P-3381

DIAGNOSIS OF BREAST TUMORS USING SONOGRAPHIC TEXTURE ANALYSIS

By

S.P.ANANDKUMAR Reg No: 0710107003

M.ARULANANDHAM Reg No: 0710107009

N.KARTHICK Reg No: 0710107040

S.MEIGNANAMOORTHI Reg No: 0710107052

of

KUMARAGURU COLLEGE OF TECHNOLOGY,
COIMBATORE - 641 049.
(An Autonomous Institution affiliated to Anna University of Technology, Coimbatore)

A PROJECT REPORT

Submitted to the

FACULTY OF ELECTRONICS AND COMMUNICATION ENGINEERING

In partial fulfillment of the requirements
for the award of the degree

of

BACHELOR OF ENGINEERING
in
ELECTRONICS AND COMMUNICATION ENGINEERING

KUMARAGURU COLLEGE OF TECHNOLOGY

COIMBATORE -641 049

(An Autonomous Institution Affiliated to Anna University of Technology, Coimbatore)

BONAFIDE CERTIFICATE

Certified that this project report **"DIAGNOSIS OF BREAST TUMORS USING SONOGRAPHIC TEXTURE ANALYSIS"** is the bonafide work of **"S.P.ANAND KUMAR (0710107003), M.ARULANANDHAM (0710107009), N.KARTHICK (0710107040), and S.MEIGNANAMOORTHI (0710107052)"** who carried out the project work under my supervision.

	18/4/2011
SIGNATURE	**SIGNATURE**
DR.RAJESWARI MARIAPPAN Ph.D.,	MRS.S.SASIKALA M.Tech.,
HEAD OF DEPARTMENT &	**PROJECT GUIDE**
PROFESSOR	**ASSISTANT PROFESSOR**
Electronics & Communication	Electronics & Communication
Engineering	Engineering
Kumaraguru College of Technology	Kumaraguru College of technology
Coimbatore.	Coimbatore.

The candidates with University Register Nos. 0710107003, 0710107009, 0710107040, 0710107052 were examined by us in the project viva-voce examination held on..18..April..2011...

ABSTRACT

The main aim of our project is to increase the ability of ultrasonographic technology for the differential diagnosis of solid breast tumors; we describe a novel computer-aided diagnosis (CADx) system using neural networks for classification of breast tumors. The basic classification of breast tumors called benignancy and malignancy can be easily detected using the proposed algorithm. Wavelet transformation is used for decomposition of the given ultrasonographic image into various sub bands for easy segmentation. Tumor regions and surrounding tissues are segmented from the physician-located region-of-interest (ROI) images by applying our proposed segmentation algorithm. Cooperating with the segmentation algorithm, three feasible features, including variance contrast, autocorrelation contrast and distribution distortion of wavelet coefficients, were extracted from the ROI images for further classification.

ACKNOWLEDGEMENT

We are extremely grateful to **Dr.J.Shanmugam, Ph.D., Director,** Kumaraguru College of Technology for having given us a golden opportunity to embark on this project.

We are extremely grateful to **Dr.Ramachandran, M.Tech., Ph.D., Principal**, Kumaraguru College of Technology for having given us a golden opportunity to embark on this project.

We are deeply obliged to **Dr.Rajeshwari Mariappan, M.E., Ph.D., B.Tech.Ed., F.I.E, MISTE, Head of the Department** of Electronics& Communication for her concern and implication during the project course.

We extend our heartfelt thanks to our project coordinator **Asst.Prof.V.Jeyasri Arokiamary, M.E.,** and our respectable Guide **Asst.Prof.S.Sasikala M.Tech.,** Department of Electronics and Communication for their helpful guidance and valuable support given to us throughout this project. Our thanks to all **Teaching and non-teaching staffs** of our department for providing us the technical support for our project.

We also thank **our college Management**, **friends and family** who helped us to complete this project with flying colours.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

A tumor or tumour is the name for a neoplasm or a solid lesion formed by an abnormal growth of cells (termed neoplastic) which looks like a swelling.

Tumors are classified into three types

1. Benign(mild and non progressive disease)
2. Pre-malignant(condition that may lead to cancer)
3. Malignant(progressively worse and to potentially result in death)

Breast cancer also known as malignant breast neoplasm is cancer originating from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk.

There are two main types of breast cancer:

- Ductal carcinoma starts in the tubes (ducts) that move milk from the breast to the nipple. Most breast cancers are of this type.
- Lobular carcinoma starts in the parts of the breast, called lobules, that produce milk.
- In rare cases, breast cancer can start in other areas of the breast.

Breast cancer may be invasive or noninvasive. Invasive means it has spread from the milk duct or lobule to other tissues in the breast. Noninvasive means it has not yet invaded other breast tissue. Noninvasive breast cancer is called "in situ".

- Ductal carcinoma in situ (DCIS), or intraductal carcinoma, is breast cancer in the lining of the milk ducts that has not yet invaded nearby tissues. It may progress to invasive cancer if untreated.
- Lobular carcinoma in situ (LCIS) is a marker for an increased risk of invasive cancer in the same or both breasts.

CHAPTER-2

VARIOUS STAGES IN BREAST CANCER

Stage	Definition
Stage 0	Cancer cells remain inside the breast duct, without invasion into normal adjacent breast tissue.
Stage 1	Cancer is 2 centimeters or less and is confined to the breast (lymph nodes are clear).
Stage IIA	No tumor can be found in the breast, but cancer cells are found in the axillary lymph nodes (the lymph nodes under the arm) OR The tumor measures 2 centimeters or smaller and has spread to the axillary lymph nodes OR The tumor is larger than 2 but no larger than 5 centimeters and has not spread to the axillary lymph nodes.
Stage IIB	The tumor is larger than 2 but no larger than 5 centimeters and has spread to the axillary lymph nodes OR The tumor is larger than 5 centimeters but has not spread to the axillary lymph nodes

tage IIIA	No tumor is found in the breast. Cancer is found in axillary lymph nodes that are sticking together or to other structures, or cancer may be found in lymph nodes near the breastbone OR The tumor is any size. Cancer has spread to the axillary lymph nodes, which are sticking together or to other structures, or cancer may be found in lymph nodes near the breastbone.
Stage IIIB	The tumor may be any size and has spread to the chest wall and/or skin of the breast AND May have spread to axillary lymph nodes that are clumped together or sticking to other structures, or cancer may have spread to lymph nodes near the breastbone. Inflammatory breast cancer is considered at least stage IIIB
Stage IIIC	There may either be no sign of cancer in the breast or a tumor may be any size and may have spread to the chest wall and/or the skin of the breast AND the cancer has spread to lymph nodes either above or below the collarbone AND the cancer may have spread to axillary lymph nodes or to lymph nodes near the breastbone.

tage IV	The cancer has spread — or metastasized — to other parts of the body.

