زانکۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

A REVIEW ON BREAST CANCER DETECTION BY USING IMAGE PROCESSING TECHNIQUE

Research Project

This research project has been written under my supervision and has been submitted for the award of the degree of B.Sc. in **(PHYSICS)**.

BY:
Niyan Rzgar Baiz

SUPERVISED BY:

Assistant Prof. Dr. Sardar P. Yaba

April

2024

بِسْمِ اللَّهِ الرَّحْمنِ الرَّحِيمِ

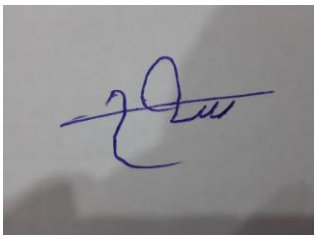
قَالُواْ سُبْحَانَكَ لاَ عِلْمَ لَنَا إِلاَّ مَا عَلَّمْتَنَا إِنَّكَ أَنتَ الْعَلِيمُ الْحَكِيمُ

صدق الله العظيم

سورة البقرة الاية32

This research project has been written under my supervision and has been submitted for the award of the degree of B.Sc. in (Physics)

Supervisor:

Signature: 

Name: **Assistant Prof. Dr. Sardar P. Yaba**

Date: / /2024

I confirm that all requirements have been completed.

Signature:

Name:

Head of the Department of physics

Date: / /2024

This thesis is

dedicated:

- ✓ *To My Parents*
- ✓ *To My Brother and Sisters*

ACKNOWLEDGEMENTS

First and foremost, I must acknowledge my limitless thanks to **Allah**, the Ever-Magnificent; the Ever-Thankful, for His help and bless. I owe a deep debt of gratitude to the physics department and Salahaddin university for giving us an opportunity to complete this work. I am very appreciative to my colleagues in the physics department. I am grateful to some people, who worked hard with me from the beginning till the completion of the present research particularly my supervisor **Assistant Prof. Dr. Sardar P. Yaba** who has been always generous during all phases of the research. I would like to take this opportunity to say warm thanks to all my beloved friends, who have been so supportive along the way of doing my thesis. Last but not least, deepest thanks go to all people who took part in making this thesis real.

SUMMARY

Breast cancer is considered to be one of the most common types of cancer affecting women worldwide. Early detection of breast cancer is a critical requirement for reducing the mortality rate among women. At present, digital X-ray mammography is the most reliable screening device for suspicious breast masses and microcalcifications in the early stages. Classification of mass plays a role of vital importance in the diagnosis of breast cancer. The existing computer-aided diagnosis (CAD) methods used to benefit a lot from low-level or middle-level features which are not that good at the simulation of real diagnostic processes. This project aims to demonstrate the best methods used for categorizing and identifying breast cancer in its early phases by deep learning methods. In this project, a comparative analysis of different techniques based on machine learning for the detection of breast cancer has been presented. We have reviewed journal papers from 2015 to 2019 about this topic for the detection of breast cancer. The results show that Beura et al. achieved the highest performance ratio of 98.8%. Moreover, Al-antari et al., achieved the lowest performance ratio was %93.5.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION AND THEORETICAL BACKGROUND

1.1 Introduction

Breast cancer is one of the most common cancers worldwide among women. The causal agent of breast cancer is still under research. But there are some jeopardy factors such as age, genes, obesity, taking birth control pills and smoking. Normally, breast cancer is a malignant tumor that initiates in the cells of the breast and eventually extends to the surrounding tissues. The disease can be preserved if it is detected early. Factors that appear to be implicated in decreasing the risk of breast cancer are regular breast examinations by health care professionals, regular mammograms, self-examination of breasts, healthy diet, and exercise to decrease excess body fat.

Mammography is the most effective imaging modality for early detection of breast cancer in the early stages. Early detection of breast cancer is a significant step toward efficient treatment of the disease. Different computer-aided detection algorithms have been developed to help radiologists provide an accurate diagnosis (da Cruz, 2011).

This project analyses various breast cancer detection techniques based on image processing techniques, data mining methods, various features used and a brief comparative study of the existing breast cancer detection system.

1.2 Theoretical background

1.2.1 The breast

The breasts are paired structures located on the anterior thoracic wall, in the pectoral region. They are present in both males and females, yet are more prominent in females following puberty. In females, the breasts contain different types of tissue:

glandular tissue, which includes the breast lobes and breast ducts fibrous, or supportive or connective tissue, which is the same tissue that ligaments and scar tissue are made of fatty tissue fills in the spaces between glandular and fibrous tissue and largely determines your breast size. All non-fatty tissue is fibro glandular tissue. There are also bands of supportive, flexible connective tissue called ligaments, which stretch from the skin to the chest wall to hold the breast tissue in place. Muscle plays an important role too. The pectoral muscle lies against the chest wall underneath both breasts, giving them support. Blood vessels provide oxygen to the breast tissue and carry away waste, as shown in Figure (1-1).

