



Classification of Circular Mass of Breast Cancer Using Artificial Neural Network vs. Discriminant Analysis in Medical Image Processing

Karzan Faidhi Hamad¹, Bulent Celik² and Rizgar Maghded Ahmed³

^{1,2}Department of Statistics, Faculty of Science, Gazi University, Ankara, Turkey

³Department of Statistics and Informatics, College of Administration & Economics, Salahaddin University, Erbil, Iraq

Article information

Article history:

Received January 18, 2024

Revised: May 24, 2024

Accepted Ma 26, 2024

Available June 1, 2024

Keywords:

Artificial Neural Network

Linear Discriminant Analysis

Medical imaging (Mammogram image), ROI, Image processing

Correspondence:

Karzan. F. Hamad

karzan.hamad89@outlook.com

Abstract

In recent years, there has been a notable increase in interest regarding intelligent classification techniques rooted in Machine Learning within the domain of medical science. Specifically, machine learning, a pivotal area of artificial intelligence, has been extensively utilized to aid medical professionals in predicting and diagnosing various diseases. This study applies two distinct machine learning algorithms to address a medical diagnosis concern related to circular masses in breast cancer. The dataset encompasses 150 cases of breast cancer patients. The primary objective is to assess and compare the effectiveness of artificial neural networks (ANNs) and linear discriminant analysis (LDA) classifiers based on key criteria: accuracy, sensitivity, specificity, and the kappa coefficient in predicting circular masses within breast cancer. Results indicate that the performance of the ANN classifier surpasses that of the LDA model, achieving an accuracy of 97.7%, sensitivity of 95%, specificity of 100%, and a kappa coefficient of 95.31%. Additionally, the final fitted models unveil the pivotal factors significantly influencing circular masses in breast cancer, highlighting Solidity and Entropy as the most critical variables.

DOI [10.33899/IQJOSS.2024.0183231](https://doi.org/10.33899/IQJOSS.2024.0183231), ©Authors, 2024, College of Computer Science and Mathematics University of Mosul.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the escalating incidence of breast cancer cases has underscored the necessity for comprehensive research into this disease and its influencing factors. This has led to a heightened focus on utilizing statistical methods and artificial intelligence techniques to explore and address breast cancer. Notably, artificial techniques have become particularly prominent in current research endeavors. The risk of experiencing more severe forms of breast cancer tends to rise over a woman's lifetime. Particularly, middle-aged women show a notably higher prevalence of this illness compared to younger women (Henriksen et al., 2019). Breast cancer begins within the body's breast cells, which can swiftly invade nearby tissues. The impact of this disease can vary based on the type of cancer, risk factors, and the patient's age. Typically, detecting breast cancer occurs either through the discovery of a breast lump or via mammogram screenings (Falk et al., 2018). Presently, machine learning algorithms leveraging computational techniques provide a streamlined path for analyzing and classifying cancer, significantly enhancing the process. In 2010, Siraj, Salahuddin, and Mohd Yusof explored the application of neural networks and logistic regression techniques in image processing to discern the distinctive features of flower images (Siraj et al., 2010). In their 2014 study, George and colleagues introduced a diagnostic system for breast cancer employing various support vector machines and neural networks. Their findings indicated accuracy rates ranging from 76% to 94% when tested on a dataset comprising 92 images (George et al., 2014). In their 2022 research, Marion and colleagues delved into the classification of breast cancer types, distinguishing between "malignant" and "benign". They employed combinations of

(LDA) with (SVM) and (LDA) with Random Forest (RF), achieving impressive accuracy rates of 96.4% and 95.6%, respectively. Notably, their findings underscored the superiority of SVM with LDA over the use of LDA with RF (Marion, O. et al., 2022).

This paper employs (LDA) and (ANN) to accurately predict the classification of circular masses in breast cancer. The aim is to determine which model is more effective in assisting doctors with early-stage disease diagnosis, thereby reducing the risk of fatal outcomes.

2. Methodology

2.1. Artificial neural network

Artificial intelligence has been a topic of interest in scientific and medical research for many years. The first use of a biological neural network in computer simulation dates to the 1950s (McCulloch et al., 1943). In 1951, McCulloch and Pitts introduced the first artificial neuron. Between 1982 and 1987, Hopfield, Kohonen, McClelland, and Rumelhart developed mathematical models for practical applications (Kohonen T., 1982). Neural networks have four main applications in medicine: modelling, processing bioelectrical signals, diagnosis, and prognosis.

