Ministry of Higher Education
and Scientific Research
University of Diyala
College of Science

Breast Cancer Diagnosis based on Fuzzy Probabilistic Neural Network

A Research
Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By
Ala'a Jalal Abdullah

Supervised By

Assist.Prof.Dr. Taha M. Hassan ***Assist.Prof. Dr. Jumana W. Salih***

2019 A.D. **1440 A.H.**

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

وَمَا أُوتِيتُم مِّنَ الْعِلْمِ إِلَّا قَلِيلًا

صدق الله العظيم

سورة الاسراء

آية (85)

Dedication

To...

My lovely family

My Supervisors

My teachers

My friends…

Linguistic Certification

I certify that this research entitled **"Breast Cancer Diagnosis based on Fuzzy Probabilistic Neural Network "** was prepared by **Ala'a Jalal Abdullah** and was reviewed linguistically. Its language was amended to meet the style of English language.

Signature :

Name : Assist.Prof. Dr. Mohammed N. Husain

Date : / / 2019

Supervisor's Certification

We certify that this research entitled **" Breast Cancer Diagnosis based on Probabilistic Neural Network"** was prepared by **Ala'a Jalal Abdullah** under our supervisions at the University of Diyala Faculty of Science Department of Computer Science, as a partial fulfillment of the requirements needed to award the degree of **Master of Science in Computer Science**.

(Supervisor)

Signature:
Name: Assist. Prof. Dr. Taha M. Hassan
Date: / / 2019

(Supervisor)

Signature:
Name: Assist. Prof .Dr. Jumana W. Salih
Date: / / 2019

Approved by University of a Diyala Faculty of Science Department of Computer Science.

Signature:
Name: Assist. Prof. Dr. Taha M. Hassan
Date: **/ / 2019**

(Head of Computer Science Department)

Examination Committee Certification

We certify that we have read this research entitled **"Breast Cancer Diagnosis based on Fuzzy Probabilistic Neural Network "**, and as an examining committee, examined the student **"Ala'a Jalal Abdullah"** in its contents and that in our opinion, it is adequate as fulfill the requirements for the Degree of Master in Computer Science at the Computer Science Department, University of Diyala.

(Chairman)

Signature:
Name: Prof. Naji M. Sahib
Date: / /2019

(Member)

Signature:
Name: Assist. Prof. Dr. Jamal M. Abbas
Date: / /2019

(Member)

Signature:
Name: Assist. Prof. Dr. Thekra H. Ali
Date: : / /2019

(Member / Supervisor)

Signature:
Name: Assist.Prof.Dr Taha M. Hassan
Date: / /2019

(Member/ Supervisor)

Signature:
Name: Assist.Prof.Dr. Jumana W. Salih
Date: / /2019

***Approved by the Dean of College of Science, University of Diyala*.**

(The Dean)

Signature:
Name: Prof.Dr. Tahseen Hussein Mubarak

Date: / /2019

Abstract

Breast cancer is one of the most types of cancers that infect females all over the world. It is happened when the cells in the breast tissues start to grow in an uncontrollable way. Because it leads to death, early detection and diagnosis is very important task to save patients life. Due to the restriction of human observers, computers play a key role in detecting early signs of cancer. The challenges that face the computer-aided and diagnosis systems are the wide range of features that define abnormalities and the fact that they are often indistinguishable from the surrounding tissues.

There are many contributions that have been introduced in the field of breast cancer detection and diagnosis by using different machine learning techniques. The differences between these researches and this proposed system are the using of adaptive threshold value depended on each image, using Discrete Wavelet Transform in both segmentation and feature extraction phase which decrease complexity and time, detecting more than one tumor in the image and using a hybrid of fuzzy logic and probabilistic neural network gave accurate result and reduced time.

The proposed system uses a multi-resolution analysis and a top-hat operation to detect the suspicious regions in a mammogram image. A discrete wavelet transform feature analysis is used to extract features form region of interest. Fuzzy Logic (FL) and Probabilistic Neural Network (PNN) are used to classify the tumor into normal or abnormal. The dataset that used in the proposed system is Mammographic Image Analysis Society (MIAS). The result showed that the accuracy, sensitivity and specificity of the proposed system were equal to about 99%, 98% and 47% respectively.

