COMPUTER-AIDED BREAST CANCER DETECTION AND DIAGNOSIS FROM DIGITAL MAMMOGRAMS

By

Mugahed Ali Shawqi Al-antari

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Biomedical Engineering and Systems

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2015

COMPUTER-AIDED BREAST CANCER DETECTION AND DIAGNOSIS FROM DIGITAL MAMMOGRAMS

By

Mugahed Ali Shawqi Al-antari

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Biomedical Engineering and Systems

Under the Supervision of

Prof. Dr. Yasser M. Kadah

……..

Professor of Biomedical Engineering
Biomedical Engineering and Systems
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2015

COMPUTER-AIDED BREAST CANCER DETECTION AND DIAGNOSIS FROM DIGITAL MAMMOGRAMS

By

Mugahed Ali Shawqi Al-antari

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Biomedical Engineering and Systems

Approved by the
Examining Committee:

Prof. Dr. Yasser M. Kadah Professor of Biomedical Engineering, Faculty of Engineering, Cairo University	Thesis Main Advisor
Prof. Dr. Ahmed M. Badawi Professor of Biomedical Engineering, Faculty of Engineering, Cairo University	Internal Examiner
Prof. Dr. Mohamed I. El-Adawy Professor of Electronic Engineering, Faculty of Engineering, Helwan University	External Examiner

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2015

Engineer's Name: Mugahed Ali Shawqi Al-antari
Date of Birth: 03 / 02 / 1985
Nationality: Yemeni
E-mail: en.mualshz@gmail.com
Phone: +201152622433/+967734502097
Address: 30 Hassan Rakha St., Faisal, Giza

Registration Date: 01 / 10 / 2012
Awarding Date: … /… / ……
Degree: Master of Science
Department: Biomedical Engineering and Systems

Supervisors: **Prof. Dr. Yasser M. Kadah**

Examiners:

Prof. Dr. Yasser M. Kadah (Thesis Main Advisor)

Professor of Biomedical Engineering and Systems, Faculty of Engineering, Cairo University.

Prof. Dr. Ahmed M. Badawi (Internal Examiner)

Professor of Biomedical Engineering and Systems, Faculty of Engineering, Cairo University.

Prof. Dr. Mohamed I. El-Adawy (External examiner)

Professor of Electronic Engineering, Faculty of Engineering, Helwan University.

Title of Thesis:
Computer-Aided Breast Cancer Detection and Diagnosis from Digital mammograms

Keywords:
Computer-Aided Detection; Computer-Aided Diagnosis; Cohen-k factor; Peripheral Equalization; Receiver Operator Characteristic Curve.

Summary:
Computer-aided detection (CADe) and diagnosis (CADx) systems are emerging technologies using to help radiologists to interpret medical images. In this study, first we applied an image processing technique for peripheral region enhancement. Then a CADe and CADx systems are developed and applied to the standard MIAS and DDSM databases for classification of masses. Results have shown that SFS was the best for both CAD systems and almost all of samples were correctly classified with MIAS database, while the DDSM database results shown 98.67% agreement.

ACKNOWLEDGMENT

Thanks to God first and foremost for his blessings on my all life and guided me for seeking knowledge and completing this thesis. Then I would like to express my sincere appreciation to my thesis advisor, Prof. Dr. Yasser M. Kadah for his invaluable time that he spends it in supervision, motivation and guidance and his insightful comments that improved the quality of this thesis.

DEDICATION

I would like to sincerely thank all of my family members and especially my parents for their ceaseless love and support and for instilling a love of learning in myself. I would like to thank my wife and my daughter for enduring a seemingly endless ordeal, for sacrificing some of their best years to completely finish this research. The thesis procedures could not complete successfully without having their patience, support, love, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARCH	Architectural distortion
ASYM	Asymmetry
AUC	Area Under Curve
BI-RADS	Breast Imaging-Reporting and Data System
BI	Blurred Image
BBS	Branch and Bound Selection
CAD	Computer Aided Diagnosis
CADe	Computer Aided Detection
CADx	Computer Aided Diagnosis
CC	Cranio-Caudal
CDF	The cumulative distribution function
CIRC	Well-defined/circumscribed masses
DDSM	The Digital Database For Screening Mammography
FFDM	Full Field Digital Mammography
FDA	Food and Drug Administration
FP	False Positive
FN	False Negative
GLCM	Gray Level Co-Occurrence Matrix
GLPF	Gaussian Low Pass Filter
KS-test	Kolmogorov-Smirnov test
KNN	K-voting Nearest Neighbor
LDA	Linear Discriminant Analysis
MIAS	The Mammographic Image Analysis Society
MLO	Mediolateral-Oblique
MISC	ill-defined masses
NN	Neural Network
NTP	Normalized Thickness Profile
NPV	Negative Predictive Value
PPV	Positive Predictive Value
QDA	Quadratic Discriminant Analysis
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SBS	Sequential Backward Selection
SFFS	Sequential Floating Forward Selection

SFS	Sequential Forward Selection
SI	Segmentation Image
SPIC	Spiculated masses
SVM	Support Vector Machine
T-test	Unpaired student test
TN	True Negative
TP	True Positive
W-test	Wilcoxon rank sum test

Abstract

Breast cancer is the uncontrolled growth of abnormal cells that start in the breast. It was and still the most common cancer diagnosed in women worldwide. Mammography has been successful in improving detection of cancer in early stage, and it continues to be the standard screening tool for breast cancer detection resulting in at least a 30% reduction in breast cancer deaths.