Conventional classifiers often use a linear approach to problem-solving, which can lead to the incorrect identification of complex relationships between predictors and dependent variables. To handle this issue, an (ANN), which originates from the biological neural network, is used. ANN is a computing system that takes a non-linear approach to problem-solving. It consists of artificial neurons or nodes, like the biological neurons in the brain, which are interconnected as "edges" and transmit data from one node to another, like synapses in the brain. The output of each neuron is generally computed as a non-linear function of its inputs. These artificial neurons and edges also have some "weight" that is adjusted as the ANN learns (Sarle, 1994).

Neural networks consist of nodes that are grouped into layers. Each layer can have different types of transformational functions that are applied to their inputs. Moving from the first layer to the last in a neural network is called an "epoch," and a model can perform multiple epochs.

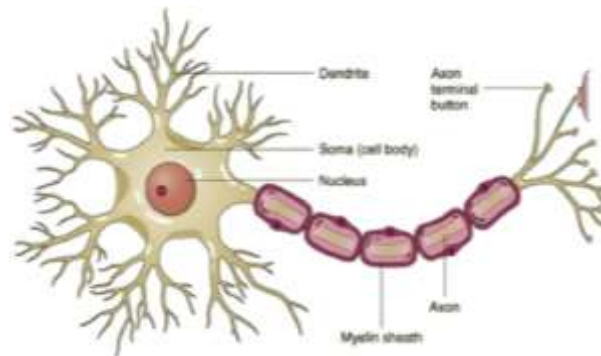


Figure (1) Biological Neural Network

Based on Figure (1), the following points can be inferred:

- 1- Dendrites:** Each neuron has entry points that receive in the form of electrical impulses from other neurons in the network.
- 2- Cell Body (Soma):** The system processes the information received from the dendrites and decides on what action to take.
- 3- Axon terminals:** outputs in the form of neurons transmit electrical impulses to communicate with other neurons (Omar & Rizgar, 2020) and (Graupe, 2013).

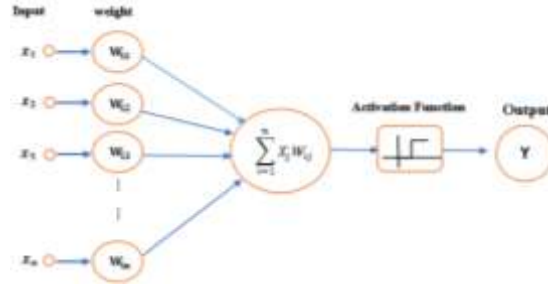


Figure (2) Artificial Neural Networks

Figure (2) illustrates that the neural network is composed of three distinct segments, which are as follows:

The Input Layer consists of the values of x .

The hidden layer contains the values of w and the resulting processes.

The output Layer of the model includes the value of y .

The practical use of artificial neural networks lies in their ability to modify the weights of the nodes connecting the neurons to produce a desired output.

$$N_{pj}^{L+1} = \sum_{i=1}^{n^L} W_{ij}^L Out_i^L + Bias_j^{L+1}$$

$$Out_{pi}^{L+1} = f(N_{pj}^{L+1}) = \frac{1}{1 + e^{-BN_{pj}^{L+1}}}$$

2.2. Artificial Neural Network Learning Algorithms

There are different learning algorithms used to train (ANNs), each with unique characteristics and applications. Here are two of the fundamental ones:

- A. Feedforward Neural Networks (FNN) / Multilayer Perceptrons (MLP):** These networks are composed of input, hidden, and output layers and are trained using backpropagation. The hidden layers allow the networks to learn complex representations (Sharma et al., 2015).
- B. Backpropagation:** Backpropagation is a crucial algorithm that is used to train (ANNs). Computing the gradient of the loss function regarding the network weights is necessary. This gradient information is then used to update the weights, which helps to improve the network's ability to make predictions over time (Reed & Marks, 1999).

$$W_{L+1} = W_L + \alpha_L \cdot g_L$$

Whereas

W_L : Weights vector.

