deep learning assisted efficient adaboost algorithm has been proposed for breast cancer detection and early diagnosis. In addition to traditional computer vision approaches, tumor classification methods using transfers are being actively developed through the use of deep convolution neural networks (CNN). The fully connected convolution layer has been utilized for feature selection, feature extraction, detection, segmentation and classification for evaluating different stages of breast cancer.

3.2 Proposed system:

In the proposed system, design of Critical and Early Breast cancer detection is done using machine learning approach to make the system detect the tumor cells automatically using advanced image processing techniques. Image is preprocessed by studying various parameters extractions such as color conversion, resizing and filtering. Watershed segmentation is carried out by segmentation algorithms. This helps to identify the amount of lesions scattered over the body. Feature extraction is by threshold and finally, Approximate reasoning method to recognize the tumor shape and position in MRI image using classification method. Finally the message box will be displayed whether it is benign or malignant.

3.3 MODULE DESCRIPTION:

Module I: Pre-processing

According to the need of the next level the pre processing step convert the image. It performs filtering of noise and other artifacts in the image and sharpening the edges in the image. RGB to grey conversion and Reshaping also takes place here. It includes shock filter for noise removal.

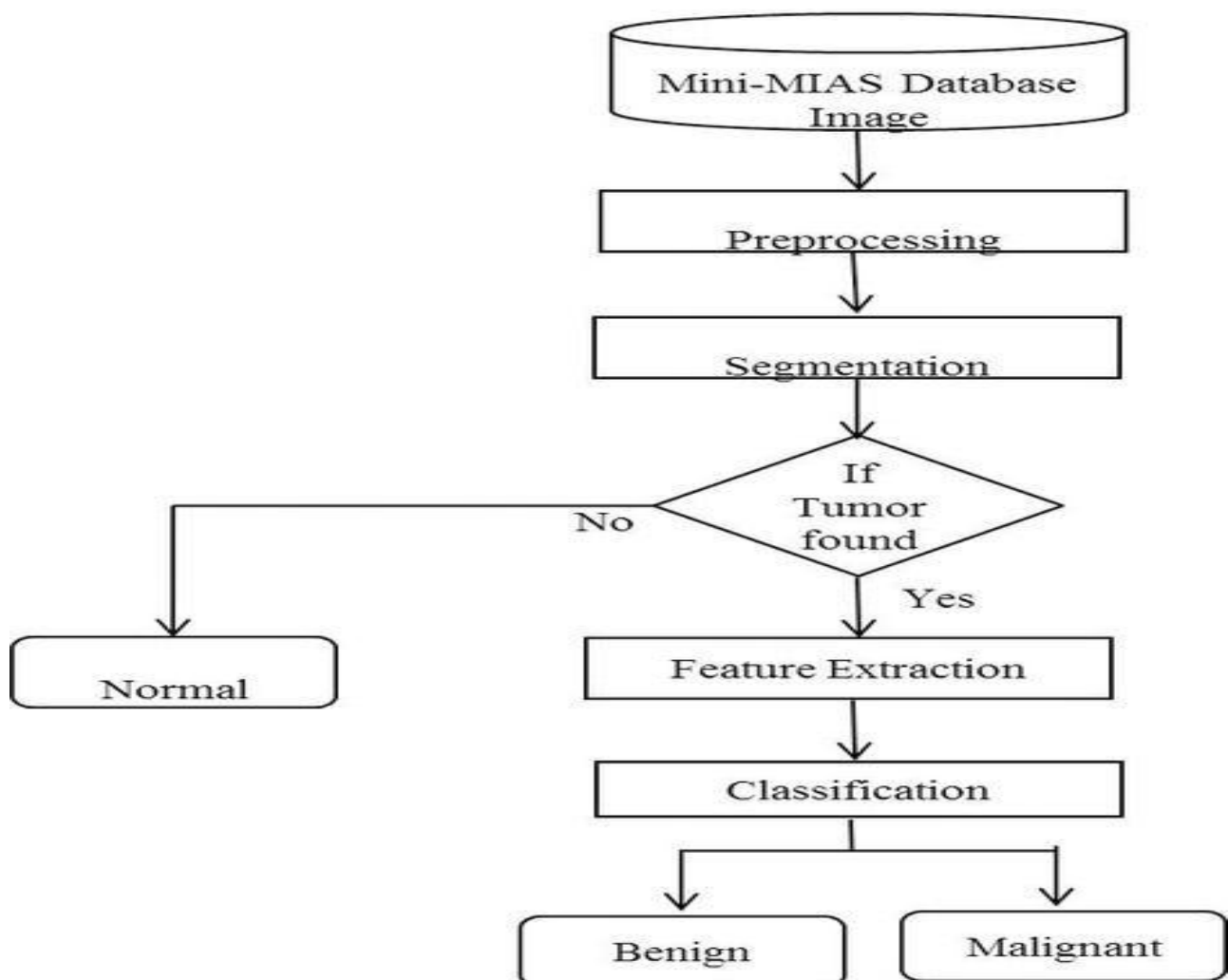
Module II: Segmentation

Segmentation is carried out by advanced medical vision learning such as active contour segmentation algorithm. The feature extraction is extracting the cluster which shows the predicted tumor at the FCM output. The extracted cluster is given to the threshold process. It applies binary mask over the entire image. In the approximate reasoning step the tumor area is calculated using the binarization method.

Module III: Classification

In each segmented part is given to the classifier recognized. We used Support vector machine (SVM) classifier in order to gain a clear understanding of the relationship between the inputs and the outputs of the models and to facilitate a comparison of the classification performance.

3.4 Architecture:



3.5 Watershed Segmentation:

In computer vision, image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images. More precisely image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture.

There are two major approaches in image segmentation: edge- and region- based. Edge based segmentation partitions an image based on discontinuities with sub-regions, while region-based segmentation does the similar function based on the uniformity of a desired property within a sub-region.

There are mainly three methods to implement watershed. They are listed below:

- Distance Transform Approach
- Gradient method
- Marker Controlled Approach

3.5.1 Distance Transform Approach