Table 1.stage of cancer

reast Cancer Statistics

- About 1 in 8 women in the United States (12%) will develop invasive breast cancer over the course of her lifetime.
- In 2010, an estimated 207,090 new cases of invasive breast cancer were expected to be diagnosed in women in the U.S., along with 54,010 new cases of non-invasive (in situ) breast cancer.
- About 1,970 new cases of invasive breast cancer were expected to be diagnosed in men in 2010. Less than 1% of all new breast cancer cases occur in men.
- About 39,840 women in the U.S. were expected to die in 2010 from breast cancer, though death rates have been decreasing since 1990. These decreases are thought to be the result of treatment advances, earlier detection through screening, and increased awareness.
- About 70-80% of breast cancers occur in women who have no family history of breast cancer. These occur due to genetic abnormalities that happen as a result of the aging process and life in general, rather than inherited mutations.
- About 5-10% of breast cancers can be linked to gene mutations (abnormal changes) inherited from one's mother or father. Mutations of the BRCA1 and BRCA2 genes are the most common. Women with these mutations have up to an 80% risk of developing breast cancer during their lifetime.
- The most significant risk factors for breast cancer are gender (being a woman) and age (growing older).

For women in the U.S., breast cancer death rates are higher than those for any other cancer, besides lung cancer.

ymptoms of Breast Cancer

Initially, breast cancer may not cause any symptoms. A lump may be too mall for you to feel or to cause any unusual changes you can notice on your own.)ften, an abnormal area turns up on a screening mammogram (x-ray of the breast), /hich leads to further testing.

In some cases, however, the first sign of breast cancer is a new lump or mass 1 the breast that you or your doctor can feel. A lump that is painless, hard, and has 1neven edges is more likely to be cancer. But sometimes cancers can be tender, oft, and rounded. So it's important to have anything unusual checked by your loctor.

According to the American Cancer Society, any of the following unusual :hanges in the breast can be a symptom of breast cancer:

- swelling of all or part of the breast
- skin irritation or dimpling
- breast pain
- nipple pain or the nipple turning inward
- redness, scaliness, or thickening of the nipple or breast skin
- a nipple discharge other than breast milk
- a lump in the underarm area

These changes also can be signs of less serious conditions that are not cancerous, such as an infection or a cyst. It's important to get any breast changes checked out promptly by a doctor.

CHAPTER-3

METHODS FOR DIAGNOSING BREAST CANCER

The various methods to diagnosing breast cancer are as follows:

- Self examination
- Mammography (low-dose amplitude-X-rays)
- Sonography(ultrasound-based diagnostic imaging technique)
- Biopsy(accurate method yet costly)

3.1 MAMMOGRAPHY

Mammography is the process of using low-dose amplitude-X-rays (usually around 0.7 mSv) to examine the human breast and is used as a diagnostic and a screening tool. The detection of breast cancer is typically through detection of characteristic masses and/or microcalcifications. Mammography is believed to reduce mortality from breast cancer. The resultant image obtained from a mammography test is called as mammogram and it is used for further analysis.

Mammography techniques are classified into two types

- Film-screen mammography
- Digital mammography

3.1.1 FILM-SCREEN MAMMOGRAPHY

Film-screen mammography is a breast radiographic technique in which a special single-emulsion film and high-detail intensifying screens are used. The technique provides a fine image at radiation exposure levels of less than 1 rad, compared with older methods that generated radiation levels of as much as 16 rad.

3.1.2 DIGITAL MAMMOGRAPHY

Digital mammography is a specialized form of mammography that uses digital receptors and computers instead of x-ray film to help examine breast tissue for breast cancer. The electrical signals can be read on computer screens, permitting more manipulation of images to theoretically allow radiologists to more clearly view the results. Digital mammography may be "spot view", for breast biopsy, or "full field" (FFDM) for screening.

3.1.3 ADVANTAGES OF DIGITAL MAMMOGRAPHY

- Can allow the technician taking the x-ray to make adjustments, such as enhancement, magnification, or other manipulation without having to take another mammogram
- Can reduce exposure to radiation, due to reduction of the need for repeat mammograms
- Takes pictures of the entire breast, even if the denseness of the breast tissue varies
- Compared to film mammography, digital mammography is more sensitive for detection of abnormalities in women with dense breasts, pre-menopausal women, perimenopausal women, and women <50 years of age[1, 2]

3.2BIOPSY

A biopsy is a medical test involving the removal of cells or tissues for examination. It is the medical removal of tissue from a living subject to determine the presence or extent of a disease. The tissue is generally examined under a microscope by a pathologist, and can also be analyzed chemically.

Breast biopsy is mainly done by any one of the following types

- Fine needle aspiration
- Core needle biopsy
- Vacuum assisted biopsy
- Direct & frontal biopsy
- Open surgical biopsy

3.2.1 TYPES

Fine needle aspiration

Fine needle aspiration (FNA) is a percutaneous ("through the skin") procedure that uses a fine needle and a syringe to sample fluid from a breast cyst or remove clusters of cells from a solid mass. With FNA, the cellular material taken from the breast is usually sent to the pathology laboratory for analysis.

Core needle biopsy

A core needle biopsy is a procedure that removes small but solid samples of tissue using a hollow "core" needle. For palpable ("able to be felt") lesions, the physician fixes the lesion with one hand and performs a freehand needle biopsy with the other.

Vacuum assisted biopsy

Vacuum assisted biopsy is a version of core needle biopsy using a vacuum technique to assist the collection of the tissue sample.

Direct & frontal biopsy

Recent innovations in tissue acquisition for the human breast have led to the development of unique direct frontal systems. Efficacy is considered optimal if the diagnosis by transcutaneous biopsy is identical to the surgical specimen in case of malignancy or in line with clinical follow-up when benign.

Open surgical biopsy

Open surgical biopsy means that a large mass or lump is removed during a surgical procedure. Surgical biopsy requires an approximately 3 to 5 centimeters incision and is normally performed in an operating room in sterile conditions.

3.3SONOGRAPHY

Diagnostic sonography (ultrasonography) is an ultrasound-based diagnostic imaging technique used for visualizing subcutaneous body structures including tendons, muscles, joints, vessels and internal organs for possible pathology or lesions. Obstetric sonography is commonly used during pregnancy and is widely recognized by the public. Ultrasound exams do not use ionizing radiation (as used in x-rays). Because ultrasound images are captured in real-time, they can show the structure and movement of the body's internal organs, as well as blood flowing through blood vessels.

3.3.1ADVANTAGES

Since ultrasound uses a different type of radiation (i.e.) sound waves it sees the body differently than other imaging modalities. Ultrasound is non-ionising radiation. X-rays, which are used in the traditional X-ray picture and in computed tomography (CT), form images based on the reaction of the X-rays to the different densities of the tissues. Sonograms are based on the speed and intensity of the sound waves returning from different body structures. For this reason, it is safer than an X-ray, and sonograms can take a different look at a particular structure.

Ultrasound can be offered as a screening tool for women who:

- Are at high risk for breast cancer and unable to tolerate an MRI examination.
- Are at intermediate risk for breast cancer based on family history, personal history of breast cancer, or prior biopsy showing an abnormal result.
- Have dense breasts.
- have silicone breast implants and very little tissue can be included on the mammogram.
- are pregnant or should not to be exposed to x-rays (which is necessary for a mammogram).

3.3.2SONOGRAPHY OVER MAMMOGRAPHY

Although the nonpalpable and minimal tumors can be detected by mammography, sonography is still suitable for palpable tumors.

