Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

# Distinguishing Shapes of Breast Cancer Masses in Ultrasound Images by Using Logistic Regression Model

**Luay Adil Abduljabbar** [1]
Ministry of Higher Education and Scientific Research, Studies & Planning & Follow Up Directorate,Baghdad, Iraq
luay.yahya1989@gmail.com

**Omar Qusay Alshebly**[2]
University of Mosul-College of Computer Sciences and Mathematics-Department of Statistics and Informatics-Mosul-Iraq
marqusay@uomosul.edu.iq

**Received:25/9/2022** **Accepted: 29/9/2022** **Published: September / 2022**



## Abstract

**The last few years witnessed great and increasing use in the field of medical image analysis. These tools helped the Radiologists and Doctors to consult while making a particular diagnosis. In this study, we used the relationship between statistical measurements, computer vision, and medical images, along with a logistic regression model to extract breast cancer imaging features. These features were used to tell the difference between the shape of a mass (Fibroid vs. Fatty) by looking at the regions of interest (ROI) of the mass. The final fit of the logistic regression model showed that the most important variables that clearly affect breast cancer shape images are Skewness, Kurtosis, Center of mass, and Angle, with an AUCROC of 88% and an Accuracy of almost 89%. We also came to the conclusion that the Fibroid mass is small and less white than the Fatty mass.**

**Paper type: Research paper**
**Keywords: Logistic Regression, Feature Extraction, Medical images, Accuracy, Area under Curve**

[1] Ass. lecturer, Ministry of Higher Education and Scientific Research, Baghdad, Iraq
[2] Ass.Lecturer, University of Mosul , College of Computer Sciences and Mathematics, Department of Statistics and Informatics, Mosul ,Iraq

## 1. Introduction

**Some people may believe that the embellishment and organization steps included in digital image processing are all that is necessary to make the final product seem different from the original image. However, this is far from the truth, and it is true that these steps are only seldom taken into account.**

**When a computer reads a picture, it splits the screen into groups of (pixels), and the recorded values for digital images are set up such that every pixel has its shade. Digital images are made up of thousands or millions of these tiny squares, which are known as elements of an image (Pixel)(Gonzalez and Wintz, 2004).**

**To get around the shortcomings of ordinary least squares (OLS) regression in handling dichotomous results, logistic regression was first developed in the 1970s. In the early 1980s, it was made available in statistical software. (Simpson and Yarandi, 1991).**

**The predictors variables of an outcome are routinely identified using automated variable selection techniques. Using automated variable selection techniques; the study's goal was to assess the reproducibility of logistic regression models. (Tu and Austin,2004)**

**Berry (2005) assessed the relative and absolute contribution of screening mammography and adjuvant therapy to the decline in breast cancer mortality using modeling approaches.**

**Kekre and etc. (2009) segmented mammographic pictures using Linde Buzo and Gray (LBG); for mammographic images, a codebook of size 128 was initially created. Using the same LBG technique, these code vectors were further grouped into 8 clusters. As a result, these 8 pictures were displayed.**

**Salahuddin and Yusof (2010) applied both techniques of neural network and logistic regression on image processing for understanding flower image features.**

**Image categorization is a technique developed by Ruusuvuori and etc.(2016) that relies on feeding synthetic features into a logistic regression classifier using spatial prior norm regularization and a Markov Random Field prior. The suggested technique produces a sizable number of artificial features and feeds them into a classifier for spatial prior regularized logistic regression. The classification outcome is further homogenized by a spatial prior. The effectiveness of the suggested method was examined for two application situations, and the findings demonstrated that the segmentation outcomes are reliable even for straightforward models with a high degree of sparsity.**

**Sharif and Mohsin (2018) employed image processing methods and algorithms to identify breast tumors and, in some circumstances, determine their stage so that the patient might receive the right care and live a better quality of life. Although digital mammography is frequently used to diagnose early-stage breast cancer, alternate safe procedures, such as infrared imaging, MRI, and biopsies, have been recommended due to their adverse consequences on human health.**

**One of the developed techniques that researchers in this field and in many different experts for featuring images, extracting information, and distinguishing patterns in common techniques of the digital image field consider the computer can distinguish the region of interest (ROI) in the image. Medical image analysis is extremely important, particularly in non-invasive and medical studies. Radiologists and doctors might arrive at a specific diagnosis using medical imaging techniques, their analysis, and diagnoses analysis.**

In this research, masses of (FIB & FAT) are taken, extracting digital features, statistical measurements and using binary logistic regression for reaching a model, which distinguishes the masses and, also, test the affecting independent variables of breast cancer masses.

## 2. Material and Methods

### 2.1 Processing of Images

While in image processing applications, the output images are for human consumption, in computer vision applications, the processed images are output for usage by a computer. These two groups are not entirely distinct from one another. Although the lines between the two are fuzzy, this definition enables us to examine the distinctions between the two and comprehend how they work together (Figure 1). While computer science was largely responsible for advancements in computer vision, the field of image processing historically developed from electrical engineering as an outgrowth of the signal processing branch (Murugaraja and Balamurali, 2014).

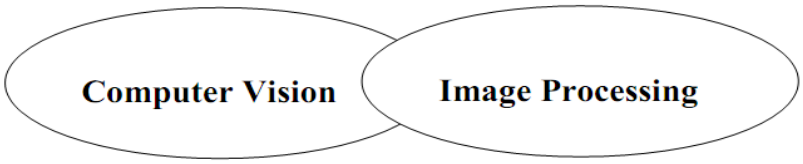


**Figure 1: Computer Imaging.**

### 2.2 Visual Computing

It is identification of the computer imaging when a person is not included in the visual process. Image analysis is a key subject in this area of computer vision (Murugaraja and Balamurali, 2014).

### 2.3 Computer Aided Diagnosis (CAD)

The skill to use a computer system for medical diagnostics is known as CAD. This diagnostic integrates a variety of approaches and tools, including big data analysis, machine learning, image processing, and databases.

The majority of CAD systems that are designed to aid in the identification of breast cancer involve mammography, ultrasound, and other image inputs. However, for photos to be used as input into a CAD system, they must first be in the proper digital format. As a result, the first function of image processing is frequently just to digitize existing mammography or MRI that is recorded in analog format. (Bankman, 2009).