A tool used commonly in conjunction with the watershed transform for segmentation is the distance transform. It is the distance from every pixel to the nearest nonzero valued pixel. The distance transform can be computed using toolbox

function `bwdist`, calling syntax, is

$D = \text{bwdist}(f)$

A binary image can be converted to a gray level image, which is suitable for watershed segmentation using different DT. However, different DT functions produce different effects. Euclidean DT has a higher possibility of “salt and pepper” over segmentation.

3.5.2 Gradient Method

The gradient magnitude is used to pre process a grayscale image prior to using the watershed transform for segmentation. The gradient magnitude image has high pixel values along object edges and low pixel values everywhere else. Watershed transform would result in watershed ridge lines along object edges. There is a problem of over segmentation in this method. The topological gradient provides a global analysis of the image then the almost unwanted contours due to the noise added to a given image can be significantly reduced by our approach. The experimental results show that the over segmentation problem, which usually appears with the watershed technique, can be attenuated, and the segmentation results can be performed using the topological gradient approach. Another advantage of this method is that it splits the segmentation process into two separate steps: first we detect the main edges of the image processed, and then we compute the watershed of the gradient detected.

3.5.3 Marker Controller Methods

Direct application of watershed transform to a gradient image can result in over segmentation due to noise. Over segmentation means a large number of segmented regions. An approach used to control over segmentation is based on the

concept of markers. A marker is a connected component belonging to an image. Markers are used to modify the gradient image. Markers are of two types internal and external, internal for object and external for boundary. The marker controlled watershed segmentation

has been shown to be a robust and flexible method for segmentation of objects with closed contours, where the boundaries are expressed as ridges. Markers are placed inside an object of interest; internal markers associate with objects of interest, and external markers associate with the background. After segmentation, the boundaries of the watershed regions are arranged on the desired ridges, thus separating each object from its neighbors.

ALGORITHM

Segmentation is used for identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation follows this basic procedure:

Step 1. Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.

Step 2. Compute foreground markers. These are connected blobs of pixels within each of the objects.

Step 3. Compute background markers. These are pixels that are not part of any object.

Step 4. Modify the segmentation function so that it only has minima at the foreground and background marker locations.