Sonography has been shown to be similar in overall effectiveness to the use of mammography and even better than mammography for women less than 35 years old.

Silicone breast implants and very little tissue included on the mammogram.

MAHATMA GANDHI
KCT
CBE-6
CENTRAL LIBRARY
P. 3381

CHAPTER-4

PROCESS INVOLVED

The various techniques involved in classifying the breast mass using the ultrasound image are as follows

1. Wavelet transform
2. Segmentation
3. Feature extraction

4.1 WAVELET TRANSFORM

In mathematics, a wavelet series is a representation of a square-integrable (real- or complex-valued) function by a certain orthonormal series generated by a wavelet. This article provides a formal, mathematical definition of an orthonormal wavelet and of the integral wavelet transform.

In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time).

The most commonly used set of discrete wavelet transforms was formulated by the Belgian mathematician Ingrid Daubechies in 1988. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale. In her seminal paper, Daubechies derives a family of wavelets, the first of which is the Haar wavelet. Interest in this field has exploded since then, and many variations of Daubechies' original wavelets were developed.

4.1.1 IMPORTANCE OF WAVELET TRANSFORM

The wavelet transform has become a useful computational tool for a variety of signal and image processing applications. For example, the wavelet transform is useful for the compression of digital image files; smaller files are important for storing images using less memory and for transmitting images faster and more reliably. The FBI uses wavelet transforms for compressing digitally scanned fingerprint images. NASA's Mars Rovers used wavelet transforms for compressing images acquired by their 18 cameras. The wavelet-based algorithm implemented in software onboard the Mars Rovers is designed to meet the special requirements of deep-space communication. In addition, JPEG2K (the newer JPEG image file format) is based on wavelet transforms. Wavelet transforms are also useful for 'cleaning' signals and images (reducing unwanted noise and blurring). Some algorithms for processing astronomical images, for example, are based on wavelet and wavelet-like transforms.

The wavelet bases ψ(x) are functions generated from a kernel function ψ(x) called the mother wavelet, by dilation and translation as follows:

$$\psi_{2^s,t}(x) = \psi_{2^s}(x - 2^{-s}t) = 2^s\Psi(2^s x - t)$$

where s and t are dilation and translation parameters. The mother wavelet _ ψ(x) must satisfy the condition that either:

$$\int \Psi(x)dx = 0.$$

(or)

$$\int \frac{|\hat{\Psi}(w)|^2}{|w|} dw < \infty.$$

Where $\hat{\Psi}(w)$ is the Fourier transform of $\Psi(x)$

Wavelet decomposition is performed by convolving the signal f(x) with the wavelet bases ψ(x); that is,

$$\langle f(x),\ \psi_{2^s,t}(x)\rangle = \int f(x)\psi_{2^s,t}(x)dx.$$

Many different wavelet bases have been constructed and correspond to multi resolution analyses. In multi resolution analysis, let $\phi(x)$ and $\psi\ (x)$ be the scaling function and associated mother wavelet; the computation of $<f(x),\ \phi_2{}^s{,}t(x)>$ is then formulated as follows:

$$\langle f(x),\varphi_{2^{s+1},t}(x)\rangle = \sum_k h(k-2t)\langle f(x),\ \varphi_{2^s,k}(x)\rangle.$$

and

$$\langle f(x),\psi_{2^{s+1},t}(x)\rangle = \sum_k g(k-2t)\langle f(x),\varphi_{2^s,k}(x)\rangle.$$

where $h(t) =< \phi_2{}^{s+1},\ \phi_2{}^s,{}_t>$ is a low-pass filter, $g(t) =< \psi_2{}^{s+1},\ \phi_2{}^s,{}_t>$ is a high-pass filter and $g(t) = (-1)^t h(1-t)$. We could find that $<f(x),\ \phi(x)>$ at 2^{S+1} scale could be achieved by passing $<f(x),\ \phi(x)>$ at 2^S scale through the low-pass filter $h(t)$. Similarly, we could find that $<f(x),\ \psi(x)>$ at 2^{S+1} scale could be achieved by passing $<f(x),\ \psi(x)>$ at 2^S scale through high-pass filter $g(t)$. On the other hand, to reconstruct $f(x)$ at $2S$ scale, $<f(x),\ \phi_2{}^S{}_{,t}(x)>$ is defined as:

$$\langle f(x),\ \varphi_{2^s,t}(x)\rangle = \sum_t (\bar{h}(k-2t)\langle f(x).\varphi_{2^{s+1},t}(x)\rangle$$

$$+\ \bar{g}(k-2t)\langle f(x).\psi_{2^{s+1},t}(x)\rangle).$$

Where $\bar{h}(t)$ and $\bar{g}(t)$ are the associated reconstruction pair of $h(t)$ and $g(t)$. We could find that the reconstruction of $f(x)$ at 2^S scale could be achieved by summing the results after low-pass filtering and high-pass filtering the signal at 2^{S+1} level from the previous decomposition stage. If the wavelet bases are orthogonal, then we could achieve an exact reconstruction by setting $h(t) = \bar{h}(t)$, and $g(t)= \bar{g}(t)$. If the fast computation is expected, the biorthogonal decomposition bases can be used to achieve the exact reconstruction by setting $\bar{g}(t) = (-1)^t h(1-t)$ and $g(t)$ $(-1)^t h(1- t)$.

A 2-D sub-band filter scheme is used when applying the wavelet transform for image applications. The figure below depicts an example of the 2-D sub-band scheme. First, a 1-D decomposition filter is applied to each row of the image signal and then two filtered sub images, L and H, are generated by down-sampling. The size of each sub-image is half of the original image size. The same 1-D decomposition filter is applied to each column of sub images L and H, and four sub images, LL, LH, HL and HH, are generated by down-sampling. The size of each sub-image is one quarter of the original image size.

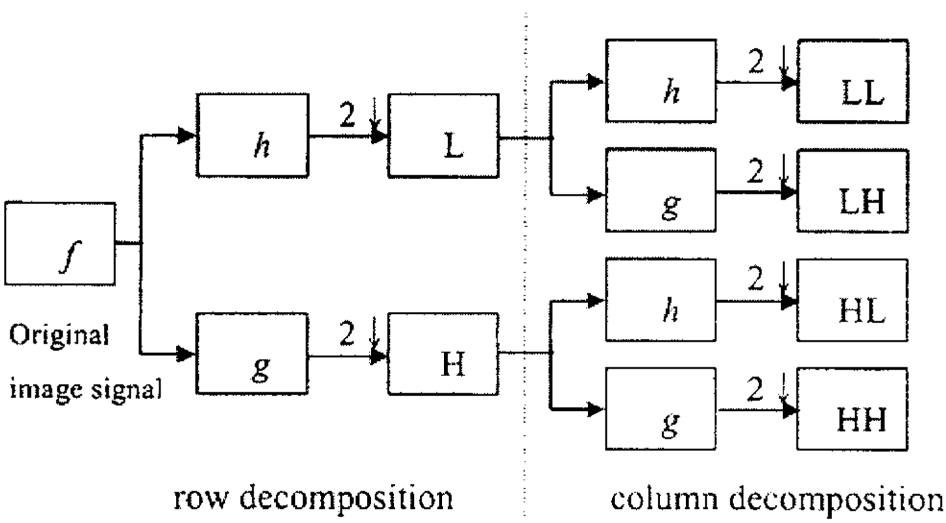


Fig.1 An example of 2-D representation of the wavelet decomposition.