### 2.4 Processing of Digital Images

A discrete two-dimensional function is a digital picture f(x,y) or a two-dimensional function f(x,y) may be used to describe a picture, and its amplitude at any given pair of coordinates (x,y) is referred to as the image's intensity or gray level at that particular location (Mohammed and Bashar, 2012) ;Figure 2 illustrates Digital Image Processing

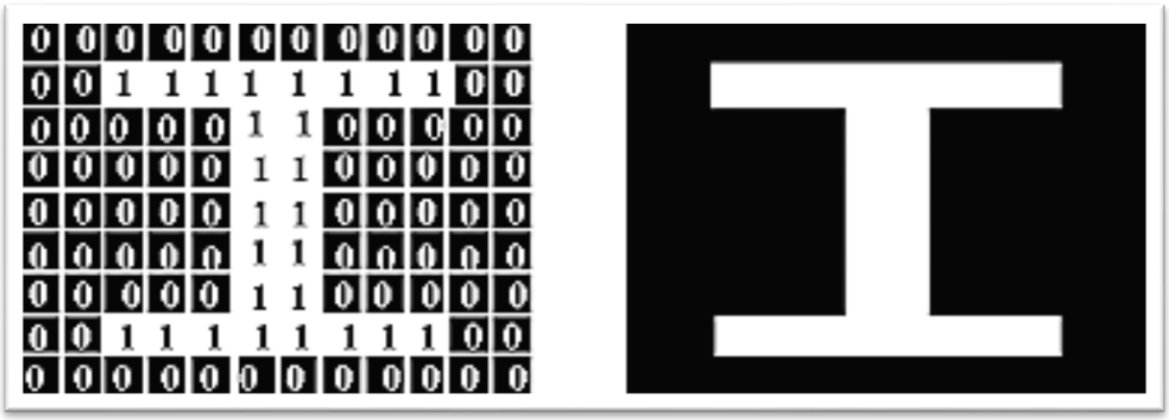

**Figure 2: Digital Image Processing**

## 2.5 Medical Imaging

**German physical scientist (Wilhelm Conrad Röntgen) was the first scientist who was able to detect X-Ray. The science of Ray was so effective first in diagnosing physicians. This science has a great development, which was used in treatments apart from diagnosing diseases.**

**Moreover, there are many other Rays (diagnoses) in which X-Ray is not used in them, but others are used such as (Ultra Sound) or (MRI); (Ultra Sound) and (MRI) are different from X-Ray, since they are harmless on the human body.**

## 2.6 Breast Cancer

**The breast, which is made up of 15 to 20 lobes made up of various lobules, is a modified sweat gland. Cooper's suspensory ligaments are fibrous bands of connective tissue that pass through the breast and enter perpendicularly into the dermis to support the structure.**

**There is more tissue in the upper outer quadrant of the breast than in the other quadrants. Individuals differ significantly in terms of the density, shape, and size of their breasts. There are 10 to 15 lactiferous ducts in each breast (Brunicardi, 2015). Figures 3 and 4 illustrate the different shapes of Fibro and Fatty masses of breast cancer.**

**Figure 3: Fibro adenomas**

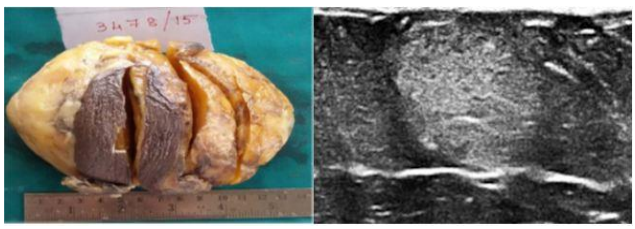

**Figure 4: Fatty (Breast lipoma)**

## 2.7 Some Statistical Measures to Describe Shape of Mass

**(Gonzales and Woods, 2008) (Ferreira and Rasband, 2012)**

### 2.7.1 Mean of Image

**Utilizing statistical moments of an image's or region's intensity histogram is one of the easiest methods for characterizing images. Let Z be a random variable representing intensity, and let p(z i), where i=0,1,2,... If L is the number of unique intensity levels, then let L-1 be the corresponding histogram.**

**The nth Z moment regarding the mean is as follows:**

$$m_n(z) = \sum_{i=0}^{L-1}(z_i - m)^n \, p(z_i) \quad (1)$$

**Where m is the mean value of Z (the average intensity):**

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i) \quad (2)$$

**The 2nd moment [the variance $\sigma^2 = m_2(z)$]**

**is especially crucial for a visual description. It is an intensity contrast metric that may be used to define relative smoothness descriptors R(z).**

**The measure**

$$R(z) = 1 - \frac{1}{1+\sigma^2(z)} \quad (3)$$

**is (0) for areas of constant intensity (the variance is zero there) and approaches (1) for large values of $\sigma^2(z)$.**

### 2.7.2 Skewness

**It is a measurement of the asymmetry in the ROI's grey scale distribution around the mean.**

$$m_3(z) = \sum_{i=0}^{L-1}(z_i - m)^3 \, p(z_i) \quad (4)$$

### 2.7.3 Kurtosis

**It is a metric for the "peakedness" of the ROI's gray value distribution around the mean. A higher kurtosis indicates that more of the variation is attributable to rare severe deviations rather than often occurring small deviations.**

$$m_4(z) = \sum_{i=1}^{L-1}(z_i - m)^4 \, p(z_i) \quad (5)$$

### 2.7.4 Entropy

**It determines how random the components of G are. The entropy equal 0 when all $p_{ij}$ are 0 and is maximum when all $p_{ij}$ are equal.**

**The largest amount is $\log_2 K$.**

$$E = -\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij} \log_2 p_{ij} \quad (6)$$

**G: Co-occurrence matrix.**

### 2.7.5 Integrated density

**The ROI's whole pixel population's gray values are added to calculate the integrated density.**

### 2.7.6 Median

**The grey value that falls in the centre or the mean of the two values that fall in the middle when all grey values are ordered by their numerical value is the median grey value of a pixel within the ROI.**

### 2.7.7 Centroid

**The selection's spatial centre x and y coordinates of every pixel in the picture or selection are averaged out to get this value utilizes the headers X and Y.**

**2.7.8 Mass Area**

**The number of pixels in the area**

$$a = \pi r^2 \; and \; \pi = \frac{22}{7} \qquad (7)$$

**2.7.9 The Perimeter**

**It is the measurement of the selection's outside border.**

$$p = 2\pi r \qquad (8)$$

**2.7.10 Width and Height**

**They are a certain mass's dimension and length.**

**2.7.11 Minimum and Maximum Grey Level**

**Grey scale values that are minimum and maximum for the selection.**

**2.7.12 Major and Minor**

**They are the best-fitting ellipse's primary and secondary axes.**

**2.7.13 Circularity**

**It represents circularity ratio given by the expression**

$$R_c = \frac{4\,\pi\, area}{p^2} \qquad (9)$$

**2.7.14 The Solidity**

**The following formula can be used to determine a block's specific convex shape's screen point ratio:**

$$s = \frac{[\text{Area}]}{[\text{Convex area}]} \qquad (10)$$

**2.7.15 Diameter Angle**

**The Ferret Angle indicator shows the diameter's angle, which ranges from 0 to 180 degrees.**

## 2.8 Logistic Regression

**The influence of several indicators that are provided at once to predict membership in either one of the two answer groups is assessed using logistic regression.**