Step 5. Compute the watershed transform of the modified segmentation function.

3.6 Gray-Level Co-Occurrence Matrices (GLCM):

A co-occurrence matrix or co-occurrence distribution is a matrix or distribution that is defined over an image to be the distribution of co-occurring values at a given offset. Mathematically, a co-occurrence matrix C is defined over an $n \times m$ image I , parameterized by an offset $(\Delta x, \Delta y)$, as:

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases}$$

Where i and j are the image intensity values of the image, p and q are the spatial positions in the image I and the offset $(\Delta x, \Delta y)$ depends on the direction used and the distance at which the matrix is computed d . The 'value' of the image originally referred to the grayscale value of the specified pixel, but could be anything, from a binary on/off value to 32-bit color and beyond.

1. Homogeneity, Angular Second Moment (ASM):

Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Range lies between [0 1]. Homogeneity is 1 for a diagonal GLCM. .

2. Contrast:

Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image. Range = $[0 \text{ (size(GLCM,1)-1)}^2]$. Contrast is 0 for a constant image.

homogeneous scene has high entropy.

3. Correlation:

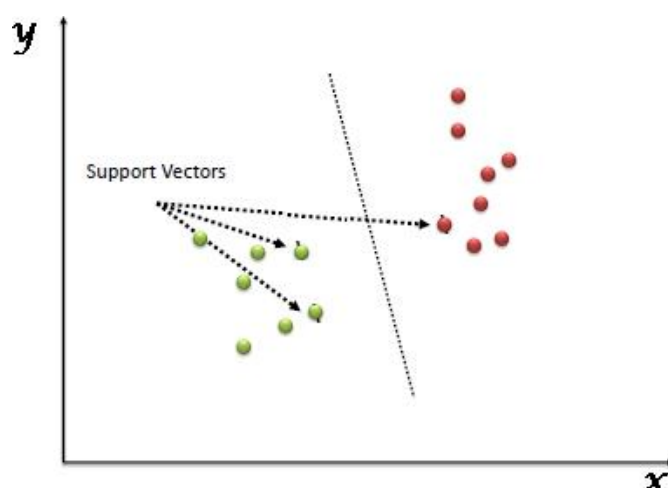
Returns a measure of how correlated a pixel is to its neighbor over the whole image. Range = $[-1 \text{ } 1]$. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant image.

4. Energy

It returns the sum of squared elements in the GLCM. Range lies between [0 1]. Energy is 1 for a constant image.

3.7 Support Vector Machine :

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for either classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n -dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).



Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

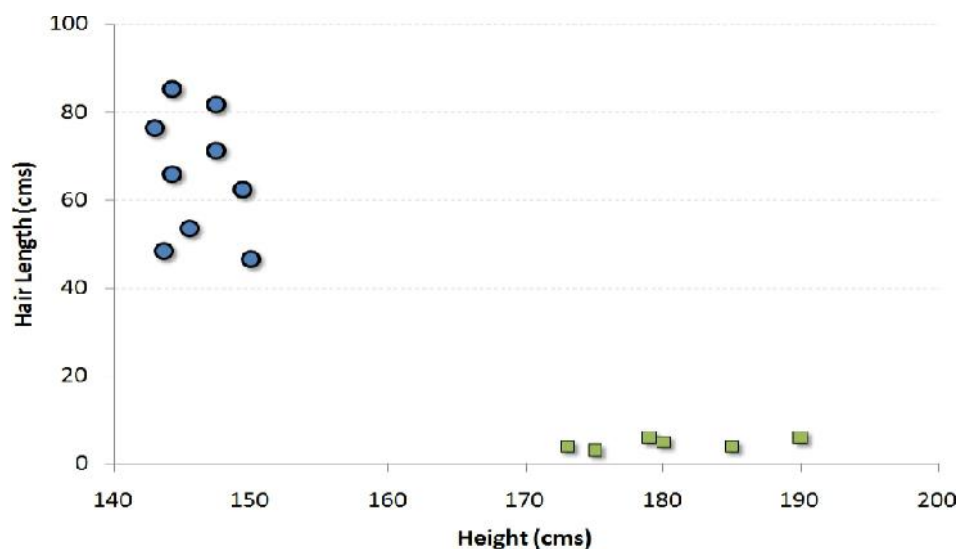
Support Vector Machine – Simplified

Let's talk about how support vector machine works. This is suitable for readers who do not know much about this algorithm and have a curiosity to learn a new technique. In following explore the technique in detail and analyze cases where such techniques are stronger than other techniques.

