Detection of Digital Mammograms Cancer using Feed-forward Neural Networks

Saman Hussein Mahmood[1] **Rizgar Maghded Ahmed**[2] **Kamaran Hassan Ahmed**[3]

saman.mahmood@su.edu.krd rizgar.ahmed@su.edu.krd kamarn.ahmed@su.edu.kr

[1,2,3] **Statistics Department/Administration and Economics College/ Salahaddin University-Erbil**

الكشف عن صور سرطان الثدي الرقمية باستخدام الشبكات العصبية ذات التغذية الامامية

م.د. سامان حسين محمود[1] **م.د. رزكار مغديد احمد**[2] **م.د. كامران حسن احمد**[3]

saman.mahmood@su.edu.krd rizgar.ahmed@su.edu.krd kamarn.ahmed@su.edu.kr

[1,2,3] **قسم الإحصاء/كلية الإدارة والإقتصاد/جامعة صلاح الدين-أربيل**

تاريخ استلام البحث	2023/2/24	تاريخ قبول النشر	2023/4/8	تاريخ النشر	2024/ 12/ 30

Abstract

The aim of this paper is bring together two areas, which are Artificial Neural Networks (feed-forward neural networks) and digital image processing techniques to early detect of malignant and benign shape of masses in digital breast cancer images using a set of statistical, engineering features and measurements extracted from the image for the purpose of analysis and detection. The neural networks will be trained and tested using data is divided so that 70% is used for training and 30% for testing. The proposed model of neural network after the training and testing is (13-5-2) (The input layer has 13 inputs; the hidden layers have 5 neurons and the output layer).

The results of the performed classification showed that the predictive ability of neural networks achieved (95% of accuracy, 96% of Sensitivity and 94% of Specificity) of classification rate of mammograms cancer into malignant and benign, and also extract four important variables to the diagnosis which are mean, median, Standard deviation and Solidity.

Keywords: *Breast cancer; Image processing, Feed-forward neural networks, Feature extraction*

المستخلص

الهدف من هذا البحث هو الجمع بين مجالين وهما الشبكات العصبية الاصطناعية (الشبكات العصبية ذات التغذية الامامية) وتقنيات معالجة الصور الرقمية للكشف المبكر لشكل الكتل الخبيثة و الحميدة للصور الرقمية لسرطان الثدي باستخدام مجموعة من الخصائص والمقاييس الإحصائية والهندسية المستخرجة من الصور لغرض التحليل والكشف. ومن أجل تدريب واختبار الشبكات العصبية تم تقسيم البيانات إلى 70٪ من العينات المستخدمة للتدريب و 30٪ لاختبار النموذج. النموذج المقترح للشبكة العصبية بعد التدريب والاختبار هي (13-5-2) (طبقة المدخلات لديها 13 مدخل، وطبقه المخفية لديها 5 عقد وطبقة المخرجات).

اظهرت نتائج التصنيف أن أداء القدرة التنبؤية للشبكات العصبية حققت (95٪ من الدقة، 96٪ من الحساسية و 94% من الخصوصية) من معدل تصنيف صور سرطان الثدي إلى خبيثة وحميدة، وأيضا تم استخراج أربعة متغيرات مهمة فى التشخيص وهي الوسط الحسابي، الوسيط، الانحراف المعياري والصلابة.

الكلمات المفتاحية: *سرطان الثدي، معالجة الصور، الشبكات العصبية ذات التغذية الأمامية ، استخراج الميزات*

1. Introduction

Moreover, 30% of cancer-related deaths in women occur from breast cancer, which is a malignant tumor that begins in the cells of the breast. This accounts for around 1% of all deaths globally. Since breast cancer accounts for 1.5% of all male cancer deaths, it is a cause for concern for men as well. Analyzing breast images can help in catching breast cancer early (Kadhim, 2012:116).

As a replacement for the issues caused by traditional screening programs, digital mammography has become one of the most popular ways to find breast cancer. To overcome these obstacles, an automated system can reduce the number of incorrect positive and negative results from radiologists, increasing the chances of detecting abnormalities at an early stage. Because aberrant masses tend to blend into homogeneous breast tissue, computerized mammography has had a hard time correctly identifying masses until now (Kadhim, 2012:116).

Decisions in modern medicine rely heavily on detection and classification. It is becoming increasingly important to use automatic algorithms due to the ever-increasing amount and complexity of patient data that must be examined when important judgments are made. They may just back a doctor's decision; however, they can't take their place (Berka, et.al, 2009:104).

The goal of this study is to find a way to identify and categorize breast cancer using image processing and artificial neural networks. An integral part of digital mammography is image processing, which comprises a number of methods for creating digital mammograms, such as identifying and classifying masses (both benign and malignant) using multilayer feed-forward neural networks (FFNNs), and extracting features (metrics and statistics) for each image independently (Naranje, 2016:677).

There are a variety of categorization problems that multilayer FFNNs can solve. A neural network's weights dictate how the network operates. Thus, the most important issue is how to find the weights for a specific problem so that the neural network can behave as intended. Parameter learning from examples is at the heart of the neural network technique. Neural networks may take in information from their surroundings and use it to get smarter. Training or learning a neural network entails an adaptive technique wherein the network's weights are incrementally changed to enhance a pre-specified performance metric. The process of adaptation can be either supervised or unsupervised, and it is referred to as a learning rule (learning algorithm). The training examples used in supervised learning are sets of known input-output pattern pairs (Berka, et.al, 2009:83).