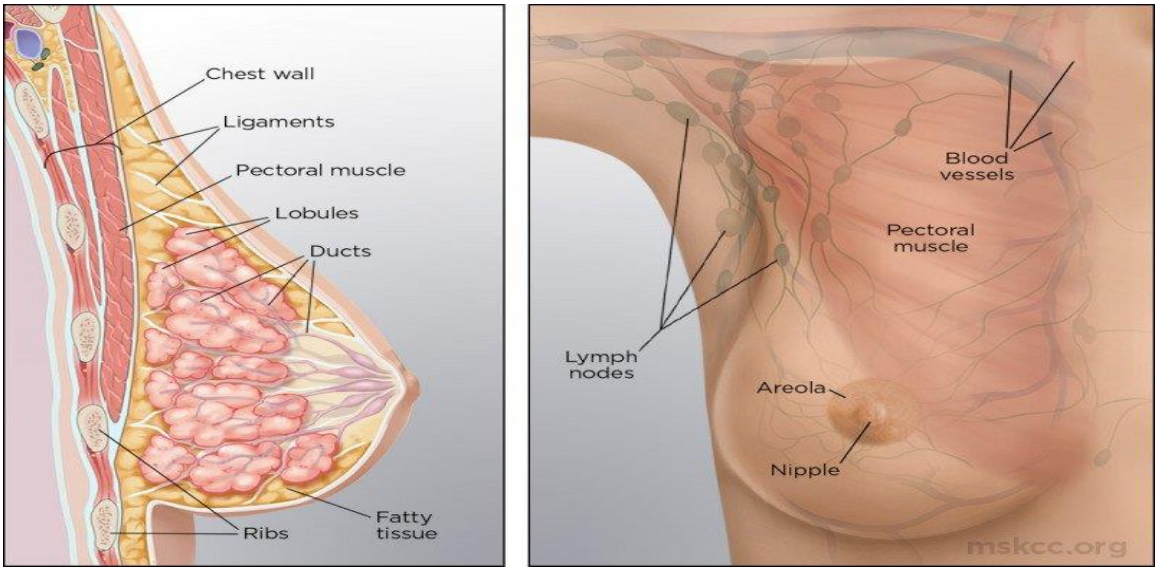


Figure (1-1) shows breast anatomy.(BONADONNA et al, 2006)

1.2.2 The breast cancer

A breast is made up of three main parts: lobules, ducts, and connective tissue. The lobules are the glands that produce milk. The ducts are tubes that carry milk to the nipple. The connective tissue (which consists of fibrous and fatty tissue) surrounds and holds everything together. Breast cancer is a disease in which cells in the breast grow out of control. There are different kinds of breast cancer. The kind of breast cancer depends on which cells in the breast turn into cancer. Most breast cancers begin in the ducts or lobules. Breast cancer can spread outside the breast through blood vessels and lymph vessels. When breast cancer spreads to other parts of the body, it is said to have metastasized.

1.2.2.1 Kinds of breast cancer

The most common kinds of breast cancer are:

1-Invasive ductal carcinoma. The cancer cells begin in the ducts and then grow outside the ducts into other parts of the breast tissue.

2-Invasive lobular carcinoma. Cancer cells begin in the lobules and then spread from the lobules to the breast tissues that are close by.

About 80% of breast cancer cases are invasive, meaning a tumor may spread from your breast to other areas of your body.

3-There are several other less common kinds of breast cancer, such as Paget's disease, medullary, mucinous, and inflammatory breast cancer.

1.2.2.2 Symptoms of breast cancer

Some breast cancer symptoms are very distinctive symptoms may include:

A change in the size, shape or contour of your breast.

A mass or lump may feel as small as a pea.

A lump or thickening in or near your breast or in your underarm that persists through your menstrual cycle.

A change in the look or feel of your skin on your breast or nipple. Your skin may look dimpled, puckered, scaly or inflamed. It may look red, purple or darker than other parts of your breast.

A marble-like hardened area under your skin.

A blood-stained or clear fluid discharge from your nipple (Board, 2022).

1.2.2.3 Causes of breast cancer

Experts know breast cancer happens when breast cells mutate and become cancerous cells that divide and multiply to create tumors. Research shows several risk factors may increase your chances of developing breast cancer. These include:

Age: Being 55 or older.