Classification analysis

Let's consider an example to understand these concepts.

We have a population composed of 50%-50% Males and Females. Using a sample of this population, you want to create some set of rules which will guide us the gender class for rest of the population. Using this algorithm, we intend to build a robot which can identify whether a person is a Male or a Female. This is a sample problem of classification analysis. Using some set of rules, we will try to classify the population into two possible segments. For simplicity, let's assume that the two differentiating factors identified are: Height of the individual and Hair Length. Following is a scatter plot of the sample.



The blue circles in the plot represent females and green squares represents male. A few expected insights from the graph are:

1. Males in our population have a higher average height.
2. Females in our population have longer scalp hairs.

CHAPTER 4

SYSTEM DESIGN

4.1 ACTIVITY DIAGRAM

It captures the dynamic behavior of the system and is used to show message flow from one activity to another.

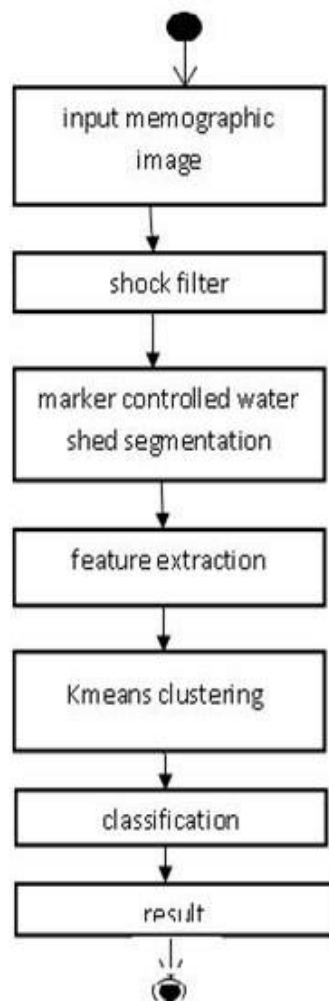


FIG 4.1 ACTIVITY DIAGRAM

4.2 SEQUENCE DIAGRAM

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these

interactions take place.