**Logistic regression cannot predict a numerical value for the dependent variable since it is a dichotomous variable, therefore the standard regression least squares deviations criterion for the best fit technique of reducing error around the line of best fit is an incorrect choice for logistic regression instead. Use the binomial probability theory, where the only possible outcomes are that probability (p) is (1 rather than 0). The maximum likelihood technique, used in logistic regression, creates the best-fitting equation or function by maximizing the likelihood that the observed data will be classified into the correct category given the regression coefficients.**

**When the logistic distribution is utilized, the conditional mean of Y given X is denoted by the quantity $\pi(x) = E(Y/x)$ to simplify notation. Our particular version of the logistic regression model is as follows:**

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \qquad (11)$$

$$1 - \pi(x) = \frac{1}{1 + e^{\beta_0+\beta_1 x}}$$

$$\text{or} \qquad \frac{\pi(x)}{1-\pi(x)} = e^{\beta_0+\beta_1 x}$$

$$g(x) = log\left[\frac{\pi(x)}{1-\pi(x)}\right]$$

$$= \beta_0 + \beta_1 x.$$

**This modification is significant because g(x) has several advantageous characteristics of a linear regression model. Depending on the range of x, the logistic g(x), may be continuous and have parameters that vary from -∞ to +∞. The conditional distribution of the outcome variable is the second significant distinction between the logistic and linear regression models. We presume that an observation of the outcome variable may be written as y=E(Y/x)+$\varepsilon$ in the linear regression model. The error is a quantity that measures an observation's deviation from the conditional mean. (David and Hosmer, 2013).**

## 3. Results and Discussion

### 3.1 Real Dataset Collection

**Two-dimensional image data sources are used in the field of digital image processing; in this study, 100 ultrasound digital images of disease tumors were collected from Baghdad Medical City -Oncology and Nuclear Medicine Hospital ,including 56 images of fibroids and 46 images of fatty tissues. All the implementations of the study on real data applications are carried out using R version (4.4.1).**

**Note that the images were all laboratory test procedures. (Pathology) was diagnosed type of tumor by a specialist doctor.**

**The Variables are:**

**1. Response Variable: contains the (FIB & FAT) shape, which is coded as follows:**
**Y=0 (FIB)**
**Y=1 (FAT)**
**2. Predictors Variables: consist of the following elements, all drawn from ROI:**
**(Area ,Mean ,StdDev. ,Mode ,Min ,Max ,Width ,height ,Perim. ,Major ,Minor ,Median ,Round ,Skew. ,Kurt. ,Angle ,Circularity )**
**Figures 5 and 6 illustrate Ultrasound Images for Fatty and Fibroid masses of breast cancer.**

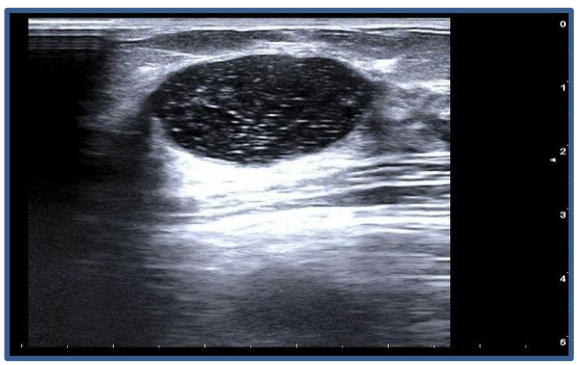

**Figure 5: Fatty Mass**

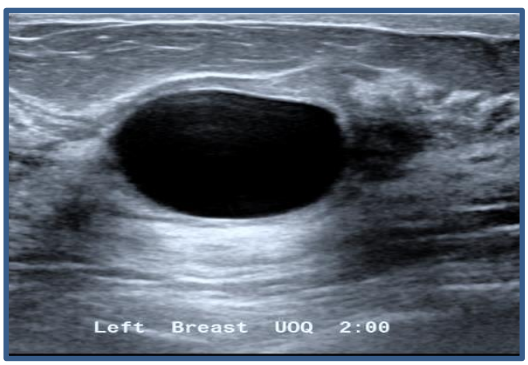


**Figure 6: Fibroid Mass**

## 3.2 Applying Logistic Regression of Ultrasound Image

**The model must now be examined to determine how well it generalizes to additional observations and matches the data. To create the most effective model, the evaluation process comprises evaluating three different areas: the quality of fit, testing of individual predictors, and validation of predicted values.**

**As seen in Table 1, the coefficients, when negative and positive values show up in the results of the logistic regression coefficients, this means that the changes in the model were both expulsion and inverse. Their standard errors, the z-statistic, and the corresponding p-values are displayed in the next section of the summary findings. This specifies the model equation, making it the most important component of the model. As can be seen from the p-value of 0.05, not all coefficients are significant. The logistic regression coefficients (Center x, Center y, Skewness, Kurtosis, Angle) show the change in the log chances of the result for a one unit increase in the predictor variables.**

**Table (1): Logistic Regression Summary**

| variables | Coefficients Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -2.96E+01 | 2.48E+01 | -1.194 | 0.23259 |
| Area | -4.80E-04 | 3.37E-04 | -1.424 | 0.15456 |
| Mean | 2.37E-01 | 2.51E-01 | 0.946 | 0.34431 |
| StdDev | 2.40E-03 | 1.17E-01 | 0.021 | 0.98359 |
| Mode | -1.84E-03 | 2.71E-02 | -0.068 | 0.94589 |
| Min | -1.92E-02 | 5.25E-02 | -0.367 | 0.71379 |
| Max | -2.38E-02 | 2.14E-02 | -1.109 | 0.26734 |
| Centr.X | 2.96E-02 | 1.20E-02 | 2.478 | 0.01323 |
| Centr.Y | -4.81E-02 | 2.20E-02 | -2.186 | 0.02879 |
| Width | 9.82E-02 | 9.04E-02 | 1.086 | 0.27735 |
| Height | 9.66E-02 | 8.98E-02 | 1.076 | 0.28179 |
| Perim. | 8.10E-02 | 4.14E-02 | 1.958 | 0.05029 |
| Major | -2.48E-01 | 1.33E-01 | -1.865 | 0.06223 |
| Minor | -1.28E-01 | 9.49E-02 | -1.352 | 0.1763 |
| Median | -1.36E-01 | 2.21E-01 | -0.617 | 0.53742 |
| Skew | -6.97E+00 | 2.36E+00 | -2.949 | 0.00319 |
| Kurt | 8.52E-01 | 2.64E-01 | 3.229 | 0.00124 |
| Angle | -1.44E-02 | 6.97E-03 | -2.064 | 0.03901 |
| Circ. | 2.14E+01 | 1.50E+01 | 1.433 | 0.15174 |
| Round | -1.07E+01 | 1.05E+01 | -1.02 | 0.30758 |
| Solidity | 2.11E+01 | 2.63E+01 | 0.801 | 0.42331 |