Sex: Women are much more likely to develop the condition than men.

Family history: If your parents, siblings, children or other close relatives have breast cancer, you're at risk of developing the disease.

Genetics: Up to 15% of people with breast cancer develop the disease because they have inherited genetic mutations.

Smoking: Tobacco use has been linked to many different types of cancer, including breast cancer.

Drinking beverages containing alcohol.

Having obesity.

Radiation exposure: If you've had prior radiation therapy — especially to your head, neck or chest — you're more likely to develop breast cancer.

1.2.2.4 Diagnosed of breast cancer

The following tests to diagnose the breast cancer:

Mammograms

Breast ultrasound.

Breast magnetic resonance imaging (MRI) scan.

Breast biopsy.

Immunohistochemistry test to check for hormone receptors.

Genetic tests to identify mutations that cause breast cancer (Bonadonna et al., 2006, Agur Anne and AF, 2009).

1.2.3 Mammography

Mammography is an x-ray imaging method used to examine the breast for the early detection of cancer and other breast diseases. It is used as both a diagnostic and screening tool. A mammogram is an X-ray picture of the breast. Doctors use a mammogram to look for early signs of breast cancer. Regular mammograms can detect breast cancer early. While **screening mammograms** are routinely performed to detect breast cancer in women who have no apparent symptoms, **diagnostic mammograms** are used after suspicious results on a screening mammogram or after some signs of breast cancer alert the physician to check the tissue. A diagnostic mammogram can help determine if these symptoms are indicative of the presence of cancer (James, 1978). The earlier the cancer is detected, the greater the likelihood of a successful outcome. For early detection of masses and abnormalities in breast cancer, the mammogram is the most important technique that can detect 85 to 90 per cent of all breast cancers. Computer-aided detection (CAD) techniques by processing and analyzing mammogram images can help radiologists with

mass detection and classification. The detection sensitivity without CAD is 80% and with CAD up to 90%1. Limitations in the human eye-brain visual system, reader fatigue, distraction, and the vast number of normal cases seen in screening programs cause radiologists to not always correctly characterize abnormalities in the medical images. A CAD tool can help to improve the diagnostic accuracy of radiologists and lighten the burden of increasing workload. Most image processing algorithms include steps such as Preprocessing, Segmentation, Feature extraction, Feature selection and Classification. Preprocessing is the first step in image processing. To reduce the noise and improve the quality of the image, it has to be done on digitized images. To find suspicious regions of interest (ROIs) containing abnormalities, the segmentation step is applied. The most studied version of grouping in computer vision is image segmentation. Image segmentation is one of the most important tasks in automatic image processing. In general, image segmentation is the process of dividing an image into homogenous groups such that each region is homogenous but the union of no two adjacent regions is homogenous. Image segmentation techniques can be classified into two broad categories (a) region-based, and (b) contour-based approaches. Image segmentation plays a crucial role in many medical imaging applications by automating or facilitating the delineation of anatomical structures and other regions of interest. In the feature extraction step, the features are calculated from the characteristics of the region of interest. In feature selection, the best set of features is selected for eliminating false positives and for classifying lesion types. Feature selection is selecting a smaller feature subset that leads to the largest value of some classifier performance function. Finally, based on selected features, the

false positive reduction and lesion classification is performed in the classification step (Gonzalez, 2009).

1.3 Literature Review

Cancer detection continues to be a challenge for medical professionals. Cancer that is caught early enough can be treated. Various image processing methods are used by different techniques for noise reduction, segmentation, feature extraction, and classification of cancer in mammograms.

(Beura et al., 2015) proposed a mammogram classification scheme to classify the breast tissues as normal, benign or malignant. The feature matrix is generated using GLCM to all the detailed coefficients from 2D-DWT of the region of interest (ROI) of a mammogram. The relevant features are used in a BPNN classifier for classification. MIAS, and DDSM databases are used for the proposed scheme. The accuracy measures are computed for normal vs. abnormal and benign vs. malignant. An accuracy of **98.0% and 94.2%** has been obtained for normal–abnormal and benign–malignant respectively in the MIAS database. Similar parameters are **98.8% and 97.4%** are achieved in the DDSM database.