4.1.2 ADVANTAGES OF USING WAVELET TRANSFORM

- Wavelet transform of a function is the improved version of Fourier transform.
- Fourier transform is a powerful tool for analyzing the components of a stationary signal. But it is failed for analyzing the non stationary signal where as wavelet transform allows the components of a non-stationary signal to be analyzed.

4.2 SEGMENTATION

In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). 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.[1] Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) 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 visual 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 are similar with respect to some characteristic or computed property, such as color, intensity, or texture.

Adjacent regions are significantly different with respect to the same characteristic(s). When applied to a stack of images, typical in Medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like Marching cubes.

4.2.1 CLASSIFICATION

Segmentation is a vast area and there are many algorithms can be used for segmentation. A broad classification is given below.

- Thresholding
- Clustering methods
- Compression-based methods
- Histogram-based methods
- Edge detection
- Region growing methods
- Split-and-merge methods
- Partial differential equation-based methods
  - Level set methods
- Graph partitioning methods
- Watershed transformation
- Model based segmentation
- Multi-scale segmentation
  - One-dimensional hierarchical signal segmentation
  - Image segmentation and primal sketch
- Semi-automatic segmentation
- Neural networks segmentation

4.2.2 SEGMENTATION OF TUMORS FROM THE ROI IMAGE

An ROI image is an appropriate image containing the important contents in which we are interested. Especially, it is more useful to make use of the ROI images or medical image processing.

An ROI image is also known as the segmented image which varies from the original image or it is only a part of the original image.

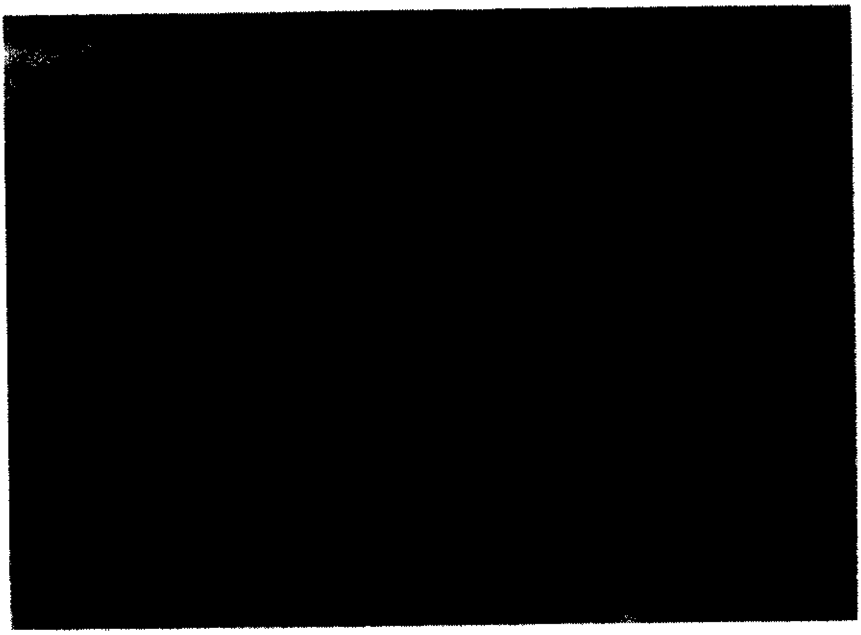


Fig.2 Ultrasound image of a tumor

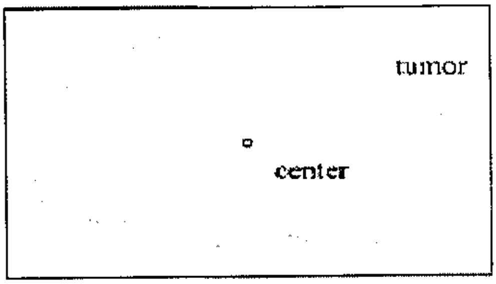
From the image it can be clearly found that the tumor is obviously exhibited in the central region of the image. For the ROI image, it is clear that color, brightness and texture characteristics of the tumor are apparently distinguishable from those of surrounding healthy tissues.

However, there still exists a considerable number of nontumor regions around the tumor in an ROI image. In this case, the result would reduce the accuracy rate of diagnosis for CADx systems.

Therefore, we propose a segmentation method for the purpose of removing the nontumor regions from an ROI image.

The segmentation algorithm applied in the ROI images to segment the tumor is described as follows.

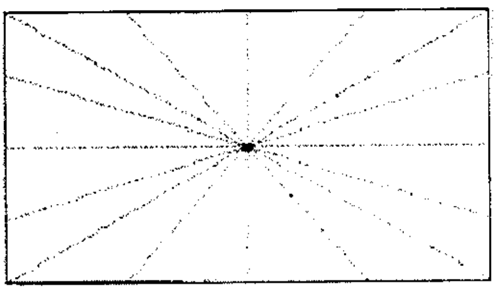
Step 1: For the reason that an ROI image of a tumor is selected by physician, the center of the tumor is expected to approximate the center of the ROI image.



Step 1: Define the center of the tumor.

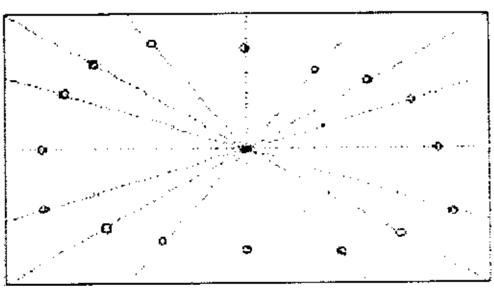
Hence, we assume that the center of the ROI image is the center of the tumor.

Step 2: A number of radial lines are depicted from the approximated center of the tumor defined previously to the boundary of the ROI image.



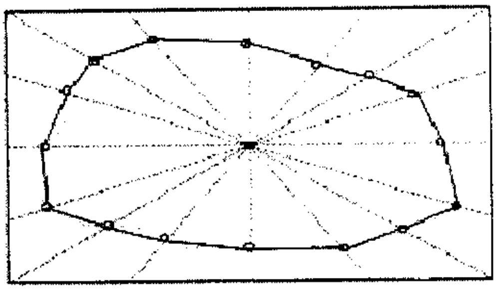
Step 2: Depict a number of radiate lines.

Step 3: Along with each radial line, we find a critical point that approximates the boundary of the tumor and surrounding tissues.



Step 3: Find critical points.

Step 4: Finally, the critical points of the adjacent radial lines are connected to depict the contour of the tumor.



Step 4: Connect critical points.

The major difficulty of the proposed segmentation algorithm is the way of finding the critical points along with the radial lines at step 3. We describe the method of detecting critical points as follows.

Step 1: Apply the wavelet transform on ROI image to obtain four subband images.