2. Related Work

Many research studies have been conducted earlier in regards to breast cancer detection and classification; the following introduces some of them.

Medical researchers Ganesan et al. (2010) conducted pre-clinical studies on carcinogenesis using neural networks. With the goal of creating diagnostic algorithms that can enhance emergency department triage techniques, this study examined demographic data from patients with lung cancer. Both the training and test datasets show that the network's simple rules achieve excellent accuracy in the lung cancer diagnostic problem.

For the purpose of computer-assisted breast cancer diagnosis utilizing textural features, Singh et.al. (2011) introduced an ANN-based classification system. Classification accuracy for benign and malignant breast cancers is 94% according to the artificial neural network (ANN) classifier. Our findings demonstrate that texture features are a powerful tool for accurate mammography picture classification.

Zhang et.al. (2012) created a neural network classifier model that utilizes a floating centroid method combined with a particle algorithm swarm optimization incorporating inertia weight to improve performance. This model outperforms other classification techniques in distinguishing between benign and malignant tumors.

Bălănică, et.al. (2013)'s this paper proposes four novel methods to extract speculation features from breast lesions in mammograms. These features are then used to train an ANN to accurately classify benign and malignant tumors.

Utomo et al. (2014) a paper was published that looked into how ANN with extreme learning techniques for diagnosing breast cancer. The study utilized the Wisconsin Breast Cancer Dataset to train and evaluate the model. The results demonstrated that extreme learning outperformed backpropagation in neural networks in terms of generalization.

An automatic support system was proposed by (Pavitha, R. and Hephzibah, S.T.J. 2014) for stage classification using Probabilistic Neural Networks. Radial basis function network will be used for classification automation in breast cancer. This method involves detecting and classifying cancerous regions using thresholding techniques. It provides faster and more accurate classification and this makes it a potentially useful tool for cancer diagnosis.

Lowis, et.al, (2015) used Wavelet Transform with Dual Tree Complexity image for mammography image analysis based on neural network classifier. The results show that the proposed method performs better than DWT, achieving an accuracy of 96.3%.

In the context of breast cancer, Naranje (2016) proposed the use of ANN as decision-making tools. Classification of mammograms relies heavily on statistical image analysis. With this technology, you may get the most accurate diagnosis possible quickly, regardless of whether the image is malignant or not.

3. Methodology

The following diagram (Fig1) illustrates the main steps involved in image classification.

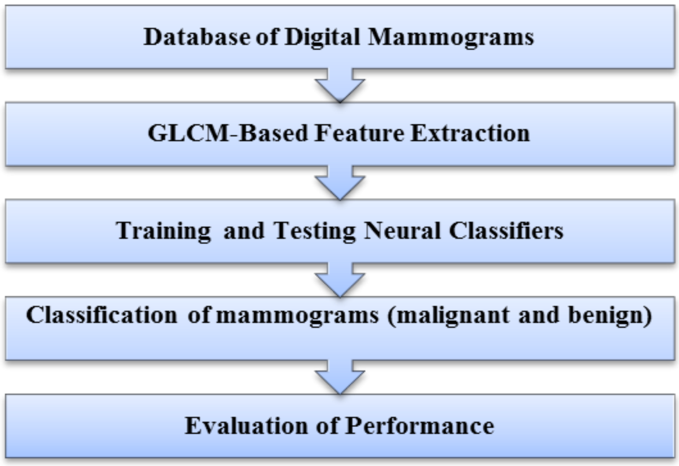


Fig1: A Summary of the Proposed Method

3.1 Digital Image Fundamentals:

We will represent images using two-dimensional functions, with the value at each spatial coordinate representing the intensity or amplitude of the signal at that position. This value is a positive scalar with physical significance determined by the image source. For example, in images created by physical processes, the intensity values correspond to the energy emitted by a physical source, such as electromagnetic waves. As a result, f(x,y) must be both nonzero and finite. (Gonzales & Woods, 2008:73).

$$0 < f(x,y) < \infty$$

These components, known as illumination and reflectance, are indicated as i(x,y) and r(x,y), respectively. They multiply to produce the image function f(x,y):

$$f(x,y) = i(x,y)\, r(x,y)$$

Where

$$0 < i(x,y) < \infty \quad \text{And} \quad 0 < r(x,y) < 1$$

The grey level value at (x,y) differs from that at (x+1,y+1), and we represent it with the (L) symbol.

$$L_{min} \leq L \leq L_{max}$$

$$r_{min} \times i_{min} \leq L \leq r_{max} \times i_{max}$$

$$Level\ 0 \leq L \leq 255$$

Sampling and quantization produce a matrix of real numbers. We shall use two essential representations for digital images. A digital image is created by sampling a continuous image into a M x N grid of pixels. In this discrete representation, Each pixel is identified by integer coordinates (X=0,1,2,..,M-1) and (Y=0,1,2,..,N-1). Accordingly, the coordinate values at the origin are (x,y)=(0,0). The coordinate values for the picture's first row are (x, y)=(0,1). The row and column coordinates, as well as the intensity value (a digital number), are unique identifiers for each pixel. (Gonzales and Woods 2008:79).