**"Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1"**

**Table 2 shows the classification table,**

**Table (2): Classification Table**

| Classification | | Observation | |
|---|---|---|---|
| | | FIB | FAT |
| Prediction | FIB | 49 | 5 |
| | FAT | 6 | 40 |
| Model Accuracy | | 89% | |
| Model Sensitivity | | 88% | |
| Model Specificity | | 89% | |
| Error Rate of Classification | | 11% | |

**Table (2) displays the percentages of correctly classified (FIB) and (FAT) items as well as the overall %. Based on the table's findings, we can see that the percentages of correctly classified (FIB) items were greater than (FAT). As it is possible to verify, the model has accuracy (89%), but also the sensitivity and the specificity are greater than 80 percent (88%, and 89%). Classification errors were therefore only 11%.**

**The Receiver Operator Characteristic is an additional tool for evaluating the performance of the model (ROC). A classification model's accuracy at a client threshold value is determined via ROC. The area under the curve is used to assess the model's accuracy (AUC).**

**0.8885 is the area under the curve as shown in the figure 6.**

**Call:**

**roc.default(response = datacancer$class, predictor = LGModelPred)**

**Data: LGModelPred in 54 controls (datacancer$class 0) < 46 cases (datazr$class 1).**
**Area under the curve: 0.8885**

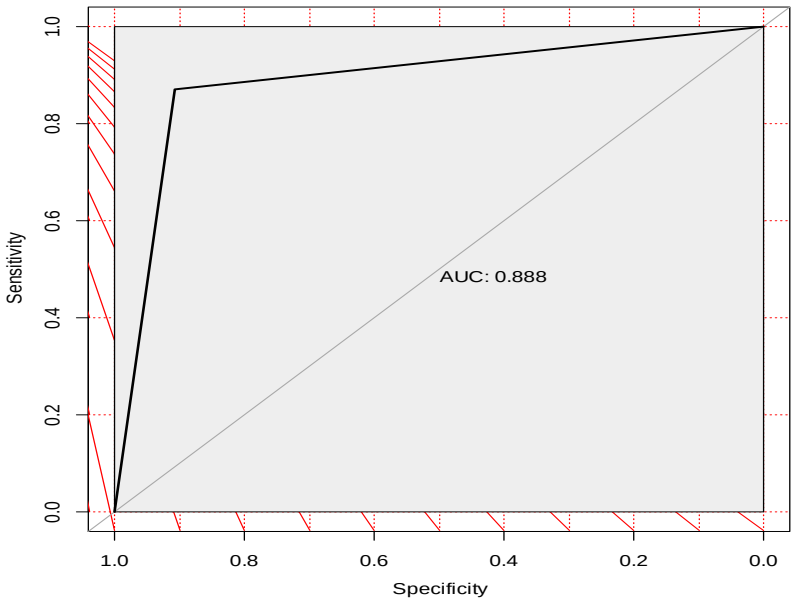


**Figure (7: Receiver Operating Characteristics (ROC)**

**ROC is plotted between the sensitivity (y axis) and the specificity (x axis). Figure 6 shows area under curve value is 89% and value of cutoff-point is 0.5.**

## 3.3 Fitted Final Model and Variables Importance

**The final model was adjusted using logistic regression with significant independent variables affecting of breast cancer masses patients**

$$logodds = \beta 0 + \beta 7X7 + \beta 8X8 + \beta 15X15 + \beta 16X16 + \beta 17X17$$
$$logodds = -2.960 + 2.962x7 - 4.811x8 - 6.966x15 + 8.524x16 - 1.438x17$$

**whereas:-**

**$B_0$ : Intercept. , X7 : Center x_axis ,X8 : Center y_axis ,X15 : Skewness X16 : Kurtosis ,X17 : Angle measurement.**

**On the other hand, for general & generalized linear models, they were using the varImp function in the caret package.**

**Table 3 shows the degree of importance for each independent variables, where the variable of kurtosis is the largest effect followed by variable skewness and then center x_axis etc.**

**Table (3): Variables Importance**

| variables | Importance |
|---|---|
| Kurt | 100 |
| Skew | 91.1 |
| CentrX | 66.4 |
| CentrY | 63.45 |
| Angle | 63.48 |
| Perim | 54.2 |
| Majo | 53.2 |
| Circulaity | 50.2 |
| Area | 49.33 |
| Minor | 45.24 |
| Max4 | 5.3 |
| Width | 40.4 |
| Height | 35.3 |
| Round | 33.4 |
| Mean | 24.5 |
| Solidity | 22.5 |
| Median1 | 0.2 |
| Min | 4.3 |
| Mode | 3.1 |
| StdDev | 2.12 |

### 3.4 Using Testing hypothesis to compare significant parameter

**The Independent-Samples T- Test procedure tests the significant differences between two sample means, of (FIB, FAT) mass.**
**We test two hypotheses:**

$H_0 : \mu_{FB} = \mu_{FT}$

$H_1 : \mu_{FB} \neq \mu_{FT}$

**The above table indicates the differences between the two variables (Kurtosis and Skewness) are significant differences because the significance value of the statistic is (0.000, 0.000) respectively and this value is less than 0.05.**
**We concluded that the mass of Fibroid has a small size and also less white than the mass of Fatty.**

## 4. Conclusion

**The most significant findings drawn by the study are as follows, according to the results of the application part:**
**1- Proper implementation of the logistic model in the comparison and quality analysis of digital medical picture masses (FIB & FAT), depending on a set of statistical and geometric measurements to achieve a high degree of classification, which was (89%).**
**2-The distinction between (FIB & FAT) image masses relies on statistical measures, which we discovered to be efficient in the process of distinction by the estimated model. This is one of the study's key results to differentiate between (FIB & FAT) masses ك the model's significance variables are (Center x, Center y, Skewness, Kurtosis, and Angle).**
**3-Where we test hypotheses (t-test) for significance parameter in logistic model (FIB & FAT) images of masses, we concluded that the (FIB) masses are (blacker and have more skewness and kurtosis) of the mass (FAT) (whiter and less skewness and kurtosis larger size) و and these results match with the medical diagnosis part**

## References

**1.Al-Ghamdi, SA. (2002).Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis & Prevention, Vol.34, Issue 6, pp. 729-741.**