α_L : learning rate.

g_L : current slope.

2.3. Discriminant analysis

Discriminant analysis is a crucial tool in multivariate statistical analysis, employed to differentiate between two or more groups of societies. Its primary purpose is to distinguish identical or overlapping societies with similar characteristics. In essence, discriminant analysis is a statistical approach using a set of variables to discern between multiple groups via a specific discriminant function, a linear combination of explanatory variables. These coefficients are derived through measurements or standardization of the items (David, 1978).

This method, also referred to as the Fisher function, serves as a valuable classification tool, contingent upon meeting certain conditions. Notably, the explanatory variables should exhibit a normal distribution, and the variance-covariance matrices should be equal. Within this methodology, researchers select distinct variables capable of differentiating between two populations of interest, generating a discriminant function. This function aids in classifying future observations into the respective groups (Tharwat et al., 2017). Optimal discrimination is achieved when the ratio of variance between groups to variance within groups is high (Kan et al., 2015) and (Morrison, 1976). The five general steps of the LDA classifier are:

- I. Compute the d-dimensional mean vectors for different classes in the dataset.
- II. Compute the scatter matrices of both (within-class and between-class).
- III. Compute the Eigen vectors (e_1, e_2, \dots, e_d) and corresponding Eigen values $(\lambda_1, \lambda_2, \dots, \lambda_d)$ for the scatter matrices calculated above.
- IV. Sort Eigenvectors by decreasing Eigenvalues and select k with maximum Eigenvalues to generate $d \times k$ matrix W .
- V. Use the eigenvector matrix W to transform the samples on to different subspaces. This can be done by matrix multiplication as $Y = X \times W$.

2.4. Representation of Digital Image Processing

To represent images, we use two-dimensional functions denoted by $f(x, y)$. The spatial coordinates (x, y) are positive scalar quantities whose physical meaning is determined by the source of the image. The values of the image span the grayscale. When an image is created from a physical process, its values are proportional to the energy radiated by a physical source, such as electromagnetic waves. Therefore, $f(x, y)$ must be nonzero and finite.

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

These components are called illumination (i) and reflectance (r) and are denoted by $i(x, y)$ and $r(x, y)$. Moreover, the product of two functions combines results in $f(x, y)$:

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

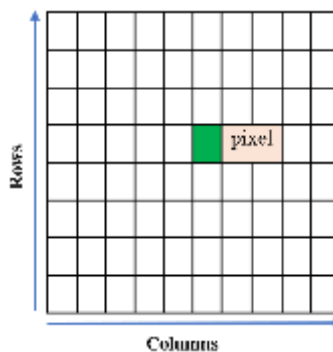
Where

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

The grey level value at point (x, y) differs from $(x + 1, y + 1)$, symbolized by (G) .

$$\begin{aligned} G_{max} &\geq G \geq G_{min} \\ r_{max} \times i_{max} &\geq G \geq r_{min} \times i_{min} \\ \text{Level } 255 &\geq G \geq 0 \end{aligned}$$

When an image $f(x, y)$ is sampled, it results in a digital image with M rows and N columns. As a result, the coordinates (x, y) become discrete values. For easy notation, we use integer values for these coordinates, namely $(X = 0, 1, 2, \dots, M - 1)$ and $(Y = 0, 1, 2, \dots, N - 1)$. The coordinate values at the origin are $(x, y) = (0, 0)$. The coordinate values along the first row of the image are represented by $(x, y) = (0, 1)$ (Gonzales, C. & Woods, E., 2008).



A digital image is a 2D grid of pixels, each with an intensity value and a location address specified by its row and column number.

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

An image is made up of small units called pixels, also known as image elements or picture elements.

$$M, N \geq 0, \text{ Number of gray levels} \geq 0$$

2.5. Medical Images

Medical images are a unique type of image that is primarily used for diagnosing patient cases. There are multiple methods to obtain medical images, with X-ray being the most popular option due to its simplicity, low cost, and minimal radiation exposure. The acquisition of medical images is an essential tool in healthcare, playing a critical role in disease diagnosis and treatment planning. Clinicians utilize various image modalities, including Computed Tomography (CT) Scans, Magnetic Resonance Imaging (MRI), X-ray, Ultrasound images, Mammography, and others to diagnose diseases accurately and efficiently plan treatments. Despite advancements in medical imaging, it still poses challenges from image formation to the final analysis (Al-Samaraie & Al- Saiyd, 2008).