Step 2: For the LL and HL subband images, evaluate the associated local variance for each pixel by sliding a variance-evaluating window throughout the whole image. The local variance is defined so as to evaluate the variance in a predefined $m \times n$ rectangle with the currently processed pixel in the center.

Step 3: Scan the pixels along with the radial line from the center of the tumor to the image boundary to find the pixel whose local variance in LL or HL subband is maximal. The maximum pixel is then considered the most likely boundary point of the tumor because of its high local variance. One thing we should note carefully is that we must avoid some nonboundary high-complexity areas inside the tumor being misjudged as boundary points. To avoid such misjudgment, we make some modifications during the scanning procedure. If there exists a pixel scanned after the currently found maximum point whose local mean is larger than the average of all local means and local variance is greater than average of all local variance, the pixel is then considered still to be inside the tumor. This inside point is defined as the new maximum point and the scanning procedure continuously works for the pixels in the radial line.

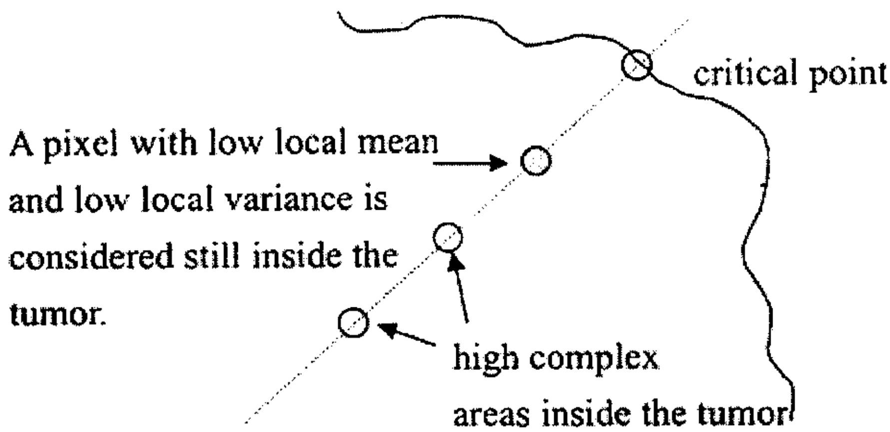


Fig. 4 Finding the boundary of the tumor.

Step 4: After scanning all the pixels in the radial line, the maximum point is considered to be the critical point of this radial line. The main reason we adopt wavelet transform for the segmentation procedure is that wavelet transform could reinforce local characteristics of textures, and the local variance of wavelet coefficients is useful to extract local texture characteristics.

Meanwhile, the local texture characteristics are the most important properties to differentiate two different regions in an image. In our method, the LH subband is not used because of massive noise and the HH subband is also not used because signals are weak and ambiguous.

Hence the LL and HL subbands are used for further computations.

4.3 FEATURE EXTRACTION

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector).

Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved.

Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples.

Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

4.3.1 **FEASIBLE FEATURES FOR DIAGNOSIS**

The features that are extracted for classification are as follows

1. Mean
2. Variance contrast
3. Autocorrelation contrast
4. Distribution distortion of wavelet coefficients

4.3.1.1 **MEAN**

The mean of a data set is simply the arithmetic average of the values in the set, obtained by summing the values and dividing by the number of values. With this in mind, it is natural to define the mean of a frequency distribution by

$$\mu = \frac{1}{n}\sum_{i=1}^{n} f_i x_i = \sum_{i=1}^{n} p_i x_i$$

The mean is a measure of the center of the distribution. As you can see from the algebraic formula, the mean is a weighted average of the class marks, with the relative frequencies as the weight factors. We can compare the distribution to a mass distribution, by thinking of the class marks as point masses on a wire (the x-axis) and the relative frequencies as the masses of these points. In this analogy, the mean is literally the center of mass--the balance point of the wire.

The mean of an image can be calculated by using the grey level matrix and hence the value is used for further computations.

4.3.1.2 VARIANCE CONTRAST

4.3.1.2.1 **VARIANCE**

The variance of a data set is the arithmetic average of the squared differences between the values and the mean. On summarizing a data set in a frequency distribution, we are approximating the data set by "rounding" each value in a given class to the class mark. Thus, the variance of a frequency distribution is given by

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \mu)^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

The standard deviation is the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

The variance and the standard deviation are both measures of the spread of the distribution about the mean. The variance is the nicer of the two measures of spread from a mathematical point of view, but as you can see from the algebraic formula, the physical unit of the variance is the square of the physical unit of the data. For example, if the variable represents the weight of a person in pounds, the variance measures spread about the mean in squared pounds. On the other hand, standard deviation measures spread in the same physical unit as the original data, but because of the square root, is not as nice mathematically. Both measures of spread are useful.

The relative frequency distribution as the probability distribution of a random variable X that gives the mark of the class containing a randomly chosen value from the data set. With this interpretation, the variance and standard deviation of the frequency distribution are the same as the variance and standard deviation of X.

4.3.1.2.2 VARIANCE CONTRAST

But in our case, since we have segmented the tumor region from the surrounding tissues, the original image could be divided into two regions: inside region (the tumor) and outside region (the surrounding tissue).

Thus, variance contrast is defined as the ratio of variance of the inside region and variance of the outside region.

Variance of a region is defined as

$$\sigma = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}.$$

Thus variance contrast can be written as

$$\sigma_{inside} / \sigma_{outside}$$

4.3.1.3 AUTOCORRELATION CONTRAST

4.3.1.3.1 NORMALISED CROSS CORRELATION

The brightness of the image and template can vary due to lighting and exposure conditions, the images can be first normalized. This is typically done at every step by subtracting the mean and dividing by the standard deviation. That is, the cross-correlation of a template, $t(x,y)$ with a subimage $f(x,y)$ is

$$\frac{1}{n-1} \sum_{x,y} \frac{(f(x,y) - \overline{f})(t(x,y) - \overline{t})}{\sigma_f \sigma_t}.$$

where n is the number of pixels in t(x,y) and f(x,y), $\overline{f}$ is the average of f and σ_f is standard deviation of f. In functional analysis terms, this can be thought of as the dot product of two normalized vectors. That is, if

$$F(x,y) = f(x,y) - \overline{f}$$

and

$$T(x,y) = t(x,y) - \overline{t}$$

then the above sum is equal to

$$\left\langle \frac{F}{\|F\|}, \frac{T}{\|T\|} \right\rangle$$

where $\langle \cdot, \cdot \rangle$ is the inner product and $\| \cdot \|$ is the L^2 norm. Thus, if f and t are real matrices, their normalized cross-correlation equals the cosine of the angle between the unit vectors F and T, being thus 1 if and only if F equals T multiplied by a positive scalar.

Normalized correlation is one of the methods used for template matching, a process used for finding incidences of a pattern or object within an image.

4.3.1.3.2 AUTOCORRELATION

Autocorrelation is the cross-correlation of a signal with itself. Informally, it is the similarity between observations as a function of the time separation between them. It is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal which has been buried under noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies. It is often used in signal processing for analyzing functions or series of values, such as time domain signals.