(Jiao et al., 2016) designed a deep feature-based framework for breast mass classification tasks. It mainly contains a convolutional neural network (CNN) and a decision mechanism. The classifiers based on different features were analyzed jointly to determine the types of test images. Deep features extracted from different layers illustrate effectiveness in classification performance and diagnosis simulation. In addition, this method was applied to DDSM datase

and achieved high accuracy under two objective evaluation measures. The classification accuracy is **96.7%.**

(Platania et al., 2017) proposed framework for automated breast cancer detection and diagnosis, called BC-DROID, which provides an automated region of interest detection and diagnosis using convolutional neural networks, and multiple GPUs and computation nodes features were used. The resulting network can detect and classify regions of interest as cancerous or benign in one step. Demonstrate the accuracy of the framework's ability to both locate the regions of interest as well as diagnose them. DDSM dataset are used for the proposed scheme, and achieves a classification accuracy of **93.5%**.

(Al-Antari et al., 2018) proposed CAD system involving detection, segmentation, and classification of breast masses via deep learning methodologies, to detect breast mass from entire mammograms, You-Only-Look-Once (YOLO), a regional deep learning approach is used. To segment the mass, a full resolution convolutional network (FrCN), a new deep network model, is proposed and utilized. Finally, a deep convolutional neural network (CNN) is used to recognize the mass and classify it as benign or malignant, by using deep features and the high deep level features. To evaluate the proposed integrated CAD system in terms of the accuracies of detection, segmentation, and classification, the publicly available and annotated INbreast database was utilized. The evaluation results of the proposed CAD system via CNN classification achieved an overall accuracy of Benign - Malignant **95.64% - 94.78 %**.

(Vijayarajeswari et al., 2019**)** presented the classification of mammograms using features extracted using the Hough transform. Hough transform is a two dimensional transform. It is used to isolate features of a particular shape in an image. Miniaturized scale characterization and masses are the two most vital markers of threat, and their mechanized identification is exceptionally important for early breast cancer diagnosis. Since masses are regularly undefined from the encompassing parenchymal, computerized mass location and arrangement is significantly additionally difficult. Here, Hough transform is used to detect features of mammogram images and it is classified using SVM. DDSM dataset are used for the proposed scheme, and achieves a classification accuracy is **94%** by the use of SVM classifier.

CHAPTER TWO

MATERIALS AND METHODS/THEORY

2.1 Materials

Five studies were evaluated in this review article for their accuracy in classifying normal and abnormal breast cell, from 2015 to 2019 (Beura et al., 2015, Jiao et al., 2016, Platania et al., 2017, Al-Antari et al., 2018, Vijayarajeswari et al., 2019)**.** The objective of this research work is to process and analyze the images that are framed from the mammography and generate results on whether the images contain cancer cells or not.

2.2 Methods

The sequence of operations for detection of tumor in the mammography of the breast, by using various steps preprocessing, segmentation, feature extraction, and classification of the breast mass, which are shown in the Figure (2.1).

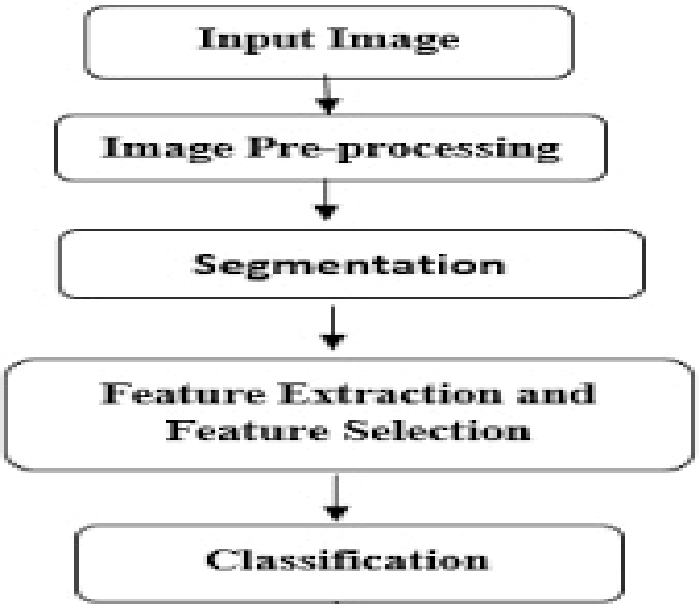


Figure (2-1). Basic Block Diagram

Image Enhancement Stage

The methods used for image enhancements are categorized into traditional methods and methods based on Deep Learning.

Image Segmentation Stage

The detection of tumors in the breast depends on segmentation techniques. Segmentation plays a significant role in image analysis and includes detection, feature extraction, classification, and treatment. Segmentation helps physicians quantify the volume of tissue in the breast for treatment planning.

classification Stage

The purpose of classification is to select the best treatment. The effectiveness of a specific treatment is demonstrated for a specific breast cancer (usually by randomized, controlled trials). That treatment may not be effective in a

different breast cancer. Some breast cancers are aggressive and life-threatening, and must be treated with aggressive treatments that have major adverse effects. Other breast cancers are less aggressive and can be treated with less aggressive treatments, such as lumpectomy Grade. Grading focuses on the appearance of the breast cancer cells compared 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 have a worse prognosis. (Gonzalez, 2009).