**2.Austin, PC. and Tu, JV. (2004) .Automated Variable Selection Methods for Logistic Regression Produced Unstable Models for predicting acute myocardial infarction mortality. Journal of Clinical Epidemiology, Vol. 57, pp. 1138–1146.**

**3.Bankman, I.N. (2009) .Handbook of Medical Image Processing and Analysis. Second Edition, Burlington, MA, USA: Elsevier.**

**4.Berry, DA. (2005) .Effect of screening and adjuvant therapy on mortality from breast cancer. Journal of medicine, Vol. 353. Issue 17, PP. 1784- 17 92.**

**5.DAVID, W. and HOSMER, JR. (2013) .Applied Logistic Regression. Third Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, Canada.**

**6.Gonzalez, R.C. & Wintz, P. (1992) "Digital Image Processing". Third Edition, Addision wesly**

**7.Kekre, H., Sarode, K. and Gharge, M. (2009) .Detection and Demarcation of Tumor using Vector Quantization in MRI images. International Journal of Engineering Science and Technology, Vol.1, No. 2, PP. 59-66.**

**8.Mohammed, N., Jebunnahar, D. and Abul Bashar, MD. (2012) .A Comprehensive Study of Digital Image Processing for finding image quality dependencies. Journal of Scientific and Research Publications, Vol. 2, ISSN. 2250-3153, pp. 1-8.**

**9.Murugaraja, S. and Balamurali, K. (2014) .An Efficient Face Tracker Using Active Shape Model. Journal of Computer Engineering, Vol. 16, Issue. 2, PP 91-96.**

**10. Ruusuvuori, P., Manninen, T. and Huttunen, H. (2016) .Image segmentation using sparse logistic regression with spatial prior. IEEE, PP. 2253 – 2257.**

**11. Siraj, F., Salahuddin, M. and Mohd Yusof, Sh. (2010) .Digital Image Classification for Malaysan Blooming Flower. IEEE, PP. 33 – 38.**

**12. Yarandi, H. and Simpson, S. (1991) the logistic regression model and the odds of testing HIV Positive. Nursing Research, Vol. 40, Issue 6, pp. 372-373.**

**13. Yasmine, M., Sharif, M. and Mohsin, S. (2018) .Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques. Journal of Recent Sciences, Vol. 2, Issue 10, PP. 88-98.**