Computer-aided diagnosis (CAD) is a system that uses the output of mammography systems to help the radiologist's decision. It has been defined as an equivalent diagnosis that was made by a radiologist who uses the output of a computer analysis of the images when making his/her interpretation.

In this thesis, computer-aided detection (CADe) and Computer-aided diagnosis (CADx) systems have been developed and applied to the standard MIAS and DDSM databases to distinguish between the different regions of breast tissues. In both CAD systems, first an image processing technique is implemented to enhance the peripheral region of breast as a preprocessing step, then regions of interest (ROI) are excerpted using window of size 32×32 pixels then a set of 422 features are extracted from ROI and normalized. For CADe, the features selection is performed using statistical methods such as t-test, Kolmogorov-Smirnov test (KS-test) and Wilcoxon signed rank test (W-test) and other methods by using Sequential Backward Selection (SBS), Sequential Forward Selection (SFS), Sequential Floating Forward Selection (SFFS) and Branch and Bound Selection (BBS) techniques. Then we used four classifiers including K-voting Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) for classification stage with leave-one-out method for testing. The proposed systems were evaluated using many indices such as overall accuracy, Cohen-k factor, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under curve (AUC) of the ROC curves.

CADx system has the same stages but the output is different where it indicates the likelihood of lesion malignancy. In CADx, SBS, SFS, SFFS and BBS methods are used for selecting the best powerful features. Both CAD systems provided encouraging results. These results were different corresponding to selection method. Finally, we compared independently between performance of all classifiers with each selection method and we concluded that the SFS selection method was the best for both CAD systems.

In both CAD systems, we used Neural Network (NN) classifier with cross validation to try to get better performance as much as possible with each selection method.

In general, with the best feature selection algorithms the results that we obtained on the MIAS dataset show that 100% of samples were correctly classified, while the DDSM dataset results show 98.67% agreement. This is the most important point in this thesis.

Chapter 1 : Introduction

1.1. Overview of the Thesis

Breast cancer was the feared disease before the 20th century as if it was shameful disease. Where breast cancer was the most common cancer diagnosed in women worldwide. However, after a little could be safely done with primitive surgical techniques, women tended to suffer silently rather than discussion care. After that when surgery advanced, and long-term survival rates improved, women began raising awareness of the disease and the possibility of successful treatment [1].

Breast cancer is the uncontrolled growth of abnormal cells that start in the breast, usually in the inner lining of the milk ducts or lobules. There are different types of breast cancer, with different stages, aggressiveness, and genetic makeup. With best treatment, ten years disease-free survival varies from 98% to 10%. Treatment includes surgery, drugs such as hormone therapy and chemotherapy and radiation. Breast cancer is the most common cancer and continues to be a significant public health problem among women around the world. Primary prevention seems impossible since the cause of this disease still remains unknown. It is believed that the most promising way to decrease the number of patient suffering from the disease is by early detection. The earlier breast cancer is detected, the better the chances that treatment will work and the better a proper treatment options can be provided [2].

Among U.S. women, breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer death, following lung cancer. Through two years ago, an estimated 232,340 new cases of invasive breast cancer and 39,620 breast cancer deaths are expected to occur among U.S. women [3].

Mammography has been successful in improving detection of cancer, particularly non-palpable breast masses and calcifications that may be malignant. Mammography continues to be the standard screening tool for breast cancer detection resulting in at least a 30% reduction in breast cancer deaths [4].

There has been some recent contention over the benefit of mammography screening and the available evidence relating mammography screening with mortality may not be definitive. However, a recent Institute of Medicine Report on Mammography (Committee on the Early Detection of Breast Cancer 2001) suggests that the reduction in mortality from breast cancer observed in recent years may be due to earlier detection through mammography screening [5].

By incorporating the expert knowledge of radiologists, the computer-based systems provide a second opinion in detecting abnormalities and making diagnostic decisions. Such a diagnostic procedure is called computer-aided diagnosis (CAD). A computerized system for such a purpose is called a CAD system. It has been shown that the performance of radiologists can be increased by providing them with the results of a CAD system. Hence, there are strong motivations to develop a CAD system to assist radiologists in reading mammograms [6].

1.2. Thesis Objective

The main objective of this thesis is to develop a computer-aided detection (CADe) system and a computer-aided diagnosis (CADx) system by developing algorithm for classification of abnormal lesions in breast cancer to distinguish between normal, benign and malignant cases using different set of features.

This algorithm includes five main stages, pre-processing stage is achieved by using image processing algorithm, Region of Interest (ROI) selected inside the suspicious area of mammogram, features extraction from ROI, features Selection to select the most powerful features and finally classification stage in order to differentiate between normal, benign and malignant group using different classifiers.

We divided the main objective of the thesis into two sub-objectives. The first sub-objective is to develop a CADe system for classifying abnormal lesions in mammograms to distinguish between normal and abnormal lesions. We will develop this system by using seven feature selection algorithms and four classifiers and compare between each classifier with each selection method.

The second sub-objective is to develop a CADx system that will be able to distinguish between normal, benign and malignant breast tissues. This system will has many advantages to support the radiologists' opinion and help them to take a truth decision in short time especially with huge cases of patients.

In each sub-objective we applied image processing technique for peripheral breast tissue enhancement by using Tao Wu' et al. algorithm [33], as a first step for each CAD system, for enhancement the peripheral area of breast. We explained this algorithm in chapter 3. In previous studies the researchers concluded that the results of CAD system are better with pre-possessing stage or peripheral enhancement technique of breast tissue. In the final we compared between the behaviors of all classifiers with each CAD system.