بسم الله الرحمن الرحيم

Sudan University of Science and Technology

College of Engineering

Biomedical Engineering Department

A dissertation submitted in partial fulfillments of the requirements for the B.Sc. Degree in Biomedical Engineering

Breast Cancer Detection in Mammographic Image

Presented By:

Dania AwadHussien Taber.

RahmaAbdulmoniem Hassan Elfadil.

Shaima Mahmoud Mustafa Onsa.

Supervisor:

Dr. /Alnazeir Osman Mohamed.

August -2014

.

Abstract

Breast cancer is one of the major causes of death in women when compared to all other cancers. This research was developed methods that can be used to detect abnormal breast tissues. Matlab programs were devolved to detect abnormal tissue in mammograms, to assist radiologist in their diagnosis as a second opinion. In this study100 image consist 100 normal ROIs, 100 abnormal ROIs (contained 61 benign ROIs and 39 ROIs malignant ROIs) was extracted from digital mammogram which obtained from mini-MAIS database. Five First order statistical features were calculated for each ROI. Fourteen Haralick features according to GLCM matrix was calculate for each ROI. The best features were selected by their ability to distinguish between normal and abnormal tissue to use it in program to detect breast cancer.

المستخلص

سرطان الثدي واحد من أكبر أسباب وفاة النساء في العالم مقارنة مع باقي السرطانات .هذا البحث طور طرق تستخدم لعمل كشف لخلايا الثدي المريضة. برامج ماتلاب تم تطويرها لعمل كشف الخلايا الغير طبيعية في صور الماوقرام الرقمية لمساعدة أخصائي الأشعة في تشخيصه كخيار ثاني. في هذه الدراسة تم استخدام 100 صورة تحتوي على 100 عينة سليمة و100 عينة مريضة (تحتوي61 عينة حميدة و39 عينة خبيثة) MIAS) تم استخراجها من صور الماموقرام الرقمية التي تحصل عليها من جمعية تحليل صور الماموقرام (Haralick) .خمس معادلات من الدرجة الأولى تم حسابها لكل العينات .أربعة عشر من معادلات هارليك ( ) تم حسابها لكل العينات. أفضل المعادلات تم اختيارها على GLCMالتي تعتمد على المصفوفة الأحصائية (أساس قدرتها على التمييز بين الخلايا السليمة والمريضة .وأستخدمت في برنامج للكشف عن سرطان الثدي.

Table of Contents

Dedication	I
Acknowledgments	II
Abstract	III
المستخلص	IV
Contents	V
List of Tables	VII
List of Figures	VIII
Chapter One: Introduction	
1.1General view	1
1.2Problem statement	2
1.3Objective	2
1.3.1General objective	2
1.3.2Specific objective	2
1.4Thesis Layout	2
Chapter two : Theoretical background	
2.1Breast Anatomy	3
2.2Breast Tumor	5

2.2.1Benign Tumor 5

2.2.2Malignant Tumor 5

2.3Breast Calcification 6

2.4Types of breast cancer 6

2.5Mammography 7

2.5.1Film Mammography 8

2.5.2Digital Mammography 8

2.6Computer-aided Detection (CAD) system 9

2.7Texture Analysis 12

2.8Features Extraction 13

2.8.1First order features 14

2.8.2Second order features (Haralick) 15

2.9Matlab and image processing toolbox 18

Chapter three : Literature Reivew

Literature Review 20

Chapter four : Methodology

4.1 Region of interest (ROI) selection 23

4.2Feature Extraction 23

4.2.1First order statistical features 23

4.2.2Second order statistical features (Haralick) 24

4.3 Features Selecation 25

4.4 Algorithm for detection abnormality 26

Chapter five: Result and Discussion

5.1 Result & Discussion 28

Chapter six : Conclusion and Recommendation

6.1 Conclusion 44

6.2 Recommendation 44

Appendix 45

Refrences 50

List of Tables

Table (2.1) Shows second order (haralick) features equations.