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1)\ldots & f(0,N-1) \\ f(1,0) & f(1,1)\ldots & f(1,N-1) \\ \vdots & \vdots & \vdots \\ f(M,0) & f(M,1)\ldots & f(M-1,N-1) \end{bmatrix}$$

This matrix's individual elements are referred to as image elements, picture elements, or pixels.

$$M,N > 0\,, G\ (number\ of\ gray\ levels) \geq 0$$

3.2 X-Ray Mammography Image

X-ray mammography image is a frequently used screening and diagnostic technique in clinical practice. Mammography is the gold standard for detecting early-stage breast cancer. It detects fatty breast tissue very effectively and clearly reveals microcalcifications, which are an indication of early cancer. Because of its low cost, it is ideal for a program that screens a wide population. Mammography is not ideal. The approach is less reliable on thick breasts in younger women or those who have had breast surgery since glandular and scar tissues are as radiopaque as abnormalities. There is some X-ray radiation at low doses (Basha and Prasad, 2009:708).

3.3 Feature Extraction

Feature extraction is the process of translating an image into a set of properties, numerous methods, including statistical, structural, model, and transform-based approaches, have been put forth in the literature for the purpose of feature extraction. This research will focus on a statistical method for detecting and classifying (malignant and benign) form masses in breast mammography images using ROI (Reign of Interest), specifically on first- and second-order statistics (Pratt, 2007:550), See Fig(2).

3.3.1 Statistical Features Analysis

There are a number of ways to quantify the behaviors of image textures; one of the most popular and straightforward is to employ a statistical feature. One or more of two approaches are typically employed by image analysis systems to extract texture features from images: first-order histogram features and second-order Co-occurrence Matrix Features. Additionally, Higher-order invariant moment characteristics may also be employed (Zhang,1996:1336).

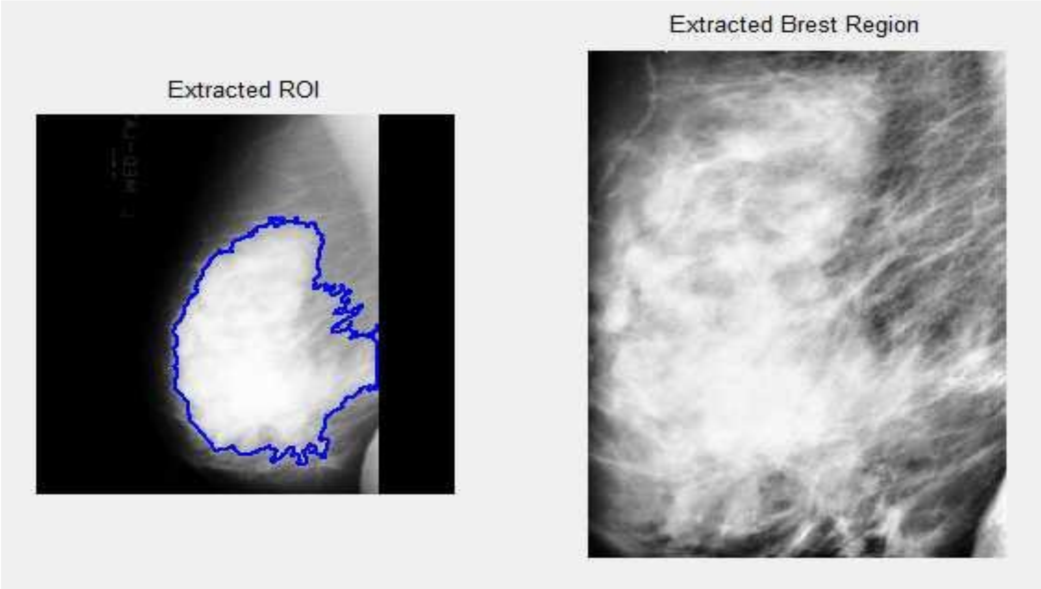


Fig 2: ROI (Reign of Interest) Extraction

3.3.1.1 First-Order Statistics Features

Without taking spatial dependence into account, statistical texture analysis relies on the statistical features of intensity histograms. The image's histogram summarizes the image's statistical data. You can use the picture histogram to get the image's first-order statistical data. The likelihood, the frequency with which the intensity occurs, as defined by the histogram and the total amount of pixels in the picture, is given as (Chevrefits and Cheriet, 2009: 608 - 620):

$$P_{ij} = \frac{g_{ij}}{N * M}$$

Where:

i,j=0,1,2..........L-1 $1 \leq i,j \leq L$

One of the most direct ways to describe texture involves using statistical properties of the image density distribution (Gonzales & Woods, 2008:853).

Let :

z: be r.v denoting intensity

p (z_i): is the matching histogram (probability of z_i)

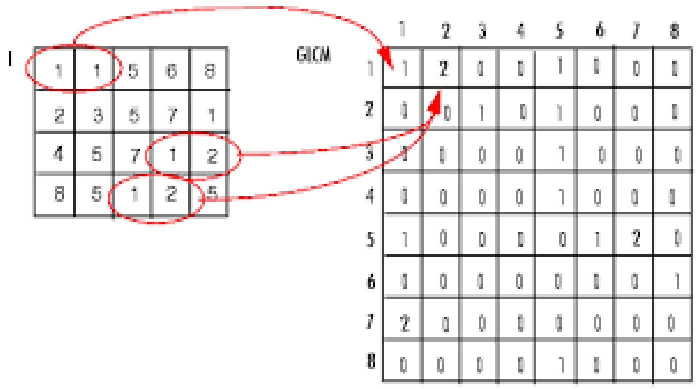
L: represents the number of distance intensity levels.

G:The dimension of the co-occurrence matrix is K*K, with element gij reflecting the number of times that pixel pairs with intensity Zi and Zj occur in f.