Let X be some repeatable process, and i be some point in time after the start of that process. (i may be an integer for a discrete-time process or a real number for a continuous-time process.) Then X_i is the value (or realization) produced by a given run of the process at time i. Suppose that the process is further known to have defined values for mean μ_i and variance σ_i^2 for all times i. Then the definition of the autocorrelation between any two time s and t is

$$R(s,t) = \frac{\mathrm{E}[(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \sigma_s},$$

where "E" is the expected value operator.

4.3.1.3.3 GREY LEVEL CO-OCCURANCE MATRIX(GLCM)

A co-occurrence matrix, P ϕ,d(i, j), is a matrix where the (i,j)th element describes the frequency of occurrence of two pixels that are separated by distance d in the direction _ with grey levels i and j. For an $N \times N$ region with M grey levels $(0, \ldots, M\_1)$, the grey-level differences that single pairs of pixels can exhibit are given below

$$
\begin{array}{cccc}
(0,0) & (0,1) & \dots & (0,M-1) \\
(1,0) & (1,1) & \dots & (1,M-1) \\
\vdots & \vdots & \vdots & \vdots \\
(M-1,0) & (M-1,1) & \dots & (M-1,M-1)
\end{array}
$$

We can find that the co-occurrence matrix can capture the texture variations in a region by various ϕ and d. The main power of the co-occurrence matrix approach is that it characterizes the spatial interrelationships of the grey levels in a texture pattern and it is invariant under monotonic grey-level transformations. In general, a minimum set of co-occurrence matrices is four (ϕ =0, 45, 90 and 135°, respectively; d =1) for the texture about which we have no prior knowledge.

However, it is obvious that the size of the matrix is a function of the number of grey levels in the image, and it would be prohibitively expensive to evaluate matrix for each pixel in a general 8-bit image (256×256 elements in a co-occurrence matrix). Thus, we use the statistical parameter matrices with the same texture reservation properties, instead of the cooccurrence matrix. The statistical method we adopted is 2-D autocorrelation, which can evaluate the texture parameters for several distances between pixels and directly from the image without using co-occurrence matrices. The main advantage of this method resides in its calculation cost, which depends only on the size of the image treated and not on the number of grey levels. Moreover, it allows the extraction of visually perceptible physical parameters from the image, such as contrast, granularity, regularity, periodicity, finesse or coarseness of the texture, and so on. The 2-D autocorrelation coefficients A(m, n) between pixel (x, y) and pixel (x +Δm, y+Δn) in an image with size $m \times n$ is defined as:

$$
A(\Delta m,\Delta n) = \frac{1}{(m-\Delta m)(n-\Delta n)} \sum_{x=0}^{m-1-\Delta m} \sum_{y=0}^{n-1-\Delta n} f(x,y) \times f(x+\Delta m, y+\Delta n).
$$

In general, Δm and Δn are set as the distances among pixels whose interrelationship is to be taken into consideration. Thus, there will be $\Delta m \times \Delta n$ coefficients.

4.3.1.3.4 **AUTOCORRELATION CONTRAST**

The autocorrelation contrast is then defined as the ratio of autocorrelation of inside the region and autocorrelation of the outside region. That is, the autocorrelation contrast is

$$A(\Delta m, \Delta n)_{inside} / A(\Delta m, \Delta n)_{outside}.$$

4.3.1.4 DISTRIBUTION DISTORTION OF WAVELET COEFFICIENTS

4.3.2.4.1 **LAPLACIAN DISTRIBUTION**

It has been shown that, for a large class of images, the transformed coefficients in each high-frequency subband can be well described by a generalized Laplacian distribution p(x) given by

$$p(x) = \frac{\lambda}{2} \cdot e^{-\lambda|x|}.$$

whose mean is zero and variance is $2/\lambda^2$.

The histogram of HL subband wavelet coefficients of an image and Laplacian distribution is shown in figure

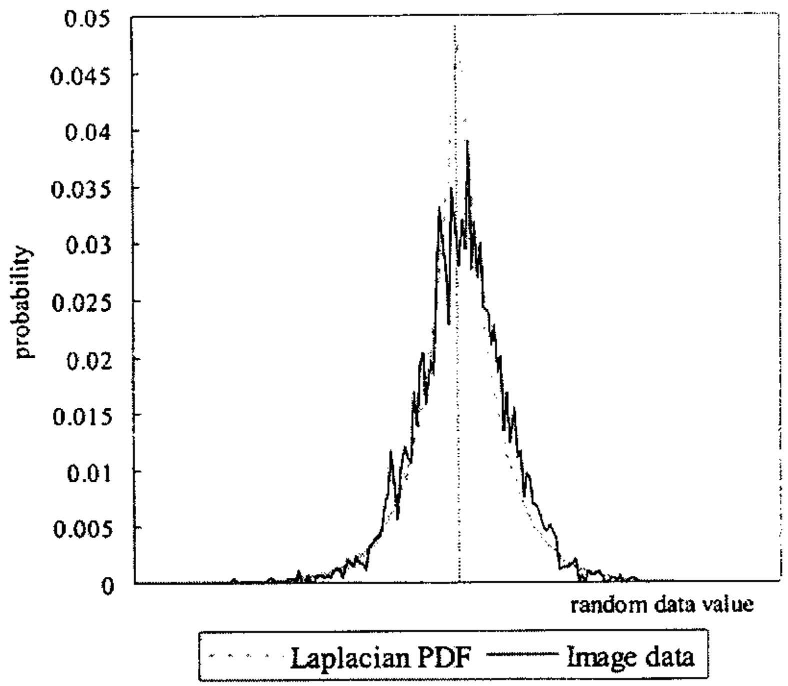


Fig.5 Histogram of laplacian distribution

4.3.1.4.2 DISTRIBUTION DISTORTION OF WAVELET COEFFICIENTS

The summation of differences among the real distribution of wavelet coefficients and the distribution of expected Laplacian distribution is adopted as our feature. That is, the distribution distortion of wavelet coefficients is defined as

$$\sum_{x} |w(x) - p(x)|.$$

Where w(x) is the real probabilitydensity function of HL subband wavelet coefficients and p(x) is the expected Laplacian distribution.

The proposed three features are extracted synchronously in the wavelet decomposition stage. Variance contrast and autocorrelation contrast are extracted from the LL subband and the distribution distortion of wavelet coefficients is extracted from the HL subband.

CHAPTER-5

RESULTS

ORIGINAL IMAGE

The below figure depicts the ultrasound image to be given as input to the prescribed algorithm.