2.6. Feature Extraction

The examination of breast cancer in this study begins with the extraction of features from mammogram images, as shown in Figure 1. Feature extraction aims to transform the raw mammograms into a set of features that can be used as input for any classification algorithm (Abirami et al., 2017). This step provides seven statistical features from the mammograms: entropy, area, perimeter, width, height, feret, and solidity. These extracted features are then used as input for the Linear Discriminant Analysis (LDA) and Artificial Neural Network Algorithm (ANN) classifiers, which are discussed in the following subsection.

2.7. Circularity measurement (Circ)

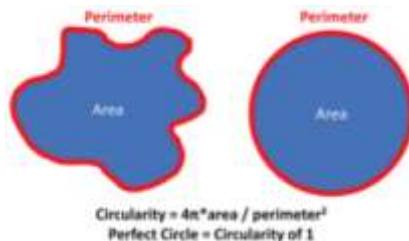
The circularity measurement (Circ.) is a mathematical formula used to assess how closely an object's shape resembles that of a perfect circle. It is computed as 4π multiplied by the object's area divided by the square of its perimeter.

$$Circ. = 4\pi * (Area / Perimeter^2)$$

Interpretation: A Circ. value of 1.0 denotes a perfect circle, indicating that the object's shape closely resembles a circle. As the Circ. value decreases toward 0.0, it signifies an increasingly elongated or less circular shape.

This metric finds applications in various scientific and technical fields, including image analysis, material sciences, and particle characterization. It aids in quantifying the roundness or circular nature of diverse objects, helping researchers and analysts to understand and categorize shapes based on their deviation from circularity (Ferreira & Rasband, 2012) (Gonzales, C. & Woods, E., 2008).

It's essential to note that for very small particles or objects, the accuracy of the Circ. measurement might be limited due to the specific constraints and considerations associated with their size, which can affect the reliability of the computed values.



2.8. ROI (Region of Interest)

refers to a specific part of an image that needs to be filtered or manipulated. This area can be represented by a binary mask image, where the pixels that correspond to the ROI are marked as 1, and the pixels outside the ROI are marked as 0 (Gonzales, C. & Woods, E., 2008).

2.9. Confusion Matrix

Evaluations are conducted using the confusion matrix to assess the success rates of classification models. This matrix provides valuable insights into both actual and predicted classifications within a classification system. Typically, the performance of these systems is thoroughly assessed by analyzing the data encapsulated within the matrix.

Table (1) Confusion Matrix

Classification	Observations		
	Negative	Positive	
Prediction	Negative	True negative (TN)	False negative (FN)
	Positive	False positive (FP)	True positive (TP)

The confusion matrix offers valuable insight into the severity of misclassifications. The numbers along the main diagonal of the matrix represent correct classifications.

A. Accuracy

The accuracy rate signifies the percentage of test set samples accurately classified by the model (Goyal, A & Mehta, 2012). It's essential that the test set remains independent of the training set to prevent overfitting. Essentially, the accuracy of a classification model represents its overall success rate, calculated by dividing the number of correct classifications by the total number of classifications (Baker, 2009):

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + FN + TP}$$

B. Sensitivity

Constructing artificial neural network models is a crucial task that involves identifying the variables that affect the model's performance and determining the parameters that hinder its performance. To evaluate the model's effectiveness, we can calculate the ratio of true positives to the sum of true positives and false negatives (Fulk & Sazonov, 2011).

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

C. Specificity

This criterion evaluates the model's proficiency in predicting true negatives across all available categories, commonly referred to as the actual negative ratio.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

D. Kappa

Kappa statistics is a valuable tool for evaluating the accuracy of individual cases and determining the reliability and validity of collected data. It measures how well-predicted values match observed values, allowing assessment of the model's performance and distinguishing between accurate results and those that occur by chance.

$$K = \frac{P_0 - P_e}{1 - P_e}$$

Here, P_0 denotes the observed concordance rate, and P_e denotes the randomly expected concordance rate (Landis & Koch, 1977).