K: The row or column dimension of square matrix G.

3.3.1.2 Second - Order Statistics Co-occurrence Features.

As previously stated, characteristics obtained from first-order histograms are only useful locally and do not account for any regional data. To do this, we created gray-level spatial co-occurrence matrices that use the joint probability distribution of pairs of pixels to define 2nd order histogram properties. Gray level co-occurrence matrices (GLCM) were presented to help define textural characteristics. The values in the co-occurrence matrix indicate two consecutive pixels separated by a distance "d" and positioned in the picture at an angle "θ" (typically d = 1, 2,..., and θ=0o, 45o, 90o, 135o). Consider the picture in Fig (3), which displays a pair of pixels f(i, j) and f(m, n) separated by a distance "d" and at an angle "θ" relative to the horizontal axis. The angular connections between neighboring pixels determine the values in this symmetric matrix (Haralick et al., 1973:612).



Fig(3) Co-Occurrence matrix

Table 1 : Explanation and formula for all variables using (1^{st} and 2^{nd} order statistics)

Descriptor	**Explanation**	**Formula**
Mean	Mean provides a measure of average intensity.	$m = \sum_{i=0}^{L-1} z_i\, p(z_i)$
Variance	The second moment [the variance $\sigma^2 = m_2(i)$]. There are differences in the levels of gray and significant changes in density in the image.	$m_2(z) = \sigma^2 = \sum_{i=0}^{L-1} (z_i - m)^2\, p(z_i)$
Skewness	A measurement of the asymmetry of the distribution of gray values around the mean within the ROI.	$\mu_3 = \sigma^{-3} \sum_{i=1}^{L-1} (z_i - m)^3\, p(z_i)$
Median	The median gray value of pixels in the ROI.	-
Energy (Uniformity)	A measure of regularity in the interval [0,1]. Constant pictures have uniformity equal to one.	$E = \sum_{i=1}^{k} \sum_{j=1}^{k} (i - j)^2 p_{ij}$
Entropy	Determines the unpredictability of G's elements. Entropy is 0 when all pij are zero, and maximal when all pij are equal. Maximum value is $2\log_2 K$	$H = -\sum_{i=1}^{k} \sum_{j=1}^{k} p_{ij} \log_2 p_{ij}$

Perimeter	The length of the ROI's exterior boundary	$p = 2\pi r$
Area of Mass	Is the count of pixels in the region	$a = \pi r^2 , \pi = \frac{22}{7}$
Contrast	A measure of the intensity contrast between one pixel and its neighbor throughout the picture. The values range from 0 (when G remains constant) to $(k-1)^2$	$C = \sum_{i=1}^{k}\sum_{j=1}^{k}(i-j)^2 p_{ij}$
Circularity	The expression represents the circularity ratio.	$R_c = \frac{4\ \pi\ area}{p^2}$
Solidity	A ratio of the screen points of the block, particular to the convex shape.	$s = \frac{[Area]}{[Convex\ area]}$
Correlation	A measure of the degree to which the intensity of a pixel is related to the intensity of its neighbors throughout the image. This value ranges from -1 to 1.	$r = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(z_i - m_r)(z_i - m_c)p_{ij}}{\sigma_r \sigma_c}$ $\sigma_r \neq 0;\ \sigma_c \neq 0$

3.4 Artificial Neural Network (ANN)

The architecture of biological nervous systems, such as the human brain, served as inspiration for the development of an information processing system called an ANN. It optimizes its output by learning from input data, coordinating internal processing, and employing a dispersed network of many highly-connected processing neurons., based on the (Schaefer, et.al. 2009:2)'s assumptions that:

1. Information is processed by many simple components known as neurons.
2. Neurons communicate with one another through connecting links.
3. Every connection between neurons has a specific weight. This weight adjusts the incoming signal before passing it along to the next neuron.
4. Every neuron takes the total of its input signals, which are adjusted by weights, and applies a nonlinear activation function to figure out what its output signal will be.

A neural network consists of three main parts: (1) the way neurons are connected, called its architecture (see Figure 4), (2) the process used to assign weights to these connections, known as the training or learning algorithm, and (3) the activation function it uses.

The many basic processing components that make up a neural network are known as neurons, units, cells, or nodes. Given a certain weight, directed communication links connect each neuron to other neurons. The question of how to accomplish the desired behavior in the neural network by determining the weights for a specific problem becomes crucial as a result. The data storage and recall, pattern categorization, input-output pattern mapping, pattern grouping, and restricted optimization problem solution capabilities of neural networks are extensive. The inputs that a neuron has received determine its activation or activity level, an internal state that the neuron holds. In a typical neural network, one activated neuron will send out signals to several additional neurons. A neuron can only send one signal at a time, but it has the ability to share that signal with many other neurons at the same time (Fausett, 1994:4) (Jalil and Mahmood, 2012:190).

3.5 Multilayer Feed-forward Networks

Multilayer FFNs are a major type of neural network. They can address different challenges across different domains. The architecture of these networks consists of nodes: an input layer, one or more hidden layers of computing nodes, and an output layer of computing nodes. Input signals transfer through the network, progressing forward from one layer to the next.