Contents

	Contents	Page No.
Chapter One: General Introduction		
1.1	Introduction	1
1.2	Literature Review	3
1.3	Problem Statement	5
1.4	Aim of Thesis	5
1.5	Thesis Outline	6
Chapter Two: Theoretical Background		
2.1	Breast anatomy	7
2.2	Breast cancer	8
2.3	Mammography	12
2.4	Image Preprocessing	13
2.4.1	Histogram Equalization Techniques	14
2.4.2	Wiener Filter	15
2.5	Segmentation	16
2.5.1	Discrete Wavelet Transform	17
2.6	Features Extraction	20
2.7	Classification	20
2.7.1	Probabilistic Neural Network	21
2.7.2	Fuzzy Logic	23
2.5	Performance Measurements	27
Chapter Three: The Proposed System		
3.1	Introduction	30
3.2	Proposed System	30
3.3	Preprocessing	33
3.4	Enhancement	36
3.5	Segmentation Technique	37
3.5.1	The Coarse Segmentation	39
3.5.1.1	Wavelet Transforms	39
3.5.1.2	Wavelet based thresholding	39
3.5.2	Fine Segmentation	40
3.5.2.1	Morphological Enhancement	40
3.6	Features Extraction	41
3.7	Classification	42
Chapter Four: Experimental Results		
4.1	Introduction	45
4.2	Mammographic Dataset and Used Tools	45
4.3	Test Material	46
4.4	Experiments	46

4.4.1	Preprocessing	46
4.4.1.1	Detection Breast region	47
4.4.1.2	Resize image	48
4.4.2	Enhancement	49
4.4.3	Segmentation	52
4.4.3.1	Coarse Segmentation	52
4.4.3.2	Fine Segmentation	55
4.4.4	Features Extraction	56
4.4.5	Classification and Pattern Recognition	56
Chapter Five: Conclusions and Future Works		
5.1	Conclusions	59
5.2	Future Works	60
References		61

List of Figures

Figure No.	Caption	Page No.
1.1	General computer-aided diagnosis	3
2.1	The structure of breast	7
2.2	Cancerous cell vs. normal cell	8
2.3	Chart of commonest ten cancers 2015	11
2.4	Views taken in screening mammography	12
2.5	Latero-Medial (LM)	13
2.6	Medio-Lateral (ML)	13
2.7	Image enhancement	13
2.8	Bank of filters iterated for DTW standard	18
2.9	2D-DWT with 3-level decomposition	19
2.10	Daubechies Families	19
2.11	PNN Structure	22
2.12	Fuzzy set of tall	25
2.13	The process of involved in a FIS	27
3.1	Digital mammogram	31
3.2	The proposed model block diagram	32
3.3	Preprocessing step block diagram	33
3.4	Enhancement step block diagram	36
3.5	Segmentation step block diagram	38
3.6	Block Diagram of Feature Extraction steps	41
3.7	Block Diagram of Classification steps	42
4.1	Sample of mammogram images	46
4.2	Architecture of mammogram image	47
4.3	Figure [4.3]: **(a)** original image, **(b)** Binary Image,**(c)** Extract breast profile, **(d)** Multiplied image **c** with image **a**.	48
4.4	Reducing the size of image	48
4.5	Pectoral muscle suppression	49
4.6	(a) before applying Wiener filter and (b) after applying the filter	50
4.7	Before and after applying CLAHE	50
4.8	Enhancement image1	51
4.9	Enhancement image2	51
4.10	After normalization image	52
4.11	Applying 2D Wavelet Transform	53
4.12	Sample of adaptive threshold of table [4.1].	54
4.13	Coarse segmentation	54

4.14	Fine segmentation	55
4.15	A summary (a) Original image, (b) apply Db6, (c) Coarse segmentation and (d) Fine segmentation	55

List of Tables

Table No.	Caption	Page No.
2.1	Normal cells vs. cancerous cells	9
2.2	The registered case of different type of cancer in Iraq 2015	10
2.3	Standard performance metrics	28
4.1	Sample of adaptive threshold	53
4.2	Confusion matrix	57
4.3	Performance of Classification	58
4.4	Comparison with some related work	58

List of Abbreviations

Abbreviations	Meaning
2D	Two Dimensional
3D	Three Dimensional
4D	Four Dimensional
AHE	Adaptive Histogram Equalization
ANN	Artificial Neural Network
BPANN	Backpropagation Artificial Neural Network
CADx	Computer-aided Diagnosis
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CT	Computed-Tomography
DWT	Discrete Wavelet Transform
FIS	Fuzzy Inference System
FL	Fuzzy Logic
GLCM	Gray Level Co-Occurrence Matrix
GONNA	Genetically Optimized Neural Network Algorithm
HE	Histogram Equalization
LM	Latero-Medial
LTEM	Law's Texture Energy Measure
MCs	Micro-calcification
MIAS	Mammographic Image Analysis Society
ML	Medio-Lateral
MLO	MedioLateral-Oblique
MRI	Magnetic-Resonance Imaging
OTUS	multi-threshold
PDF	Probability Density Function
PET	Positron Emission Tomography
PNN	Probabilistic Neural Network
SPC	Specificity
SVM	Support Vector Machines
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