## Appendix

**Part of the Matrix Data used in this study was obtained via the Image J Program after determini ng ROI.**

| y | Area | Mean | StdDev | Mode | Min | Max | Centr X | Centr Y | Width | Height | Perim. | Major | Minor | Median | Skew | Kurt | Angle | Circ. | Round | Solidity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5129 | 57.09 | 18.461 | 49 | 18 | 153 | 117.264 | 62.177 | 110 | 66 | 309.783 | 113.734 | 57.419 | 55 | 0.931 | 1.482 | 17.484 | 0.672 | 0.505 | 0.959 |
| 1 | 11051 | 46.125 | 25.406 | 37 | 0 | 164 | 153.3 | 92.764 | 153 | 88 | 417.81 | 155.478 | 90.499 | 43 | 0.883 | 1.135 | 174.67 | 0.796 | 0.582 | 0.972 |
| 1 | 5411 | 33.681 | 17.521 | 25 | 0 | 146 | 140.228 | 84.658 | 132 | 54 | 327.203 | 138.329 | 49.805 | 29 | 2.536 | 9.232 | 9.96 | 0.635 | 0.36 | 0.966 |
| 1 | 13606 | 39.678 | 26.365 | 16 | 0 | 233 | 241.393 | 115.318 | 145 | 118 | 465.221 | 154.914 | 111.828 | 36 | 1.175 | 3.323 | 159 | 0.79 | 0.722 | 0.964 |
| 1 | 20891 | 25.247 | 18.823 | 18 | 0 | 255 | 203.93 | 313.731 | 244 | 114 | 626.773 | 238.616 | 111.473 | 22 | 5.431 | 57.6 | 176.19 | 0.668 | 0.467 | 0.967 |
| 1 | 79965 | 30.449 | 19.752 | 17 | 0 | 135 | 575.969 | 187.246 | 517 | 195 | 1253.074 | 536.357 | 189.826 | 27 | 0.885 | 0.703 | 178.691 | 0.64 | 0.354 | 0.959 |
| 1 | 14361 | 42.026 | 28.718 | 30 | 0 | 208 | 196.136 | 144.545 | 165 | 117 | 473.318 | 168.773 | 108.341 | 35 | 1.553 | 3.152 | 21.656 | 0.806 | 0.642 | 0.962 |
| 1 | 34168 | 29.329 | 18.856 | 26 | 0 | 141 | 314.196 | 179.464 | 221 | 191 | 674.636 | 223.189 | 194.92 | 26 | 1.365 | 3.12 | 174.108 | 0.943 | 0.873 | 0.987 |
| 1 | 7485 | 33.828 | 17.701 | 29 | 0 | 136 | 154.035 | 65.779 | 115 | 82 | 345.819 | 115.75 | 82.335 | 32 | 0.995 | 1.701 | 167.529 | 0.787 | 0.711 | 0.964 |
| 1 | 19409 | 35.434 | 22.405 | 1 | 0 | 140 | 366.567 | 239.247 | 228 | 105 | 564.475 | 231.031 | 106.965 | 33 | 0.511 | 0.002 | 5.234 | 0.765 | 0.463 | 0.984 |
| 1 | 12803 | 18.691 | 18.98 | 0 | 0 | 136 | 348.223 | 230.639 | 142 | 122 | 435.911 | 140.767 | 115.804 | 14 | 1.889 | 4.881 | 6.725 | 0.847 | 0.823 | 0.953 |
| 1 | 5825 | 60.413 | 22.041 | 45 | 30 | 181 | 157.649 | 141.399 | 106 | 72 | 305.362 | 105.751 | 70.133 | 55 | 1.856 | 4.569 | 7.077 | 0.785 | 0.663 | 0.971 |
| 1 | 967 | 39.786 | 23.482 | 22 | 0 | 128 | 98.181 | 32.172 | 45 | 28 | 131.957 | 46.607 | 26.417 | 33 | 1.114 | 0.972 | 171.853 | 0.698 | 0.567 | 0.947 |
| 1 | 2780 | 29.182 | 23.524 | 17 | 1 | 152 | 155 | 85.262 | 72 | 51 | 216.746 | 76.847 | 46.06 | 21 | 1.768 | 3.193 | 22.527 | 0.744 | 0.599 | 0.96 |
| 1 | 3697 | 20.803 | 24.692 | 0 | 0 | 242 | 46.509 | 95.028 | 69 | 61 | 232.8 | 74.305 | 63.349 | 14 | 3.604 | 19.347 | 9.505 | 0.857 | 0.853 | 0.984 |
| 1 | 2778 | 52.369 | 26.367 | 38 | 0 | 165 | 121.663 | 55.491 | 71 | 48 | 209.329 | 71.944 | 49.164 | 44 | 1.338 | 1.56 | 167.225 | 0.797 | 0.683 | 0.967 |
| 1 | 9136 | 37.823 | 22.069 | 27 | 0 | 173 | 134.571 | 84.011 | 125 | 98 | 362.904 | 125.453 | 92.723 | 34 | 1.532 | 3.825 | 170.494 | 0.872 | 0.739 | 0.971 |
| 1 | 10709 | 68.151 | 32.684 | 47 | 0 | 251 | 158.636 | 68.102 | 125 | 108 | 391.699 | 124.129 | 109.846 | 62 | 1.379 | 2.931 | 1.961 | 0.877 | 0.885 | 0.971 |
| 1 | 5425 | 51.004 | 25.197 | 42 | 0 | 212 | 121.665 | 66.017 | 109 | 64 | 317.312 | 112.843 | 61.212 | 47 | 1.619 | 4.528 | 7.392 | 0.677 | 0.542 | 0.963 |
| 1 | 3111 | 67.951 | 35.151 | 45 | 4 | 236 | 131.828 | 74.939 | 74 | 60 | 222.055 | 71.041 | 55.757 | 61 | 1.212 | 1.759 | 8.783 | 0.793 | 0.785 | 0.951 |
| 1 | 46898 | 47.42 | 16.127 | 44 | 9 | 108 | 322.467 | 185.852 | 387 | 149 | 917.691 | 388.604 | 153.659 | 46 | 0.416 | -0.112 | 1.089 | 0.7 | 0.395 | 0.981 |
| 1 | 46791 | 56.744 | 26.155 | 40 | 0 | 166 | 319.542 | 187.471 | 334 | 175 | 870.137 | 343.587 | 173.395 | 53 | 0.609 | 0.034 | 5.521 | 0.777 | 0.505 | 0.976 |
| 1 | 29832 | 38.689 | 15.864 | 26 | 3 | 129 | 266.953 | 205.758 | 230 | 164 | 661.157 | 236.347 | 160.71 | 36 | 0.907 | 0.916 | 8.384 | 0.858 | 0.68 | 0.975 |
| 1 | 9075 | 45.801 | 13.398 | 34 | 25 | 111 | 155.353 | 199.973 | 140 | 83 | 372.259 | 139.208 | 83.003 | 43 | 1.213 | 1.425 | 176.83 | 0.823 | 0.596 | 0.972 |
| 1 | 15210 | 41.051 | 13.329 | 40 | 8 | 160 | 226.871 | 120.658 | 183 | 128 | 516.976 | 187.514 | 103.277 | 39 | 1.43 | 4.814 | 25.289 | 0.715 | 0.551 | 0.909 |
| 1 | 54704 | 52.627 | 24.051 | 35 | 2 | 182 | 596.699 | 339.29 | 319 | 277 | 977.877 | 338.413 | 205.817 | 49 | 0.967 | 1.196 | 146.14 | 0.719 | 0.608 | 0.919 |
| 1 | 21379 | 45.923 | 24.491 | 34 | 0 | 177 | 139.169 | 184.448 | 197 | 150 | 568.585 | 193.723 | 140.513 | 42 | 0.968 | 1.443 | 170.937 | 0.831 | 0.725 | 0.956 |
| 1 | 34047 | 28.421 | 15.75 | 28 | 0 | 141 | 591.191 | 253.261 | 294 | 132 | 771.906 | 311.999 | 138.943 | 26 | 1.642 | 4.697 | 3.045 | 0.718 | 0.445 | 0.984 |
| 1 | 2257 | 73.543 | 35.475 | 48 | 11 | 193 | 148.038 | 128.599 | 61 | 52 | 193.933 | 65.383 | 43.952 | 65 | 0.84 | 0.345 | 152.262 | 0.754 | 0.672 | 0.937 |
| 1 | 31450 | 100.941 | 38.143 | 82 | 2 | 255 | 213.384 | 125.534 | 370 | 116 | 872.85 | 372.933 | 107.374 | 99 | 0.176 | -0.548 | 175.68 | 0.519 | 0.288 | 0.948 |
| 1 | 18488 | 19.969 | 18.142 | 4 | 0 | 136 | 176.453 | 152.235 | 211 | 131 | 565.936 | 205.324 | 114.646 | 14 | 1.762 | 3.797 | 6.073 | 0.725 | 0.558 | 0.923 |
| 1 | 6385 | 43.922 | 24.006 | 0 | 0 | 189 | 168.29 | 93.183 | 98 | 88 | 327.674 | 100.188 | 81.144 | 41 | 1.134 | 3.632 | 170.478 | 0.747 | 0.81 | 0.954 |
| 1 | 36319 | 85.453 | 37.045 | 33 | 29 | 255 | 227.177 | 143.081 | 281 | 190 | 760.728 | 260.442 | 177.555 | 87 | 0.184 | -0.827 | 172.584 | 0.789 | 0.682 | 0.961 |
| 1 | 52948 | 82.922 | 26.255 | 79 | 34 | 254 | 179.47 | 172.889 | 312 | 212 | 912.719 | 324.104 | 208.006 | 79 | 0.994 | 1.501 | 7.995 | 0.799 | 0.642 | 0.979 |
| 1 | 12480 | 78.464 | 32.727 | 71 | 24 | 251 | 130.052 | 91.892 | 141 | 114 | 419.186 | 139.342 | 114.036 | 74 | 1.415 | 3.942 | 5.368 | 0.893 | 0.818 | 0.977 |
| 1 | 9991 | 80.622 | 33.656 | 81 | 9 | 240 | 135.349 | 53.833 | 228 | 63 | 522.914 | 238.593 | 53.317 | 75 | 1.052 | 1.373 | 173.251 | 0.459 | 0.223 | 0.927 |
| 1 | 5496 | 79.099 | 17.275 | 70 | 52 | 180 | 79.216 | 49.998 | 94 | 74 | 290.438 | 97.958 | 71.436 | 75 | 2.049 | 7.249 | 16.775 | 0.819 | 0.729 | 0.969 |
| 1 | 14695 | 61.384 | 26.852 | 42 | 31 | 255 | 115.187 | 115.989 | 116 | 204 | 582.4 | 204.3 | 91.582 | 57 | 4.201 | 26.332 | 86.944 | 0.544 | 0.448 | 0.808 |