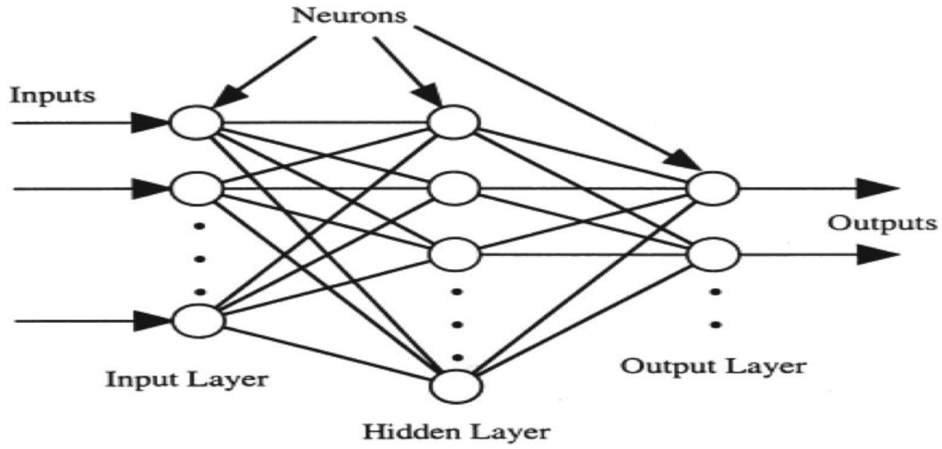


Fig 4: Neural network feed-forward architecture

For medical image processing, ANNs like FFNN are frequently utilized. Supervised and unsupervised networks are the usual components of neural network learning models. The training data set for supervised training contains several source-target pattern pairs. To increase the accuracy of the final outputs, the network processes the input data, compares them to the desired outputs, and adjusts their weights (Schaefer, et. al. 2009:4).

On the other hand, supervised FFNNs include back-propagation (BP) networks. A major step forward in the field of neural networks, it uses the gradient descent technique to reduce the overall squared error of the network's calculated output. As a result, its mistakes end up in the inner neurons rather than the output ones. When a stopping requirement is met, the procedures of recalculating the output and modifying the set of weights between the layers continue (Schaefer, et. al. 2009:4).

In backpropagation network training, there are three steps: feeding the input training pattern forward, calculating and backpropagating the associated error, and adjusting the weights. In order to properly categorize the inputs, this approach modifies the weights in the backpropagation network (BPN) for a particular set of training input-output pairings. For basic perceptron networks with differentiable units, the gradient-descent approach is the basis of this weight updating process. Once trained, the net's computations during the feed-forward phase are all that are needed for application. A learning rule that takes into account each training pattern and attempts to minimize the squared error; also called Least Mean Square (LMS). (Fausett, 1994:290), (Haykin, 1999:224).

Here is an overview of the steps involved in the backpropagation algorithm:
Step0. Initialize learning rate α and weights (set to small random values).
Feed-forward:
Step1. The input signal is received by the input unit (X_i, i=1,..., n) and sent to every unit in the hidden units.
Step2. Every hidden unit adds up its input signals with weights assigned to them

$$(H_j,\ j = 1,\ldots,p),\ H_j = v_{oj} + \sum_{i=1}^{n} X_i v_{ij}$$

Use the activation function to determine the output signal, $h_j = f(H_j)$ and send this signal to every unit in the layer output.

Step3. The output unit (Y_k, where k=1 to m) adds up the weighted signals it receives.

$$Y_k = W_{0k} + \sum_{j}^{p} h_j W_{jk}$$

To find its output signal, we use its activation function, $y_k = f(Y_k)$
Backpropagation of error:
Step 4. A unit of output the target value is sent to an output unit (Yk, k=1... m), which computes its error.

$$\delta_k = (t_k - y_k) f'(Y_k)$$

Weight correction or update Wjk is used, $\Delta W_{jk} = \alpha.\ \delta_k h_j$

Bias correction or update W_{0k} is used, $\Delta W_{0k} = \alpha.\delta_k$

It then transmits δ_k to the units in the layer beneath.

Step 5. Every hidden unit *(H_{jo}* j = 1, ... *,p)* adds up the delta inputs it receives from the units in the layer above.

$$\Delta_j = \sum_{k=1}^{m} \delta_k W_{jk}$$

To calculate its error information term, it is multiplied by the activation function's derivative. $\delta_j = \Delta_j\ f'(H_j)$

Weight correction or update V_{ij} is used, $\Delta V_{ij} = \alpha.\ \delta_j . X_i$

Bias correction or update V_{0j} is used, $\Delta V_{oj} = \alpha\ .\delta_j$

Update weights and biases:

Step 6. Every output unit (Y_k k=1,…m) *u*pdates its weights (j = 0,…,p) and bias.

$W_{jk}(new) = W_{jk}(old) + \Delta W_{jk}$

Every hidden unit *(Y_j j=1,...p)* updates its weights(i=1,…,n) and bias

$V_{ij}(new) = V_{ij}(old) + \Delta V_{ij}$

The network will continue to update the weights of any training process to obtain the optimum weights and then get the desired output to reach any better reconcile the model (test stop condition).

When we add a momentum term (μ) to the backpropagation, the adjustment to the weights takes into account both the current gradient and the gradient that preceded it. This helps in speeding up the convergence to the solution. (Fausett, 1994:305)

$$w_{jk}(t+1) = w_{jk}(t) + \alpha.\delta_k.h_j + \mu[w_{jk}(t) - w_{jk}(t-1)]$$

3.7 Classification Using Feed-forward Networks (FFNN)

It is also possible for neural networks to compute through classification. The collection of input patterns is comprised of a number of categorized groups or classes. Detection and categorization play a significant role in the decision-making process in contemporary medicine. The number and complexity of patient data that must be taken into consideration when crucial decisions are made continue to increase, which in turn increases the importance of autonomous algorithms. It is possible that they will support the decisions made by a physician; nonetheless, they cannot take their place. It would appear that the strategy of backpropagation of errors, which involves training multilayer feed-forward structures, was the most successful in overcoming classification issues (Berka et al. 2009:104).