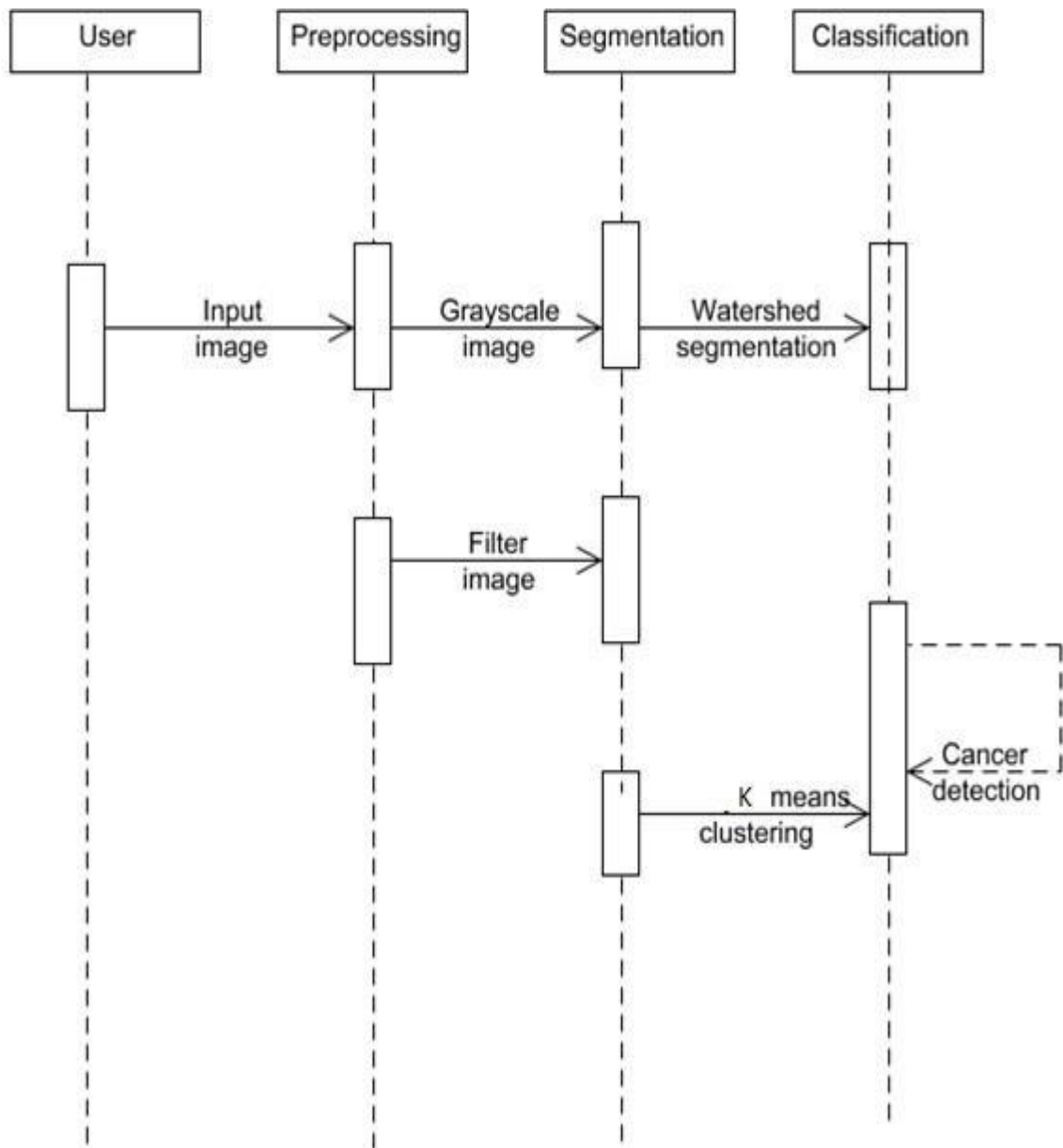


FIG 4.2 SEQUENCE DIAGRAM

4.3 USE CASE DIAGRAM

A use case diagram is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.