3. Result

In this section, we will discuss the proposed steps to process medical mammogram images of infected breast masses. We will use a combination of statistical and geometric measurements to analyze the regions of interest (ROI) of digital images. Our approach involves (LDA) and (ANN), which has yet to be used in previous research studies. The following programs were used to analyze digital medical images:

1. Image J 1.45r by the National Institutes of Health (USA).
2. SPSS (V: 26).

3.1. Data Collection

Digital image processing deals with two-dimensional image data sources. For this research, 100 digital mammogram images were collected - 100 images of breast cancer. These images were stored as TIFF files with 8-bit color depth. The data was collected from the maternity hospital in Erbil.

3.2. Selection ROI of Masses

This research involved processing digital images and extracting measurements using Image J1.45r software. The steps are outlined below:

We need to identify the (ROI), which includes the benign and malignant mass.

To accurately describe the shape of the mass, we need to identify and measure statistical and geometric variables.

Microsoft Excel (2010) is used to save and export data to the SPSS program.

Then, Perform the necessary statistical analyses using multiple linear regression.

Dependent variable: We made circularity the dependent variable and coded any value above 0.5 as 1 and any value below 0.5 as 0.

$$Y = 0 \text{ (irregular shape or increasingly elongated shape)}$$
$$Y = 1 \text{ (perfect circle)}$$

Independent variable: The following variables are derived from ROI: (Entropy, Area, Perimeter, Width, Feret, and Solidity).

3.3. Applying Artificial neural network in mammogram Images

The dataset was randomly split into training and test datasets, with 70% (107 patients) allocated for training and the remaining 30% (43 patients) for testing. To ensure a fair assessment and mitigate data partitioning effects, all classification methods underwent evaluation using 10-fold cross-validation, averaged across 10 separate partitions. This approach allowed for a comprehensive comparison of classification performance metrics.

3.4. Performance Evaluation of Models Applied

After partitioning the data into training and testing sets, the model development commenced using the training dataset, consisting of 107 cases: 69 cases categorized as 0 (irregular shape) and 48 cases categorized as 1 (perfect circle). Despite this partition, the distribution of the response remained unchanged, revealing 69 cases identified as irregular shapes and 48 cases recognized as perfect circles.

Table (2) Confusion Matrix and Statistics for Training Dataset of ANNs

Classification		Observation	
		irregular shape 0	perfect circle 1
Prediction	irregular shape 0	69	1
	perfect circle 1	0	37
Model Accuracy		99.1%	
Model Sensitivity		97.36%	
Model Specificity		100%	
Error Rate of Classification		0.99%	
Kappa Coefficient		97.93%	

Table 2 demonstrates the exceptional performance of the Neural Network model in effectively classifying 106 out of 107 cases available. The model boasts an impressive Accuracy of 99.1%. Notably, both its sensitivity and specificity surpass 95%, standing at 97.36% and 100%, respectively. This indicates the model's ability to accurately predict outcomes based on the entered independent variables, achieving 97.36% accuracy in sensitivity and an impeccable 100% in specificity. The kappa coefficient, a measure of agreement, stands at an outstanding 97.93, showcasing near-perfect alignment with the model's predictions.

Table (3) Confusion Matrix and Statistics for Training Dataset of ANNs

Classification		Observation	
		irregular shape 0	perfect circle 1
Prediction	irregular shape 0	23	1
	perfect circle 1	0	19
Model Accuracy		97.7%	
Model Sensitivity		95%	
Model Specificity		100%	
Error Rate of Classification		2.3%	
Kappa Coefficient		95.31%	

Table 3 showcases the impressive performance of the neural network model, accurately classifying 42 out of 43 cases a remarkable feat. While compared to Table 1, there was a decline in all coefficients, the model still maintains a strong level of accuracy and sensitivity. With an accuracy rate of 97.7% and a sensitivity of 95%, our model reliably predicts breast cancer occurrences by 95% based on independent variables. Notably, the model achieves a flawless specificity of 100%, ensuring precise predictions. The Classification Error Rate remains minimal at 2.3%. Moreover, the kappa coefficient, scoring 95.31%, signifies an almost perfect alignment with the model's predictions. In summary, the artificial neural network models applied to the testing dataset have delivered exceptionally promising results.