| 1 | 51979 | 45.553 | 28.115 | 0 | 0 | 165 | 251.633 | 201.861 | 495 | 187 | 1301.909 | 475.539 | 139.172 | 43 | 0.608 | 0.244 | 170.172 | 0.385 | 0.293 | 0.772 |
| 1 | 11822 | 40.775 | 21.564 | 0 | 0 | 129 | 59.647 | 112.592 | 120 | 126 | 421.978 | 126.122 | 119.347 | 38 | 0.573 | 0.379 | 40.57 | 0.834 | 0.946 | 0.97 |
| 1 | 39608 | 30.577 | 15 | 29 | 0 | 109 | 322.382 | 253.602 | 242 | 218 | 758.391 | 237.218 | 212.591 | 29 | 0.865 | 1.443 | 16.129 | 0.865 | 0.896 | 0.971 |
| 1 | 36743 | 90.712 | 32.402 | 75 | 0 | 245 | 139.313 | 105.338 | 258 | 185 | 732.871 | 255.538 | 183.075 | 87 | 0.554 | 0.577 | 172.393 | 0.86 | 0.716 | 0.979 |
| 1 | 5339 | 47.176 | 26.836 | 0 | 0 | 146 | 134.715 | 74.434 | 92 | 80 | 284.645 | 85.622 | 79.394 | 43 | 0.661 | 0.016 | 177.343 | 0.828 | 0.927 | 0.946 |
| 1 | 4570 | 69.247 | 48.258 | 0 | 0 | 234 | 71.649 | 67.387 | 115 | 55 | 312.049 | 116.926 | 49.764 | 61 | 0.505 | -0.428 | 8.019 | 0.59 | 0.426 | 0.946 |
| 1 | 19325 | 37.451 | 22.375 | 0 | 0 | 179 | 185.401 | 101.159 | 214 | 125 | 559.362 | 208.11 | 118.233 | 36 | 0.779 | 2.022 | 176.377 | 0.776 | 0.568 | 0.955 |
| 1 | 19342 | 19.624 | 20.443 | 0 | 0 | 134 | 152.828 | 87.377 | 187 | 133 | 542.589 | 187.675 | 131.221 | 15 | 1.216 | 1.549 | 8.301 | 0.826 | 0.699 | 0.964 |
| 1 | 13398 | 23.251 | 19.993 | 0 | 0 | 229 | 142.069 | 99.527 | 180 | 114 | 499.065 | 162.755 | 104.813 | 19 | 1.635 | 5.319 | 19.417 | 0.676 | 0.644 | 0.897 |
| 1 | 6822 | 43.359 | 31.913 | 0 | 0 | 204 | 163.344 | 99.844 | 90 | 111 | 357.8 | 109.763 | 79.134 | 36 | 1.839 | 4.238 | 99.621 | 0.67 | 0.721 | 0.862 |
| 1 | 6166 | 26.719 | 21.256 | 0 | 0 | 111 | 175.875 | 79.971 | 112 | 80 | 333.85 | 115.797 | 67.798 | 25 | 0.852 | 0.663 | 16.245 | 0.695 | 0.585 | 0.916 |
| 2 | 29123 | 110.224 | 22.753 | 113 | 42 | 179 | 186.169 | 123.569 | 253 | 147 | 736.307 | 267.185 | 138.782 | 111 | -0.22 | -0.211 | 175.065 | 0.675 | 0.519 | 0.933 |
| 2 | 4286 | 157.72 | 43.885 | 171 | 0 | 255 | 177.028 | 84.103 | 81 | 74 | 263.638 | 80.611 | 67.697 | 169 | -1.872 | 4.251 | 28.514 | 0.775 | 0.84 | 0.929 |
| 2 | 11311 | 123.2 | 39.62 | 118 | 0 | 254 | 211.919 | 90.929 | 131 | 109 | 411.334 | 134.032 | 107.449 | 124 | -0.028 | -0.403 | 163.396 | 0.84 | 0.802 | 0.971 |
| 2 | 19694 | 83.544 | 28.452 | 92 | 0 | 189 | 209.757 | 91.007 | 194 | 128 | 555.583 | 197.577 | 126.914 | 84 | -0.15 | 0.38 | 6.241 | 0.802 | 0.642 | 0.967 |
| 2 | 6530 | 97.938 | 18.912 | 91 | 44 | 162 | 111.987 | 117.238 | 98 | 81 | 300.915 | 103.8 | 80.099 | 98 | 0.093 | -0.33 | 25.838 | 0.906 | 0.772 | 0.981 |
| 2 | 9725 | 89.09 | 35.997 | 75 | 12 | 227 | 241.051 | 71.892 | 141 | 85 | 391.774 | 142.826 | 86.695 | 87 | 0.469 | 0.003 | 171.193 | 0.796 | 0.607 | 0.978 |
| 2 | 20609 | 65.396 | 22.685 | 63 | 0 | 190 | 189.183 | 133.878 | 220 | 156 | 629.097 | 240.425 | 109.141 | 63 | 0.452 | 1.336 | 28.052 | 0.654 | 0.454 | 0.925 |
| 2 | 2376 | 103.399 | 35.308 | 104 | 9 | 210 | 127.957 | 114.261 | 61 | 51 | 191.93 | 59.907 | 50.499 | 105 | -0.14 | -0.187 | 15.069 | 0.811 | 0.843 | 0.956 |
| 2 | 2960 | 62.893 | 31.183 | 59 | 0 | 255 | 166.532 | 71.81 | 75 | 50 | 207.512 | 74.887 | 50.327 | 60 | 1.048 | 2.941 | 173.13 | 0.864 | 0.672 | 0.968 |
| 2 | 2195 | 89.536 | 21.369 | 96 | 9 | 225 | 127.991 | 56.312 | 59 | 48 | 174.28 | 59.35 | 47.09 | 92 | -0.438 | 1.893 | 6.761 | 0.908 | 0.793 | 0.963 |
| 2 | 17804 | 99.255 | 37.477 | 99 | 1 | 255 | 82.118 | 124.379 | 141 | 161 | 514.632 | 166.738 | 135.955 | 97 | 0.301 | 0.537 | 68.481 | 0.845 | 0.815 | 0.976 |
| 2 | 3852 | 92.86 | 34.442 | 91 | 13 | 230 | 153.908 | 62.135 | 98 | 47 | 255.676 | 103.362 | 47.45 | 92 | 0.469 | 0.431 | 9.911 | 0.74 | 0.459 | 0.977 |
| 2 | 10654 | 99.805 | 19.55 | 96 | 0 | 255 | 126.916 | 63.475 | 161 | 90 | 432.035 | 161.735 | 83.872 | 99 | 0.621 | 2.139 | 179.806 | 0.717 | 0.519 | 0.952 |
| 2 | 27099 | 78.695 | 20.649 | 81 | 7 | 192 | 234.475 | 169.102 | 212 | 157 | 635.628 | 218.506 | 157.906 | 79 | 0.009 | 1.342 | 4.338 | 0.843 | 0.723 | 0.976 |
| 2 | 23195 | 103.889 | 43.805 | 78 | 10 | 246 | 434.352 | 177.3 | 224 | 131 | 636.566 | 239.569 | 123.275 | 98 | 0.628 | -0.074 | 179.843 | 0.719 | 0.515 | 0.908 |
| 2 | 11219 | 50.952 | 23.425 | 52 | 0 | 130 | 275.061 | 184.294 | 156 | 97 | 440.797 | 151.572 | 94.242 | 50 | 0.191 | -0.631 | 168.422 | 0.726 | 0.622 | 0.957 |
| 2 | 26269 | 57.051 | 18.451 | 64 | 12 | 113 | 318.521 | 206.121 | 277 | 129 | 699.647 | 263.857 | 126.761 | 57 | 0.095 | -0.507 | 2.035 | 0.674 | 0.48 | 0.938 |
| 2 | 9286 | 106.38 | 29.582 | 102 | 0 | 224 | 105.398 | 87.469 | 134 | 95 | 390.505 | 131.485 | 89.921 | 104 | -0.907 | 3.089 | 2.036 | 0.765 | 0.684 | 0.935 |
| 2 | 10180 | 64.109 | 34.09 | 33 | 10 | 255 | 120.31 | 54.424 | 148 | 86 | 472.453 | 148.645 | 87.198 | 59 | 0.924 | 1.305 | 4.929 | 0.573 | 0.587 | 0.97 |
| 2 | 5415 | 58.981 | 33.809 | 0 | 0 | 190 | 436.734 | 172.982 | 109 | 63 | 299.556 | 112.542 | 61.262 | 54 | 0.59 | -0.012 | 177.191 | 0.758 | 0.544 | 0.94 |
| 2 | 3458 | 97.569 | 32.506 | 60 | 34 | 193 | 55.172 | 75.662 | 73 | 62 | 232.106 | 71.358 | 61.701 | 93 | 0.427 | -0.651 | 155.136 | 0.807 | 0.865 | 0.95 |
| 2 | 8199 | 117.274 | 34.292 | 98 | 9 | 213 | 418.586 | 117.366 | 117 | 90 | 359.005 | 116.267 | 89.788 | 116 | 0.121 | -0.339 | 1.612 | 0.799 | 0.772 | 0.949 |
| 2 | 6222 | 66.625 | 25.981 | 53 | 0 | 170 | 184.271 | 46.908 | 98 | 82 | 305.018 | 97.509 | 81.245 | 60 | 0.826 | 1.532 | 171.922 | 0.84 | 0.833 | 0.953 |
| 2 | 17089 | 65.507 | 22.103 | 65 | 0 | 189 | 310.625 | 110.908 | 197 | 113 | 535.667 | 203.749 | 106.79 | 64 | 0.695 | 2.487 | 170.997 | 0.748 | 0.524 | 0.965 |
| 2 | 5702 | 73.447 | 33.552 | 0 | 0 | 207 | 138.427 | 48.997 | 118 | 90 | 352.348 | 121.235 | 59.884 | 72 | 0.158 | 0.06 | 41.608 | 0.577 | 0.494 | 0.866 |
| 2 | 3120 | 54.597 | 37.421 | 0 | 0 | 166 | 155.676 | 61.995 | 56 | 67 | 233.857 | 68.832 | 57.713 | 42 | 0.82 | -0.228 | 93.43 | 0.717 | 0.838 | 0.954 |
| 2 | 10639 | 46.331 | 32.288 | 17 | 4 | 195 | 155.157 | 72.101 | 135 | 104 | 425.221 | 133.225 | 101.678 | 36 | 1.18 | 1.273 | 23.045 | 0.739 | 0.763 | 0.949 |
| 2 | 24096 | 22.864 | 19.796 | 3 | 0 | 155 | 190.815 | 123.851 | 240 | 148 | 678.905 | 231.484 | 132.536 | 18 | 1.787 | 4.717 | 10.57 | 0.657 | 0.573 | 0.937 |
| 2 | 11181 | 92.424 | 36.614 | 0 | 0 | 255 | 157.525 | 110.061 | 169 | 94 | 473.731 | 164.193 | 86.704 | 87 | 1.214 | 3.685 | 6.311 | 0.626 | 0.528 | 0.886 |
| 2 | 12434 | 48.979 | 22.091 | 40 | 0 | 185 | 169.66 | 108.951 | 171 | 97 | 473.055 | 167.439 | 94.551 | 47 | 0.685 | 1.617 | 2.875 | 0.698 | 0.565 | 0.96 |