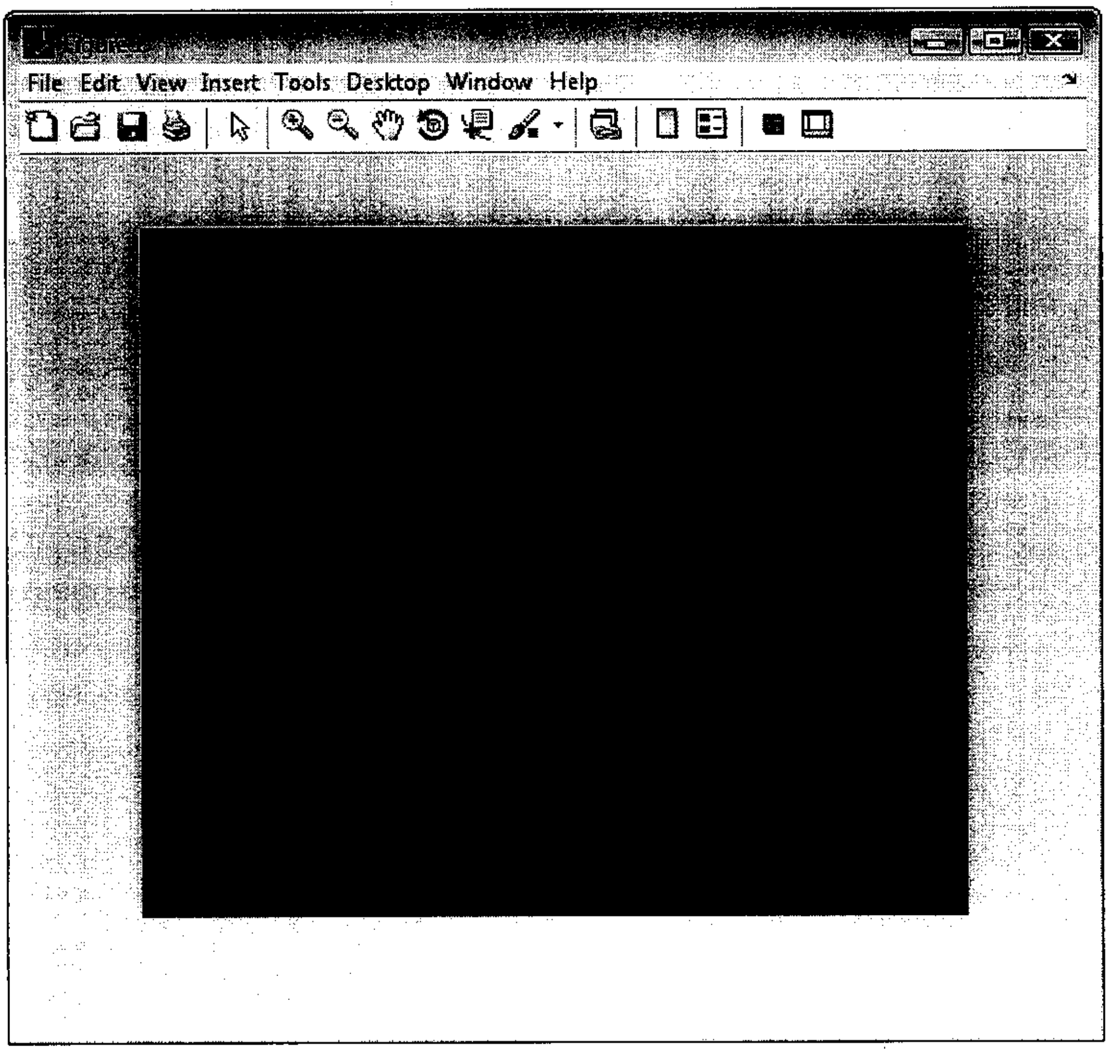


Fig.6 Ultrasound image to be classified

RESULT OF WAVELET TRANSFORM

On the application of wavelet transform, four subbands are obtained due to low pass and high pass filters. The details obtained are divided into approximation, horizontal, vertical and diagonal which are LL, LH, HL and HH subbands respectively.

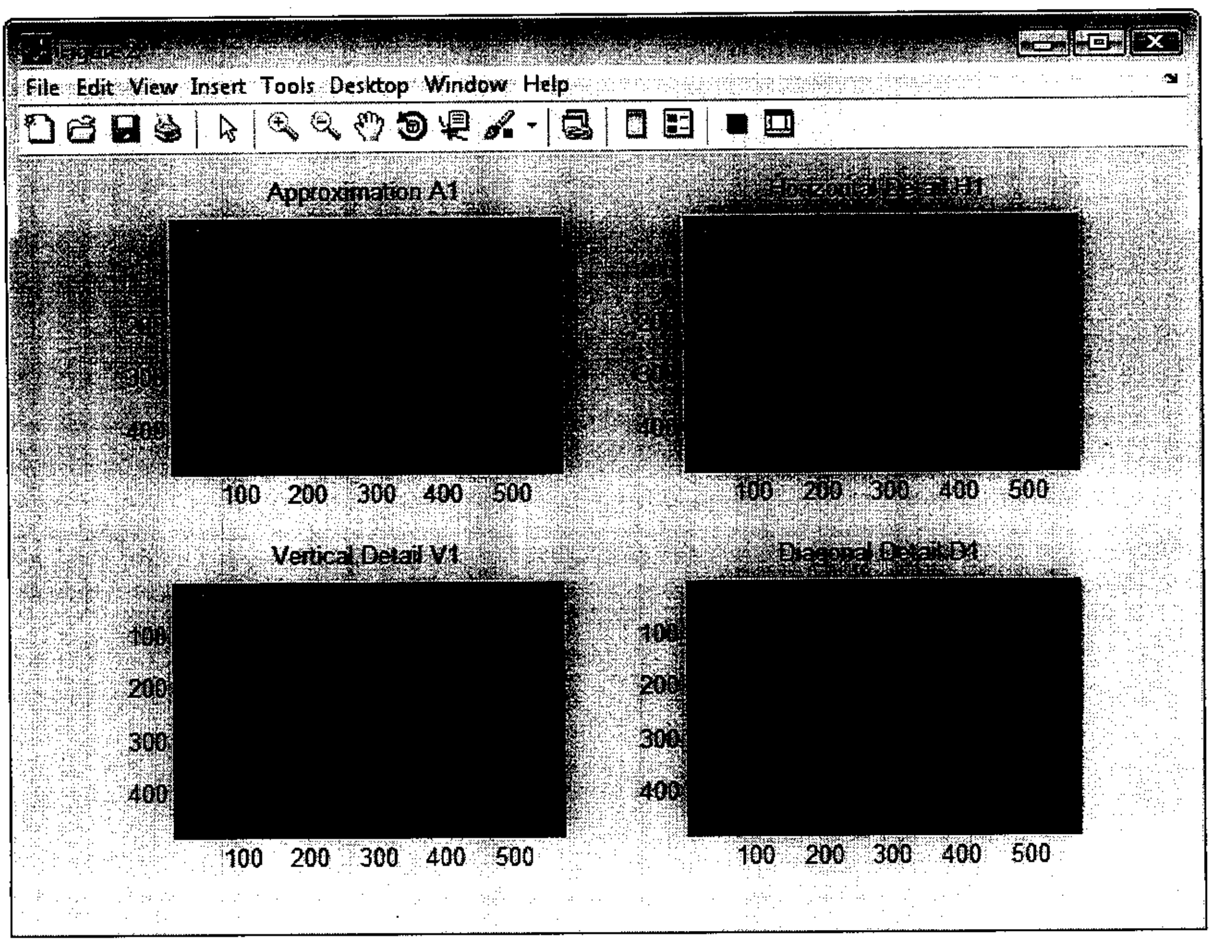


Fig.7 Subbands after applying wavelet transform

RESULT OF SEGMENTATION

The approximation subband image is considered and hence the process of segmentation is carried out initilising the area inside the tumor region.