Table (4): Independent Variable Importance for ANNs

Garson algorithm for variable importance	
overall	
Solidity	0.25
Perimeter	0.22
Entropy	0.15
Width	0.14
Area	0.11
Feret	0.10

Table 4 delineates the significance of various factors impacting the circularity of breast cancer, as analyzed through of (ANNs). Among these factors, Solidity emerges as the most influential, commanding a substantial 25% influence on the dependent variable, Circularity. This dominance is followed closely by the impactful presence of Perimeter at 22%, Entropy at 15%, and Width at 14%, each contributing distinctly to the modulation of the observed circularity. Conversely, within this spectrum of influence, the variables of Area and Feret emerge as the least affected among the independent variables, registering at 11% and 10% respectively. Though their impact may seem comparatively subdued, their presence adds valuable nuances to the multifaceted understanding of Circularity's dependencies within this computational framework.

Table (5) Confusion Matrix and Statistics for Training Dataset of ANNs

Classification		Observation	
		irregular shape 0	perfect circle 1
Prediction	irregular shape 0	70	4
	perfect circle	6	41
Model Accuracy		91.74%	
Model Sensitivity		92.11%	
Model Specificity		91.11%	
Error Rate of Classification		8.26%	
Kappa Coefficient		82.47%	

Table 5 demonstrates the exceptional performance of the Neural Network model in effectively classifying 111 out of 121 cases available. The model boasts an impressive Accuracy of 91.74%. Notably, both its sensitivity and specificity surpass 90%, standing at 92.11% and 91.11%, respectively. This indicates the model's ability to accurately predict outcomes based on the entered independent variables, achieving 92.11% accuracy in sensitivity and an impeccable 91.11% in specificity. The kappa coefficient, a measure of agreement, stands at an outstanding 82.47, showcasing near-perfect alignment with the model's predictions.

Table (6) Confusion Matrix and Statistics for Training Dataset of ANNs

Classification		Observation	
		irregular shape 0	perfect circle 1
Prediction	irregular shape 0	18	3
	perfect circle	0	8
Model Accuracy		89.66%	
Model Sensitivity		100%	
Model Specificity		72.73%	
Error Rate of Classification		10.34%	
Kappa Coefficient		76.8%	

Table 6 showcases the impressive performance of the neural network model, accurately classifying 26 out of 29 cases a remarkable feat. While compared to Table 1, there was a decline in all coefficients, the model still maintains a strong level of accuracy and sensitivity. With an accuracy rate of 89.66% and a sensitivity of 100%, our model reliably predicts breast cancer occurrences by 95% based on independent variables. Notably, the model achieves specificity of 72.73%, ensuring a good prediction. The Classification Error Rate remains minimal at 10.34%. Moreover, the kappa coefficient, scoring 76.8%, there is Substantial agreement with the model's predictions. In summary, the artificial neural network models applied to the testing dataset have delivered good promising results.

Table (7): Independent Variable Importance for DA

LD1	
Entropy	-1.1770
Area	-0.0015
Perim	0.0057
Width	0.0024
Feret	-0.0014
Solidity	12.5749

Table 7 outlines the factors influencing breast cancer circularity using (LDA). Among these, Solidity stands out as the most influential, significantly impacting circularity by 12.5749 units. The following closely are entropy (-1.1770), perimeter (0.0057), and width (0.0024), each contributing distinctly to circularity. In contrast, Area and Feret emerge as less impactful variables in this analysis. Despite their comparatively smaller influence, their inclusion enriches the comprehensive understanding of Circularity's dependencies in this computational framework.

3.5. Comparison between models

Table 8 displays a comparison solely for the testing dataset. This comparison identifies the superior classifier based on this dataset. The findings indicate that the ANN model outperformed LDA, demonstrating higher accuracy and efficiency in classification.

Table (8) Comparison between models

Model	ANN	LDA
Model Accuracy	97.7%	89.66%
Model Sensitivity	95%	100%
Model Specificity	100%	72.73%
Kappa Coefficient	95.31%	76.8%

Table 8 provides a summary of the performance metrics. The ANNs achieved an accuracy of 97.7%, surpassing the LDA model's accuracy of 89.66%. This underscores the preference for the ANNs due to their higher accuracy, a critical factor in determining the best model. Regarding sensitivity, the ANNs scored 95%, slightly lower than the LDA's perfect 100%. Conversely, in specificity, the ANNs excelled at 100%, while the LDA lagged at 72.73%. Additionally, the Kappa

Coefficient for ANNs stood notably higher at 82.47% compared to LDA's 76.80%. This means that the ANN model's last model has a complete preference.