Chapter One

General Introduction

Chapter One

General Introduction

1.1 Introduction

In last few decades cancer is one of the most critical and deadly diseases all over the world. The cancer starts in cell, which is the building block that form tissues. Tissues can be found in any portions of the human body, involving the breast. Generally, cells are creating and split each time the body needs them, for growing and survive. When the normal cell becomes old, it shrivels until die, then new cell will be created. Occasionally, this process does not follow a normal way, some new cells are created when they are not needed any more, and old cells don't die to let the new cells to replace them. This uncommon creation of the cells forms a chunk of tissue, also called, a tumor, lump, or growth. The cancer which forms in tissues of breast is called, breast cancer [1]. Because the cancer leads to death, most of the countries around the world, especially the industrialized countries, have directed the efforts to the early detection of breast cancer, which will improve the chances of success treatment [2].

In 2010 the reports showed that nearly 1.5 million cases of breast cancer happened to females. Additionally, reports show that among 23% of breast cancer cases, about 14% of these cases ended with death. In 2018 approximately 2.9 million new cases were diagnosed and about 627,000 of these cases were end with death. The detection of tumor in its early stages, diagnosis, and proper treatment helps in decreasing the fatality exhibits rate [3].

In the early stages tumors are noticed as teeny bright spots via mammography screening, these spots are calcium deposits called micro-calcifications (MCs). Usually they are not obvious in the image and it is difficult to recognize by the radiologists due to human vision system. The

second type of the breast cancer is called masses, they can be easy to discover because of their shape, color contrast, and size, but some types of masses might appear similar to the normal breast tissue which make them difficult to discover [4].

Due to the fast growth of technology in a field of computer science, the idea of an electronic doctor doesn't seem to be totally unimaginable. It isn’t a new technology, Ledley and Lusted depicted the theory as symbolical logic, and in 1959, the probability of helping in diagnosis was invented by physicians using complex reasoning.

Since that, CADx has been faced a lot of challenges in both medical domain and computer sciences that have been impeded its application in clinical practice. Since the 1980s, there were great progresses in technology, involving medical equipment, and computer science have played vital roles in improving the quality of these medical images, which made CADx systems more popular in the medical domain. Essentially, in the 1990s, a number of CADx programs were applied in commerce, from that we can say CADx had been started a new domain [5].

CADx is a tool used for helping radiologists for increasing the accuracy of images and for specifying possible results to avoid the wrong explanation. The CADx is consist of software and hardware for images that are obtained by Magnetic Resonance Imaging (MRI), Computed Tomography (CT), positron emission derived from (CT) and (MRI), mammography, Ultrasound, X-ray, positron emission tomography (PET), Photography, and etc., which can be in the form of 2D, 3D, and even 4D images in sometimes, they can be analyzed by using the CADx systems. In a digital format these images can't give us all data that we need. Some of the radiological equipment like tomography and MRI output digital images directly. In some cases, a film must be converted into a digital image such as with X-ray and also as in a mammogram by using a mammogram

conversion. Figure (1.1) shows the general computer aided diagnosis system [6].

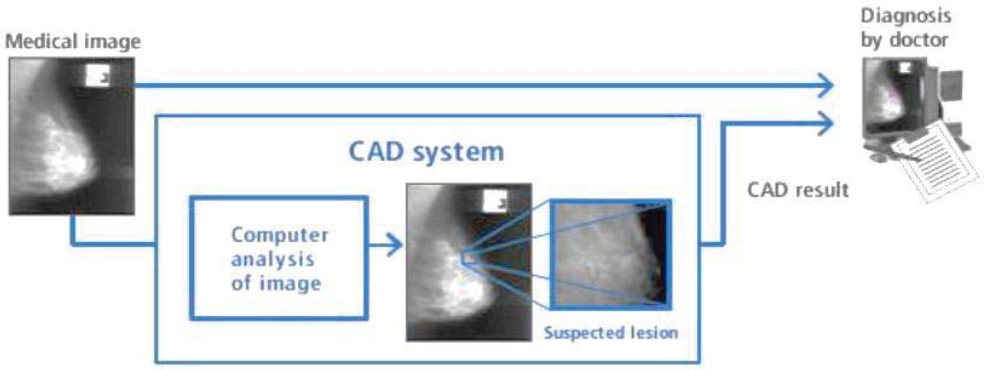

Figure (1.1): General computer-aided diagnosis [6].