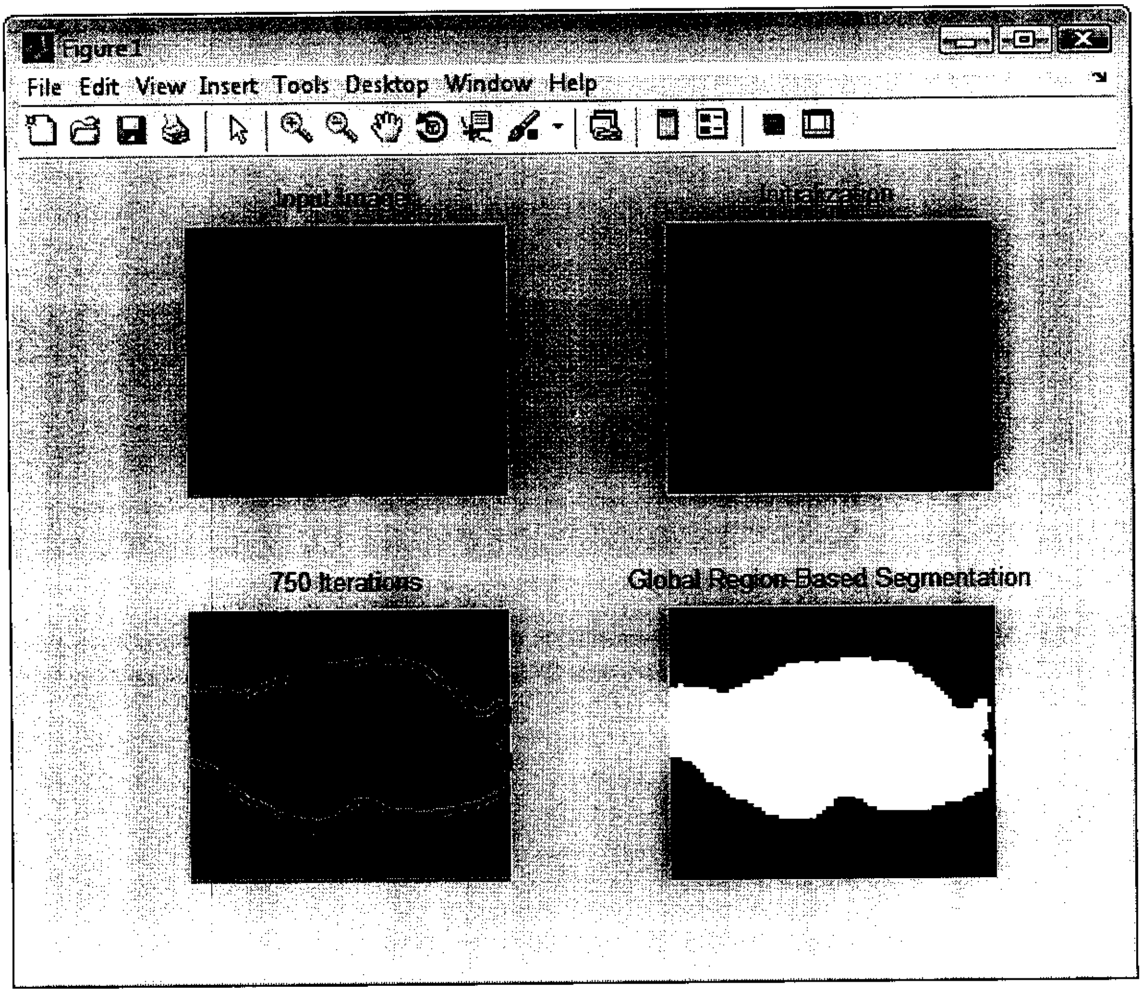


Fig.8 Result of segmentation

IMAGE / FEATURES	1	2	3
MEAN CONTRAST	0.38223285	0.14918974	0.15316761
VARIANCE CONTRAST	0.68595181	0.57982214	0.74117010
AUTOCORRELATION CONTRAST	0.30953265	0.10925755	0.07942402

Table 2. Features of benign tumors

IMAGE / FEATURES	1	2	3
MEAN CONTRAST	0.19205675	0.22014507	0.37577963
VARIANCE CONTRAST	0.04862761	0.11999332	0.18366256
AUTOCORRELATION CONTRAST	0.24548476	0.23505532	0.21664598

Table 3. Features of malignant tumors

Fig.9 Binary image of the region outside the tumor

Fig.10 Binary image of the region inside the tumor

CONCLUSION

The texture characteristics of an ultrasound image impose a great effect in the classification of tumor. Though there are several other features, the above features are solely enough for the classification of a tumor as benign or malignant. The importance of wavelet transform was clearly understood from the usage of the subbands produced as a result of it.

Thus the proposed segmentation and classification algorithm reduces the effort of the physician by automating the process, providing high accuracy. And finally the tumor is classified without extracting a tissue from the examinee or passing high frequency radiation through him.

5.1. FUTURE WORKS

A multilayered perceptron (MLP) neural network trained using error back - propagation algorithm with momentum can be used for the differential diagnosis of breast tumors on sonograms.

The main goal of an image-recognition system using neural networks is to achieve the purpose that evaluates the input signals to the desired outputs correctly. If the input feature vectors are chosen properly, a well trained MLP neural network will achieve good performance for image recognition by its nonlinear, stable and high-dimensional segmentation properties.

REFERENCES

BOOK REFERENCE

1. "Digital Image Processing Using MATLAB" by Gonzalez (Rafael C); Woods
2. (Richard E); Eddins (Steven L)
3. "Digital Image Processing Using Matlab" by Jain (Anil K); Brooks/Cole in
4. 2000
5. "Digital Signal Processing Using Matlab" by Ingle Vinay K. ;Proakis G. John
6. "Matlab For Engineers" by Biran Adrian ;Breiner Moshe, Addison Wesley
7. "Introduction To Neural Networks" by Anderson (James A), Prentice Hall Of India (P) Ltd in 1995.
8. "Introducing Neural Networks" by Carling (Alison)
9. P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," in Proc. IEEE Workshop Comput. Vision, Miami, FL, Nov.1987, pp. 16-22.
10. D. R. Chen, R. F. Chang, and Y. L. Huang, "Computer-aided diagnosis applied to US of solid breast nodules by using neural networks," Radiology, vol. 213, pp. 407–412, 1999
11. "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks" by Dar-Ren Chen, Reuy-Feng Chang, Wen-Jia Kuo and Yu-Len Huang
12. "Feature Extraction Foundations and Applications" , Isabelle Guyon, Masoud Nikravesh, Steve Gunn and Lotfi A. Zadeh

WEBSITE REFERENCE

1. http://www.mathworks.com/matlabcentral/
2. http://www.medphys.org/
3. http://www.breastcancer.org/
4. http://www.radiologyinfo.org/
5. http://ieeexplore.ieee.org/
6. http://www.sciencedirect.com/
7. http://www.springer.com/
8. http://www.wikipedia.org/