List of Figures

Figure(2.1) shows the anatomy of female breast.

Figures(2.2) Shows the Digital Mammography.

Figure(2.3) shows the steps of CAD system to diagnosis abnormality of mammogram.

Figure (4.1) A-Shows how to calculate GLCM matrix from ROI in distance 1 and direction 0°.

B-Shows the four directions of relationships between two pixels.

Figure(4.2) shows the flow chart of breast cancer detection.

Figure (5.1) shows first order features for 100 images (100 ROIs normal and 100 ROIs abnormal (61 benign , 39 malignant))using window 20*20 pixels.

Fig (5.2) shows second order (Haralick) features for 100 images (100 ROIs normal and 100 ROIs abnormal (61 benign, 39 malignant))using window 20*20 pixels and direction of GLCM 0°.

Figure (5.3) shows first order features for 100 images (100 ROIs normal and 100 ROIs abnormal (61 benign , 39 malignant))using window 40*40 pixels.

Fig (5.4) shows second order (Haralick) features for 100 images (100 ROIs normal and 100 ROIs abnormal (61 benign, 39 malignant))using window 40*40 pixels and direction of GLCM 0°.

CHAPTER ONE

Introduction

1.1 General View:

Breast cancer is one of the major causes of death in women when compared to all other cancers. Studies have demonstrated that African women are more prone to get breast cancer at an early age as compared to Western women. One study showed a high percentage occurred in Sudanese women less than 50 years of age showing a clear indication of early onset of the disease in younger women. This is especially true in the case of Sudanese women. Early detection is the key to control the breast cancer and decrease the mortality rate. Mammography is one of the imaging modality in early breast cancer detection typically through detection of characteristic masses and micro calcifications. There is variation and inconsistent between radiologists in diagnose the breast cancer. The radiologist may lose part of the tumor or faulting in determining the location, number and size of the tumor [1].

Computer-aided detection (CAD) tool developed to aid radiologist in detectingtumors. CAD system could act as a second opinion. it also offer better accuracy assists in reducing missed cancers and provide better prognosis [2] .There are many technique that used to developing CAD system. One of these techniques is features extraction from Regions of Interest (ROI) of Mammograms. Several types of features extraction from digital mammograms including position feature, shape feature and Texture feature. Textures are one of the important features used for description, analysis and classification of ROI in Mammograms .The texture features are ability to distinguish between abnormal and normal cases. Texture measures are two types, first order and second order. In the first order, texture measure are statistics calculated from an individual pixel and do not consider pixel neighbor relationships . Intensity feature are first order texture calculation [3]. In the second order, measures consider the relationship between neighbor pixels GLCM is a second order texture calculation [4]. Texture features has been extracted and used the best features that can help to detection abnormal breast cancer and distinguish between normal and abnormal breast tissues.

1.2 Problem Statement:

There is variation and inconsistent between radiologists in diagnose the breast cancer due to factoversvision,experience and skill ofthe Radiologist, the image qualityand the highamount ofcases thatare being testedanddiagnosedduring the day .The radiologist may lose part of the tumor or faulting in determining the location, number and size of the tumor.

1.3 Objectives:

1.3.1 General aim:

Developing CAD system that used as a second opinion of diagnosing Breast Cancer.

1.3.2 Specific objectives:

1. Collect mammogram Data.
2. Design code to read image and select ROI .
3. 1^{st} order Feature extraction from ROIs.
4. Gray Level Co-occurrence matrix (GLCM) calculate and Normalization of GLCM.
5. 2^{nd} order feature extraction (Haralick).
6. Choose the best features to distinguish between normal and abnormal tissues

1.4 Thesis Layout:

This project contains six chapters. Chapter one contains an introduction to the project, chapter two about literature review that related with the project , while chapter three contains a theoretical background for the project components , chapter four contains materials and methods used to achieve the objectives of the project , the results and discussion are found in chapter five ,and finally chapter six contains conclusion and recommendation.