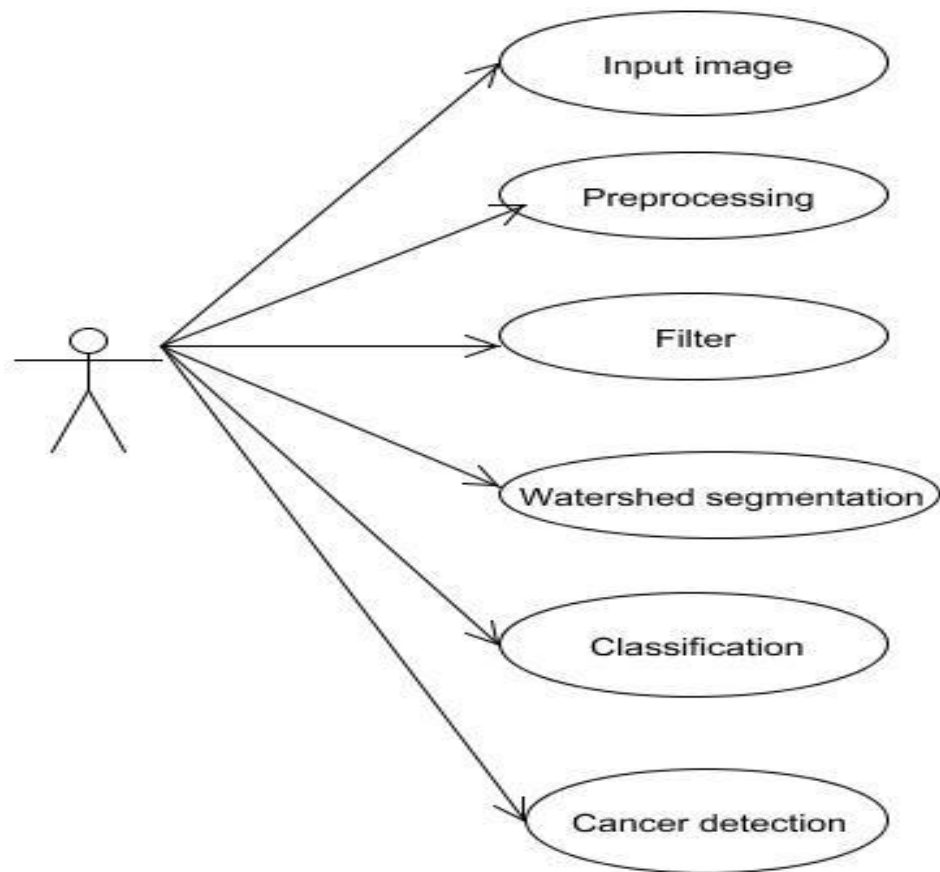


FIG 4.3 USE CASE DIAGRAM

CHAPTER 5

SYSTEM REQUIREMENTS

5.1 HARDWARE REQUIREMENTS

- Hard Disk : 40GB and Above
- RAM : 512MB and Above
- Processor : Pentium III and Above

5.2 SOFTWARE REQUIREMENTS

- MATLAB R2013a
- Windows Operating System XP and Above

MATLAB (**matrix laboratory**) is a multi-paradigm numerical computing environment and fourth-generation programming language. A proprietary programming language developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the Mu PAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems. The basic data structure in MATLAB is the array, an ordered set of real or complex elements. This object is naturally suited to the representation of images, real-valued ordered sets of color or intensity data.

MATLAB stores most images as two-dimensional arrays (i.e., matrices), in which each element of the matrix corresponds to a single pixel in the displayed image. (Pixel is derived from picture element and usually denotes a single dot on a computer display.) For example, an image composed of 200 rows and 300 columns of different coloured dots would be stored in MATLAB as a 200-by-300 matrix. Some images, such as RGB, require a three-dimensional array, where the first plane

in the third dimension represents the red pixel intensities, the second plane represents the green pixel intensities, and the third plane represents the blue pixel intensities.

This convention makes working with images in MATLAB similar to working with any other type of matrix data, and makes the full power of MATLAB available for image processing applications. For example, you can select a single pixel from an image matrix using normal matrix subscripting.

5.3 INTRODUCTION TO MATLAB R2018a:

The MATLAB System

The MATLAB system consists of these main parts:

Desktop Tools and Development Environment

This is the set of tools and facilities that help you use and become more productive with MATLAB functions and files. Many of these tools are graphical user interfaces. It includes the MATLAB desktop and Command Window, a command history, an editor and debugger, a code analyzer and other reports, and browsers for viewing help, the workspace, files, and the search path.

Mathematical Function Library

This is a vast collection of computational algorithms ranging from elementary functions, like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix Eigen values, Bessel functions, and fast Fourier transforms.

The Language

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create large and complex application

programs.

Graphics

MATLAB has extensive facilities for displaying vectors and matrices as graphs, as well as annotating and printing these graphs. It includes high-level functions for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level functions that allow you to fully customize the appearance of graphics as well as to build complete graphical user interfaces on your MATLAB applications.

External Interfaces

This is a library that allows you to write C and FORTRAN programs that interact with MATLAB. It includes facilities for calling routines from MATLAB (dynamic linking), for calling MATLAB as a computational engine, and for reading and writing MAT-files.

Image Processing Toolbox

The Image Processing Toolbox is a collection of functions that extend the capability of the MATLAB numeric computing environment. The toolbox supports a wide range of image processing operations, including:

- Spatial image transformations
- Morphological operations
- Neighborhood and block operations
- Linear filtering and filter design
- Transforms
- Image analysis and enhancement
- Image registration
- De blurring
- Region of interest operations

Many of the toolbox functions are MATLAB M-files, a series of MATLAB statements that implement specialized image processing algorithms. You can extend the capabilities of the Image Processing

Toolbox by writing your own M-

files, or by using the toolbox in combination with other toolboxes, such as the Signal Processing Toolbox and the Wavelet Toolbox.