# تمييز أشكال كتل سرطان الثدي في صور الموجات فوق الصوتية باستخدام نموذج الانحدار اللوجستي

م.م. لؤي عادل عبدالجبار[1]
وزارة التعليم العالي والبحث العلمي
دائرة الدراسات والتخطيط والمتابعة ،
بغداد ، العراق
luay.yahya1989@gmail.com

م.م. عمر قصي الشبلي[2]
جامعة الموصل/كلية علوم
الحاسوب والرياضيات/ قسم الاحصاء
والمعلوماتية ، الموصل ، العراق
omarqusay@uomosul.edu.iq

**Received:25/9/2022** **Accepted: 29/9/2022** **Published: September / 2022**



## مستخلص البحث

شهدت السنوات القليلة الماضية استخدامًا كبيرًا ومتزايدًا في مجال تحليل الصور الطبية. ساعدت هذه الأدوات أخصائي الأشعة والأطباء على التشاور أثناء إجراء تشخيص معين. في هذا البحث قمنا بتطبيق العلاقة بين القياسات الإحصائية والهندسية ورؤية الكمبيوتر والصور الطبية ، واستخدام انموذج الانحدار اللوجستي لاستخراج ميزات صور سرطان الثدي ، للتمييز بين شكل الكتلة (الليفية والدهنية) من خلال تحديد منطقة الاهمية بمساعدة الكادر الطبي المختص (ROI) من الكتلة ، ان الانموذج النهائي الذي تم توفيقه للانحدار اللوجستي الذي استخدم أهم المتغيرات التي لها تأثير واضح على صور شكل الكتلة لسرطان الثدي هي الانحراف ، والتفرطح ، ومركز الكتلة ، والزاوية. مع دقة تقارب 89٪ و مساحة تحت المنحنى (AUCROC) بنسبة 88٪ ، خلصت الدراسة أيضًا إلى أن كتلة الورم الليفي لها حجم صغير ، وأيضًا أقل بياضًا من كتلة الورم الدهني.

**نوع البحث:** ورقة بحثية.

**المصطلحات الرئيسة للبحث:** الانحدار اللوجستي, استخلاص الميزات, الصور الطبية, الدقة ,المساحة تحت المنحنى