1.2 Literature Review

There are many contributions that have been introduced in the field of breast cancer diagnosis, some of them are described briefly as follow:

- In (2011), J. Dheeba et al. proposed a supervised machine learning algorithm, Artificial Neural Network that tuned by a Modified Genetic Algorithm for detection of breast cancer. A feature extraction methodology is used for extracting the texture features of the normal tissue and cancerous tissue for increasing the accuracy of detection. The result showed that the proposed algorithm had a recognition score of 97.8% [7].
- In (2012), Swati et al. discussed the detection of breast cancer using ART1 Network method. The process in this method is occurred in three steps: recognition, comparison and search. A Backpropagation algorithm is used for training of network design and error minimization. The results showed that the ART1 Network is very efficient. For unsupervised learning pattern, the accuracy detection was 92% [8].

- In (2013), Punam S. Pawar et al. developed a technique for detecting of breast cancer also by using Backpropagation Neural Network. By increasing the number of neurons in hidden layer the accuracy gets improved. They get best results with 9 neurons in the hidden layer. The accuracy of the propose technique was 99% [9].
- Arpit Bhardwaj et al. (2014), proposed a new Genetically Optimized Neural Network Algorithm (GONNA). A new crossover and mutation processes have been introduced to reduce the devastating nature of these processes. The GONNA have been used to classify breast cancer tumours as benign or malignant. The proposed algorithm gave better classification accuracy of 98.52% [10].
- In (2015), S. Saini et al. presented a breast cancer detection technique by using image processing models by Artificial Neural Networks (ANN). Gray Level Co-Occurrence Matrix (GLCM) feature extracted is used to train Artificial Neural Network. From Artificial Neural Network they obtained an 87.5% accuracy and 67.8% accuracy from Cascade-forward Backpropagation Network [11].
- In (2016) S. Naranje employed an automatic technique using Artificial Neural Network (ANN). After a pre-processing operation on the breast cancer dataset the data obtained used as an input to the ANN to classify the tumor as a cancerous or non-cancerous tumor. The proposed technique achieved more than 90%accuracy [12].
- In (2017), Y. J. Tan et al. developed a method by using a Convolutional Neural Network (CNN) to speed up the diagnosis process to help the specialists to detect the abnormalities. A CNN is trained using an enhanced mammogram images, the classifier then produced a model to detect the cancerous tumor. The proposed method provided an 82.73 % accuracy and a fast diagnosis time [13].

- In (2018), I. Routray et al. proposed an approach to detect breast cancer by using Law's Texture Energy Measure (LTEM) method. Backpropagation Artificial Neural Network (BPANN) method is used for classifying the malignant, benign and normal tissue region. The performance of suggested method achieved a sensitivity of 90.9% with an accuracy of 94.4% for normal-abnormal classification. It is also achieved accuracy of 91.7% and 66.66% specificity for benign-malignant classification [14].

1.3 Problem Statement

A breast cancer is a fatal disease that affects a large proportion of females around the world. Early detection of this disease helps in healing and decreasing the chance of death.

X-ray mammogram is classified as the easiest and widespread technique for cancer detection in its early stages. Radiologists might miss 10% to 15% of breast cancer tumor, so CADx systems are used to help them taking the accurate decision [15]. By using CADx reduced both cost and time. The computer system used several methods and techniques including, databases, image processing, machine learning and data analysis tools to detect and diagnose breast cancer.

1.5 Aim of Thesis

The aim of this thesis is to introduce a hybrid approach for breast cancer diagnosis based on an intelligent computing model to help the specialists to specify suspicions regions in the breast mammogram images. Using image processing methods with machine learning techniques to detect the masses in digital mammogram images and develop a model for diagnosing

these masses as normal and abnormal for reducing the number of falsely classified cancers and increasing the accuracy of cancer detection and diagnosis.

1.6 Thesis Outline

In addition to this chapter, the rest of the thesis including the following chapters:

Chapter Two: Theoretical Background

This chapter is divided to the following sections:

1. Medical background. This section includes breast anatomy, and breast cancer.
2. Review of Image Processing Method and machine learning. This section involves image enhancement, segmentation, representation and extraction feature, and classification.

Chapter Three: Proposed System and Applied Methods

It contains the proposed methods and algorithms which are implemented in this work.

Chapter Four: Experiments and Results

The experiment's stages are clearly explained with snapshots of the proposed system along with the final results after evaluating the work that is represented in this chapter.

Chapter Five: Conclusions and Future Work

And finally, the last chapter contained the research conclusions and a number of recommendations for future work.