Computer Vision Toolbox

The field is of importance for such various applications as autonomous vehicles, navigating with the help of images, captured by a mounted camera, and high precision measurements using images, taken by calibrated cameras. In this paper, we will present a number of numerical routines, implemented in MATLAB, that are useful in a variety of computer vision applications. The collection of routines will be called the Computer Vision Toolbox. One of the main problems in Computer Vision is to calculate the 3D-structure of the scene and the motion of the camera from measurements in the images taken from different view-points in the scene. This problem is called structure and motion, referring to the fact that both the structure of the scene and the motion of the camera are calculated from image measurements only. A number of different sub problems, arising from different knowledge of the intrinsic properties of the camera appear. Other important problems are to calculate the structure of the 3D-scene given the motion of the camera and to calculate the motion of the camera given the structure of the scene. These problems are somewhat simpler than the general structure and motion problem, but are nevertheless important for navigation and obstacle avoidance. A related problem is to calibrate a camera, i.e. calculate the focal distance, the principal point etc., from images of a known scene. This calibration may be used to make a more precise reconstruction. The problem of estimating structure and motion from image sequences are closely related to the field of closed range photogrammetric. In this field images are used to make precise measurements of three-dimensional objects from a number of images, often taken by calibrated cameras. The main difference is that in computer vision, the focus is not on accuracy, but instead on reliability and speed of calculation. In all

these problems different kind of so called features can be used. The simplest feature is points that are easily detected in the sequence, such as corners or painted marks. Other types of features are lines and curves, which are easier to detect and track but more difficult to use in structure and motion algorithms. In order to use our computer vision routines these features must be detected in advance and the correspondence between different images must be established. The main emphasis of our computer vision toolbox is to use these detected features to solve for structure and motion.

Image Acquisition Toolbox

The Image Acquisition Toolbox's is a collection of functions that extend the capability of the MATLAB numeric computing environment. The toolbox supports a wide range of image acquisition operations including

- Acquiring images through many types of image acquisition devices, from

professional grade frame grabbers to USB-based Webcams

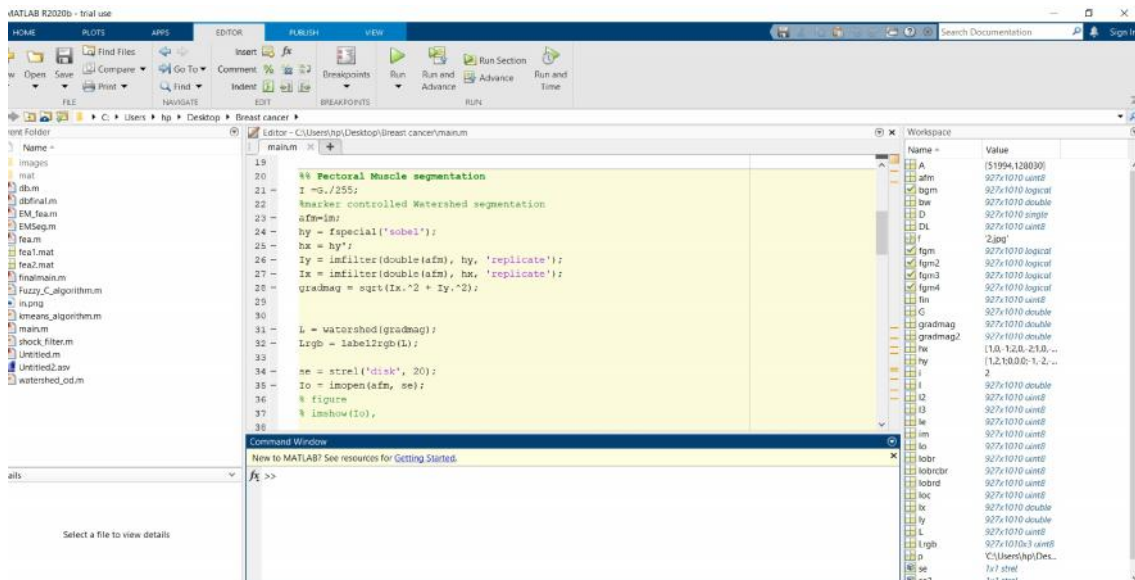
- Viewing a preview of the live video stream
- Triggering acquisitions (includes external hardware triggers)
- Configuring call back functions that execute when certain events occur
- Bringing the image data into the MATLAB workspace