The hierarchy of significance for variables in both the ANN and LDA models is arranged from the most significant to the least significant in the following order:

Table (9) The most important variables

Variables in ANN Mode	Variables in LDA Mode
Solidity	Solidity
Perimeter	Entropy
Entropy	Perimeter
Width	Width
Area	Area
Feret	Feret

According to table 9, it's evident that both the solidity, perimeter, and entropy variables hold utmost significance in influencing the Circle in both the (ANN) and (LDA) models, followed by the width and area variables.

4. Conclusion

This study demonstrates the remarkable potential of machine learning, specifically (ANNs), in the realm of medical diagnostics, focusing on circular masses in breast cancer. By leveraging a dataset of 150 cases, the comparison between ANNs and (LDA) reveals ANNs' superior performance in predicting and diagnosing these masses. With an accuracy of 97.7%, sensitivity of 95%, specificity of 100%, and a Kappa Coefficient of 95.31%, the ANN classifier outperforms LDA. Moreover, the study unveils critical insights into the key factors influencing circular masses in breast cancer, emphasizing Solidity, Perimeter, and Entropy as significant variables. This research underscores the potential of machine learning algorithms, particularly ANNs, as valuable tools aiding medical professionals in accurate and efficient disease prediction and diagnosis within the realm of breast cancer.

Acknowledgment

The authors are very grateful to the University of Mosul, and College of Computer and Mathematical Science, which helped improve this work's quality.

Conflict of interest

The author has no conflict of interest.

References

1. Abirami J, Jothi G, Brindha P. (2017). "Microscopic, physicochemical and phytochemical screening of *Crateva magna* (lour) DC". (leaf). *Int J Curr Pharm Res*, Vol. 9, Issue 5, pp.201-204.
2. Al-Samaraie, M. F & Al Saiyd, N. A., (2008) "Medical Colored Image Enhancement using Wavelet Transform Followed by Image Sharpening," *Ubiquitous Comput. Commun. J*, Vol. 6, pp. 1-8.
3. Baker, R. S. J. d. (2009). "Data Mining". In *International Encyclopedia of Education*, Third Edition. Oxford, UK: Elsevier.
4. David G. Kupper, L (1978), "Applied Regression Analysis and other multivariate methods", The University of North Caroline and Chapel Hill.
5. Falk D, Cubbin C, Jones B, et al. (2018) "Increasing breast and cervical cancer screening in rural and border Texas with a friend to a friend plus patient navigation." *J Cancer Edu*, 33, 798-05.
6. Ferreira, T. & Rasband, W. (2012) "ImageJ User Guide". *ImageJ/Fiji* 1.46.
7. Fulk, G. D., & Sazonov, E. (2011). "Using sensors to measure activity in people with stroke". *Topics in Stroke Rehabilitation*, Vol. 18, No. 6), pp. 746–757.
8. George, Y. M., Zayed, H. H., Roushdy, M. I. & Elbagoury, B. M. (2014) "Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Images". *IEEE Systems Journal*, 8, 949-964.
9. Gonzales, C. & Woods, E. (2008) "Digital Image Processing". Third Edition, Prentice-Hall Publishing, New Jersey, United States of America. ISBN:978-0-13-168728-8.
10. Goyal, A. & Mehta, R. (2012) "Performance Comparison of Naïve Bayes and J48 Classification Algorithms", (*International Journal of Applied Engineering Research* Vol 7, No 11).
11. Graupe, D. (2013). "Principles of artificial neural networks", World Scientific.
12. Henriksen EL, Carlsen JF, Vejborg IM, Nielsen MB, Lauridsen CA (2019). "The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review". *Acta Radiol*, 60, 13-8.
13. Kan M, Shan S, Zhang H, Lao S, Chen X (2015). "Multi-view discriminant analysis". *IEEE Trans Pattern Anal Mach Intell*, 38, pp.188-94.
14. Kohonen T. (1982) "Self-organized formation of topologically correct feature maps". *Biological cybernetics*, Publishing model Hybrid; vol. 43: pp. 59-69.
15. Landis, J. R., & Koch, G. G. (1977). "The Measurement of Observer Agreement for Categorical Data. *Biometrics*", Vol. 33, No. 1, pp. 159-174
16. Marion, A., Micheal, A., Moses, M. & Oludayo, O. (2022). "A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis", *Applied Sciences*,12(22).
17. McCulloch WS, Pitts W. (1943) "Bull Maths Biophysics"; vol. 5, pp. 115-118.
18. Morrison, D. F. (1976): "Multivariate Statistical Methods". (2nd Edition). Mc Graw-Hill Book Company, Inc. New York, USA.
19. Omar, Q. and Rizgar, A. (2019). "Prediction and Factors Affecting of Chronic Kidney Disease Diagnosis using Artificial Neural Networks Model and Logistic Regression Model", *Iraqi Journal of Statistical Science*, vol (16), issue (1): pp. 140-159.
20. Reed, R. and Marks, R. J., (1999). "Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks", Massachusetts Institute of Technology, England.
21. Sarle, W. S. (1994). "Neural networks and statistical models".
22. Siraj, F., Salahuddin, M. & Shahrul, Y., (2010) "Digital Image Classification for Malaysian Blooming Flower". *IEEE, Second International Conference on Computational Intelligence, Modelling and Simulation*, PP. 33 – 38.
23. Tharwat A, Gaber T, Ibrahim A, Hassanien AE (2017). "Linear discriminant analysis: A detailed tutorial". *AI Commun*, 30, 169-90.
24. V. Sharma, S. Sharma, N. Danish, and A. Mundari (2015) "Neural Network a Supervised Machine Learning Algorithm," *Int. J. Eng. Dev. Res. IJEDR*, vol. 3, no. 2, pp. 662–668.