With the help of multilayer FFNN, it is possible to do a wide variety of recognition and classification operations. The utilization of these networks as a means of problem resolution represents a significant divergence from the current wisdom. It is not necessary to have a formal mathematical model of the recognition or classification problem in order to use FFNN systems for training and recall purposes. Instead, a good solution may be discovered by employing the error back-propagation technique to alter the network parameters after an adequate training set and an appropriate network architecture has been constructed. This will allow for adequate network parameters to be adjusted. Rather than taking a methodical and formal approach, the answer is discovered via the process of solving the problem through experimentation and simulation. Neural network computation, as a result, offers approaches that are situated somewhere in the middle, between traditional engineering and artificial intelligence (Zurada, 1992: 221)

4. Application

The application of multilayer feed-forward neural networks for the identification and categorization of breast cancer images significantly aids oncologists in obtaining critical information. The application process Comprised of 100 gray-level mammography X-ray images for women's breast cancer, recorded in TIFF format (8-bit) with dimensions of 206 x 358 pixels, including 50 photos of benign masses and 50 images of malignant masses. The data source for the experiment in the suggested method was derived from the Digital Database for Screening Mammography (DDSM), a resource for mammographic images. Digital image processing involves identifying the region of interest (ROI) of a mass, followed by the extraction of a feature vector comprising statistical and geometric measurements to characterize the mass's shape, including Energy, Contrast, Correlation, Entropy, Area, Mean, Standard Deviation, Perimeter, Circularity, Integrated Density, Median, Skewness, and Solidity, for both benign and malignant masses, analyzed individually for each image.

Multilayer perceptron (MLP) neural networks are generally trained with the back-propagation (BP) technique, an application of the gradient method to feedforward neural networks (FFNNs) aimed at minimizing network faults, predominantly employed for supervised learning.

In the practical part using three programs to extract and analyzes data, which are: image J.V. 1.64a, SPSS software V.24 and Matlab 2015(V 8.5).

4.1 Network Structure

The constructed multi-layer back propagation neural network (BPNN) classifier comprises three layers: input, hidden, and output. A picture will be processed through the feature vector. The feature vector serves as the input to the Backpropagation Neural Network (BPNN) in picture classification. The input layer consists of 13 nodes (X1…X13), representing independent variables, which correspond to the extracted features and significantly impact the network's performance and complexity. The output nodes represent the classes of dependent variables, while the nodes in the hidden layer can be established through trial and error. Neural training in the hidden layer will commence with 2 nodes, increasing to a maximum of 10 nodes, with the optimal nodes identified through comparison based on Mean Square Error. Each connection, or synapse, between two neurons is given a weight. This weight starts as a random number between -0.5 and +0.5. This weight is adjusted during subsequent iterations during the network's training based on input and output data.

The Backpropagation (BP) algorithm applies the gradient method or other numerical optimization techniques to feed-forward artificial neural networks to decrease network defects.

This study uses the backpropagation technique to learn from the samples. The tan-sigmoid and log-sigmoid functions are utilized in the hidden and output layers, respectively, while the Levenberg-Marquardt is employed for weight adjustment during training.

4.2 Training and testing in neural network

The classification process consists of two main steps, which are the training phase and the testing phase. We begin the training stages to identify the neural network model that will help detect and classify images based on the features we extracted, by the above mentioned methods used to update the weights in FFNN with error back propagation algorithms. In the ANN based breast cancer classification, a total of 70 samples were used for training out of which include 35 benign and 35 malignant samples. During the classification process, the desired target values were “1” for benign and “2” for malignant.

Once the network is trained, it is important to test how well it performs. We take a set of samples, which contain 15 benign cases and 15 malicious cases. During the testing phase, we feed the input data into the network and check how well the output compares to what we expected. By looking at whether the results match or not, we can see how well the network is performing.

Neural network diagram consisted of 13 inputs, 5 hidden units, and 2 outputs, shown in figure 5.