Many of the toolbox functions are MATLAB M-files. You can extend the capabilities of the Image Acquisition Toolbox by writing your own M-files, or by using the toolbox in combination with other toolboxes, such as the Image Processing Toolbox and the Data Acquisition Toolbox. The toolbox also includes a Simulink interface called the Image Acquisition Block set. This block set extends Simulink with a block that lets you bring live video data into a model.

CHAPTER 6

RESULTS AND DISCUSSION

- ❖ the deep learning assisted efficient AdaBoost algorithm has been proposed for breast cancer detection and early diagnosis.
- ❖ The AdaBoost algorithm for the final prediction function to build an ensemble classifier.
- ❖ As the results of the estimation test show, our suggested approach as the greater capacity to predict and the deep-learning classifier is better than the other classifiers.
- ❖ Our discussion and analysis have shown excellent potential for fast generalization and directly enhanced the efficiency of the prediction of the result, which are derived automatically by the neural network.
- ❖ Taking advantage of the high-deep learning from the Convolutional Neural Network deep learning model proposed DLA-EABA helped improve the system performance.
- ❖ The proposed DLAEABS method has high accuracy in detecting breast cancer mass and increases the patient survival rate. The performance of the suggested approach is very high when compared to other existing methods.



CHAPTER 7

CONCLUSION

In this work, to detect the suspicious lesions from mammograms we have used watershed segmentation algorithm followed by morphological operation. After selecting the suitable filter for lesion detection, we have used watershed algorithm, which gives segmented image with pectoral muscle and to remove the pectoral muscle, we have applied morphological operation, which detected the lesion from the mammographic images more accurately. This method can help the radiologists for exact detection of breast cancer in its primary stage. Finally the support vector machine(SVM) classifier is used to check whether the given image is benign or malignant.

REFERENCES

- [1] P. Gupta and A. Kaur Malhi, "Using deep learning to enhance head and neck cancer diagnosis and classification," in Proc. IEEE Int. Conf. Syst., Comput., Autom. Netw., New York, NY, USA, Jul. 2018, pp. 1–6.
- [2] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," Expert Syst. Appl., vol. 46, pp. 139–144, Mar. 2016.
- [3] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. C. Shih, J. Tomaszewski, F. A. González, and A. Madabhushi, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," Sci. Rep., vol. 7, no. 1, Jun. 2017, Art. no. 46450.
- [4] S. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," Sensors, vol. 18, no. 9, p. 2799.
- [5] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwigelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," IEEE J. Biomed. Health Informat., vol. 22, no. 4, pp. 1218–1226, Jul. 2018.
- [6] D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," (2016), arXiv:1612.00542. [Online]. Available: <https://arxiv.org/abs/1612.00542>
- [7] S. Suzuki, X. Zhang, N. Homma, K. Ichiji, N. Sugita, Y. Kawasumi, T. Ishibashi, and M. Yoshizawa, "Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis," in Proc. 55th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE), Sep. 2016, pp. 1382–1386.
- [8] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," in Proc. Biocomputing, Jan. 2017, pp. 219–229.
- [9] M. Saha, C. Chakraborty, and D. Racocanu, "Efficient deep

learning model for mitosis detection using breast histopathology images,” *Comput. Med. Imag. Graph.*, vol. 64, pp. 29–40, Mar. 2018.

[10] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, “Deep learning for identifying metastatic breast cancer,” 2016, arXiv:1606.05718. [Online]. Available: <http://arxiv.org/abs/1606.05718>

[11] N. Dhungel, G. Carneiro, and A. P. Bradley, “Automated mass detection in mammograms using cascaded deep learning and random forests,” in *Proc. Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA)*, Nov. 2015, pp. 1–8.