تصنيف الكتلة الدائرية لسرطان الثدي باستخدام الشبكة العصبية الاصطناعية مقابل التحليل التمييزي في معالجة الصور الطبية.

كارزان فيدهي حمد¹ ، بولنت سيليك² و رزكار مغديد أحمد³

^{1,2} قسم الإحصاء ، كلية العلوم ، جامعة غازي ، أنقرة ، تركيا

³ قسم الإحصاء والمعلوماتية ، كلية الإدارة والاقتصاد ، جامعة صلاح الدين ، أربيل ، العراق.

الخلاصة: في السنوات الأخيرة ، كان هناك زيادة ملحوظة في الاهتمام بتقنيات التصنيف الذكية المتجذرة في التعلم الآلي في مجال العلوم الطبية. على وجه التحديد ، تم استخدام التعلم الآلي ، وهو مجال محوري للذكاء الاصطناعي ، على نطاق واسع لمساعدة المهنيين الطبيين في التنبؤ بالأمراض المختلفة وتشخيصها. تطبق هذه الدراسة خوارزميتين متميزتين للتعلم الآلي لمعالجة مشكلة التشخيص الطبي المتعلقة بالكتل الدائرية في سرطان الثدي. تشمل مجموعة البيانات 150 حالة من مرضى سرطان الثدي. الهدف الأساسي هو تقييم ومقارنة فعالية الشبكات العصبية الاصطناعية (أنس) والتحليل التمييزي الخطي (لدا) المصنفتان على أساس المعايير الرئيسية: الدقة والحساسية والنوعية ، ومعامل كابتا في التنبؤ الجماهير الدائرية داخل سرطان الثدي. تشير النتائج إلى أن أداء المصنف آن يفوق أداء نموذج لدا ، وتحقيق دقة 97.7% ، وحساسية 95% ، وخصوصية 100% ، ومعامل كابتا 95.31%. بالإضافة إلى ذلك ، تكشف النماذج النهائية المجهزة عن العوامل المحورية التي تؤثر بشكل كبير على الكتل الدائرية في سرطان الثدي ، مما يسלט الضوء على الصلابة والإنتروبيا باعتبارها المتغيرات الأكثر أهمية.

الكلمات المفتاحية: الشبكة العصبية الاصطناعية ، التحليل التمييزي الخطي ، التصوير الطبي (صورة الماموجرام) ، عائد الاستثمار ، معالجة الصور .