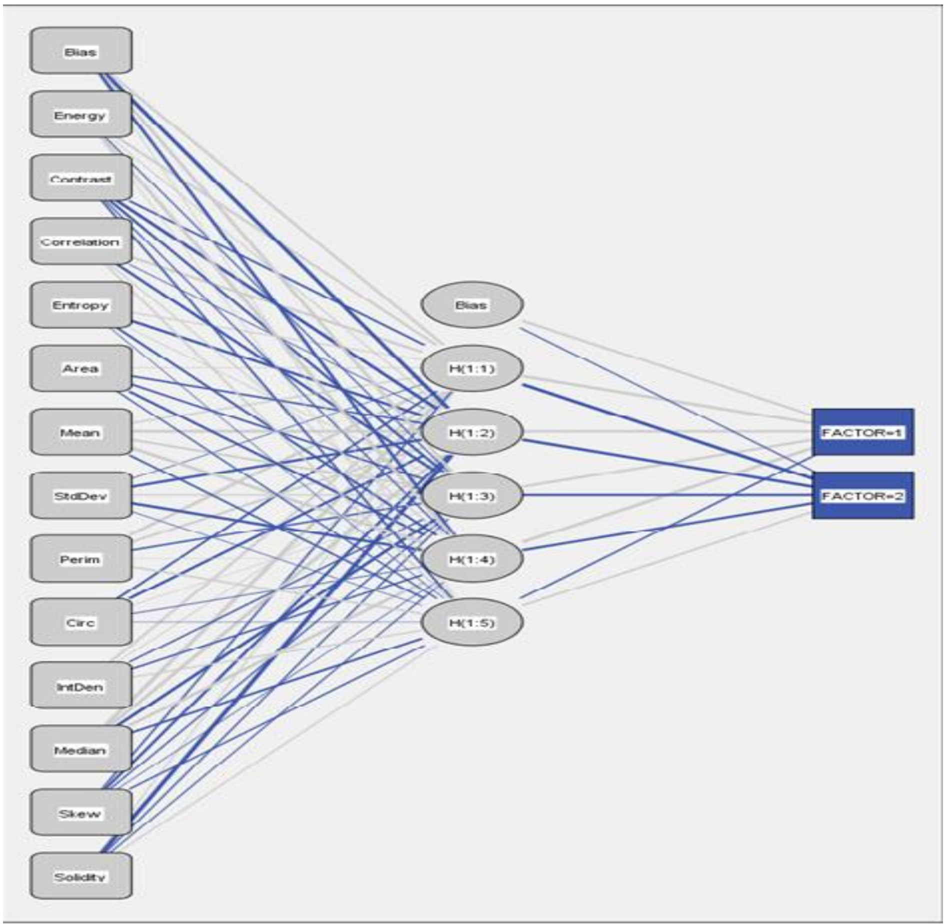


Fig 5: Neural network diagram consisted of 13 inputs, 5 hidden units, and 2 outputs.

4.3 Result of Detection and Classification

Employing multilayer feedforward neural networks to determine the optimal detection and classification of benign versus malignant breast masses in women, utilizing GLCM feature extraction from mammographic images. The conventional metric for evaluating classification performance was accuracy.

The results were presented in Table3.

Table (3) Classification results of the proposed method

		malignant		Benign		Percent Correct
		NO.	Percent	NO.	Percent	
Training	malignant	34	97.14%	1	2.86%	(34+32)/70 = **94.3%**
	benign	3	8.57%	32	91.43%	

Testing	**malignant**	**14**	**93.33%**	**1**	**6.67%**	(14+15)/30 = **96.7%**
	benign	**0**	**0%**	**15**	**100%**	
Overall Percent	**malignant**	**48**	**96%**	**2**	**4%**	(48+47)/100 =**95%**
	benign	**3**	**6%**	**47**	**94%**	

There are several ways to measure model performance. In the context of medical systems, it is common to discuss performance in terms of sensitivity, specificity, and accuracy, which can be described as follows (Schaefer, et.al. 2009:271):

Accuracy = (images classified correctly)/(all classified images) **=(95/100)*100= 95%**
Sensitivity = (images with malignant classified as malignant)/(all malignant images) = **(48/50)*100 = 96%**
Specificity = (benign images classified as benign)/(all benign images) **y=47/50)*100= 94%**

Experience focuses primarily on the ability of neural networks to analyze and classify (classification task usually involves with training and testing data). The results from the prediction neural networks using randomly selected sets are shown in Table 3, the network achieved an 95% of accuracy, 96% of Sensitivity and 94% of Specificity, of rate mammogram classification into malignant and benign, this result suggests that high performance using GLCM feature extraction, and the ability of the model to cancer detection.

4.4 Independent Variable Importance

The method determines the relative importance of explanatory variables to response variables identified in the supervised neural network by examining the model weights. The fundamental concept is that the strength of association of a particular explanatory variable with a specific response variable can be assessed by identifying all weighted connections between the relevant nodes. This involves pinpointing; all weights that connect the particular input node go through the hidden layer, and end up in the response variable must be identified. The weights reflect the relative influence of the information processed in the network, with input variables that lack relevance to the response variable being diminished by the weights. We repeat this step for each input variable until a list of weights associated with each input variable is compiled by the analyst. We measure how much each input affects the output. Then, we adjust these measurements to compare them fairly. This gives us a single number for each input, showing how important it is predicting the output (Garson,1991:48).

The importance of input variable which influences the output variable (detection of Cancer). Results of the neural network analysis, which evaluated 13 different variables to see which ones help in detecting cancer and which ones do not. These variables are ranked based on their effectiveness, as shown in Table 4

The importance of an independent variable refers to how much the predicted value of a network model changes with changes in that independent variable. The first four significant features extractions that have been found are mean (100%), median (74%), Standard deviation (67.7%) and Solidity (51.6%). The next two important variables have been perim and Area. Other variables were not very important, and the variable that emerged as the least important was Area.