CHAPTER THREE

RESULTS AND DISCUSSION

3.1 Result

According to the survey in this project, one observes various image processing methods were used for breast cancer detection, as shown in Table (3-1).

Table (3-1) Classification accuracy of different methods in literature.

Author	Database	Features	Classification	Accuracy (%)
(Beura et al., 2015)	DDSM MIAS	Combination of DWT, and GLCM.	BPNN	**98 - 94** **98.8 – 97.4**
(Jiao et al., 2016)	DDSM	Deep features from different layers	CNN	96.7
(Platania et al., 2017)	DDSM	multiple GPUs and computation nodes	CNN	**93.5**
(Al-Antari et al., 2018)	INbreast	deep features and the high deep level features	CNN	95.64 – 94.78
(Vijayarajeswari et al., 2019)	DDSM	Hough Transform	SVM	94

3.2 Discussion

Despite extensive study, there is still a lack of an accurate model since detection is a multidisciplinary role that depends on a variety of parameters. The advancement of accurate cancer detection techniques at an earlier stage has been the subject of extensive research. Among various methods discussed in this project, (Beura et al., 2015) achieved the highest performance ratio presented in Table (3-1) due to using of 2D-DWT and GLCM in succession to derive a feature matrix from mammograms, and using BPNN classification. It is observed to achieve higher classification accuracy with BPNN. An accuracy of 98.0% and 94.2% has been obtained for normal–abnormal and benign–malignant respectively in the MIAS database. Similar parameters are 98.8% and 97.4% are achieved in the DDSM database. Moreover, (Al-Antari et al., 2018), achieved the lowest performance ratio of %93.5 presented in Table (3-1), due to using multiple GPUs and computation nodes, and using CNN classification for benign–malignant in the DDSM database.

CHAPTER FOUR

CONCLUSION AND FUTURE WORK

4.1 Conclusion

1- In this project, we obtained an efficient mammogram classification scheme to support the decision of radiologists.

2- The scheme utilizes 2D-DWT and GLCM in succession to derive a feature matrix from mammograms.

3- It is observed to achieve higher classification accuracy with BPNN. An accuracy of 98.0% and 94.2% has been obtained for normal–abnormal and benign–malignant respectively in the MIAS database. The similar parameters are 98.8% and 97.4% are achieved in the DDSM database.

4.2. Future Work

In the future, we prefer to identify the influential factors for breast cancer in the human body and use the same features of cancer prediction. An IoMT application can be developed to extract the physiological parameters, physiological parameters for prediction include blood pressure, body mass index, and cholesterol.

REFERENCES

AGUR ANNE, M. & AF, D. 2009. Grant's atlas of anatomy. *Lippincott Williams & Wilkins,* 135.

AL-ANTARI, M. A., AL-MASNI, M. A., CHOI, M.-T., HAN, S.-M. & KIM, T.-S. 2018. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International journal of medical informatics,* 117**,** 44-54.

BEURA, S., MAJHI, B. & DASH, R. 2015. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing,* 154**,** 1-14.

BOARD, C. N. E. 2022. *Breast Cancer: Introduction* [Online]. Available: https://www.cancer.net/cancer-types/breast-cancer/introduction [Accessed 17/3/2024 2024].

BONADONNA, G., GIANNI, L., HORTOBAGYI, G. N. & VALAGUSSA, P. 2006. *Textbook of breast cancer: a clinical guide to therapy*, CRC Press.

DA CRUZ, C. F. 2011. Automatic analysis of mammography images. *Engineering Faculty-Porto University, Oporto, Master in Bioengineering-Biomedical Engineering*.

GONZALEZ, R. C. 2009. *Digital image processing*, Pearson education india.

JAMES, G. T. 1978. Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers. *Analytical biochemistry,* 86**,** 574-579.

JIAO, Z., GAO, X., WANG, Y. & LI, J. 2016. A deep feature based framework for breast masses classification. *Neurocomputing,* 197**,** 221-231.

PLATANIA, R., SHAMS, S., YANG, S., ZHANG, J., LEE, K. & PARK, S.-J. Automated breast cancer diagnosis using deep learning and region of interest detection (bc-droid). Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics, 2017. 536-543.

VIJAYARAJESWARI, R., PARTHASARATHY, P., VIVEKANANDAN, S. & BASHA, A. A. 2019. Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement,* 146**,** 800-805.