Table (4) Independent Variable Importance

Variables	Importance	Normalized Importance	Rank
Mean	0.22	(0.22/.22)*100= **100%**	1
Median	0.1629	(0.1629/.22)*100= **74%**	2
Standard deviation	0.1491	(0.1491/0.22)= **67.7%**	3
Solidity	0.1135	(0.1135/0.22)*100= **51.6%**	4
Perim	0.0909	(0.0909/0.22)*100= **41.3%**	5
Area	0.078	(0.078/0.22)*100= **35%**	6
Skew	0.0525	(0.0525/0.22)*100= **23.8%**	7
Circ	0.0299	(0.0299/0.22)*100= **13.5%**	8
IntDen	0.0278	(0.0278/0.22)*100= **12.6%**	9
Energy	0.0259	(0.0259/0.22)*100= **11.7%**	10
Entropy	0.0199	(0.0199/0.22)*100= **9%**	11
Contrast	0.0172	(0.01672/0.22)*100**= 7.6%**	12
Correlation	0.0029	(0.0029/0.22)*100= **1.3%**	13

5. Conclusions and Recommendations

5.1 Conclusions

1. We obtained the high performance for proposed model to detect cancerous mammogram image using feed-forward neural network.
2. Obtained four important variables of the studied variables in a diagnosis which are mean, median, Standard deviation and Solidity.
3. We conclude that the use ANN can be effectively used in the detection and classification of cancer to help oncologists.
4. Our work demonstrates a wide range of methods and strategies that can be effectively combined to produce beneficial outcomes for humans.

5.2 Recommendations

1. Using Feed-forward NN **with** other types of images processes.
2. Probabilistic Neural Networks or Radial Neural Networks are used with digital mammograms for the detection of cancer.
3. The use of other variables extracted from digital mammograms to detect cancer with artificial neural networks.

References

1. Bălănică, V., I. Dumitrache, I. and Preziosi, L.(2013): " Breast Cancer Diagnosis based on Spiculation Feature and Neural Network Techniques ", International Journal of Computers, Communications & Control, Volume 8– No.3, pp. 354-365.
2. Basha, S. S. and Prasad, K. S.(2009): "Automatic detection of breast cancer mass in mammograms using morphological operators and fuzzy means clustering", Journal of Theoretical and Information Technology, pp.704-709.
3. Berka, P., Rauch, J. and Zighed, D. A. (2009): " Data Mining and Medical Knowledge Management: Cases and Applications", IGI Global , Hershey, New York, USA
4. Chevrefits.C and Cheriet.F.J. (2009): "Texture Analysis for Automatic Segmentation of Intervertebral Disks of Scoliotic Spines from MR Images". IEEE Transactions on information technology in biomedicine, vol.13 No. 4, 608 – 620.
5. Fausett, L. V. (1994): "Fundamentals of Neural Networks: Architectures, Algorithms, and Application", Prentice-Hall, New York.
6. Ganesan, N.,Venkatesh, K., Rama, M. A. and Palani, A. M. (2010): " Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data", International Journal of Computer Applications, Volume 1– No.26, pp. 76-85.
7. Garson, G.D.(1991):"Interpreting neural network connection weights". Artificial Intelligence Expert. Vol.6 No.4.pp 46-51.
8. Gonzales, C. & Woods, E. (2008)" Digital Image Processing". Third Edition, Prentice-Hall Publishing, New Jersey, United States of America.
9. Haralick, R.M., Shanmugam,K. and Dinstein,I. (1973):"Textural Features for Image Classification", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-3, pp. 610-621.
10. Haykin, S. (1999):" Neural Networks A Comprehensive Foundation", Second Edition, Prentice Hall International, USA.
11. Jalil, T.S. and Mahmood, S.H. (2012) 'Extended Kalman Filter of Training Feed Forward Artificial Neural Network', journal of kirkuk University For Administrative and Economic Sciences, vol. 2, no. 1, pp. 183-201.
12. Kadhim, D. A. (2012) "Development algorithm- computer program of digital mammograms Segmentation for detection of masses breast using Marker-Controlled Watershed in MATLAB environment ",The First Scientific Conference the Collage of Education for Pure Sciences, pp.114-123.
13. Lowis, X., Hendra, Y. and Lavinia, Z. (2015): "The Use of Dual-Tree Complex Wavelet Transform Based Feature for Mammogram Classification", International Journal of Signal Processing, Image Processing and Pattern Recognition, Volume 8– No.3, pp. 87-96.
14. Naranje, S., (2016): "Detection of Breast Cancer using ANN", International Journal on Recent and Innovation Trends in Computing and Communication, Volume4 – issue 4, pp. 675-677.
15. Pavitha, R., Hephzibah, S.T.J(2014): " Mammographic Cancer Detection and Classification Using Bi Clustering and Supervised Classifier ", International Journal of Innovative Research in Science, Engineering and Technology, Volume 3– Issue 1, pp. 1382-1389.
16. Pratt,W.K. (2007) "Digital Image Processing", A Wiley-Inter science Publication, John Wiley & Sons, New York, 4th Edition, pp.547-555.

17. Schaefer, G., Hassanien, A.E and Jiang, J. (2009):"Computational intelligence in medical imaging techniques and applications", Chapman & Hall/CRC, USA
18. Singh, B. K., Yadav, A., Shailaja Singh, S. (2011): " ANN based Classifier System for Digital Mammographic Images", International Journal of Computer Applications, Volume 35– No.13,pp. 39-42.
19. Utomo, C. P., kardiana, A. and yuliwulandari, R. (2014): "Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques", YARSI University, Jakarta, Indonesia,IJARAI,Vol.3,No.7. pp.10-14
20. Zhang, L., Wang, L., Wang, X., Liu, K., and Abraham, A. (2012): "Research of Neural Network Classifier Based on FCM and PSO for Breast Cancer Classification", E. Corchado et al. (Eds.): Part I, LNCS 7208, pp. 647–654.
21. Zhang,Y. J. (1996): " A Survey on Evaluation Methods for Image Segmentation", Pattern. Recognition, Vol. 29, No. 8, pp. 1335–1346.
22. Zurada, Jacek M.(1992): "Introduction to Artificial Neural Systems", West publishing co., USA.