

Comparative study of logistic regression and artificial neural networks on predicting breast cancer cytology

Yousra Abdulaziz Mohammed, Eman Gadban Saleh

Institute College of medicine, University of Fallujah, Baghdad, Iraq

Article Info

Article history:

Received Apr 28, 2020

Revised Jun 15, 2020

Accepted Jul 8, 2020

Keywords:

Artificial neural networks

Breast tumor

Classification

Logistic regression

UCI ML repository

Validation

ABSTRACT

Currently, breast cancer is one of the most common cancers and a main reason of women death worldwide particularly in developing countries such as Iraq. Our work aims to predict the type of tumor whether benign or malignant through models that were built using logistic regression and neural networks and we hope it will help doctors in detecting the type of breast tumor. Four models were set using binary logistic regression and two different types of artificial neural networks namely multilayer perceptron MLP and radial basis function RBF. Evaluation of validated and trained models was done using several performance metrics like accuracy, sensitivity, specificity, and AUC (area under receiver operating characteristic ROC). Dataset was downloaded from UCI ml repository; it is composed of 9 attributes and 699 samples. The findings are clearly showing that the RBF NN classifier is the best in prediction of the type of breast tumors since it had recorded the highest performance in terms of correct classification rate (accuracy), sensitivity, specificity, and AUC (area under receiver operating characteristic ROC) among all other models.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Yosra Abdulaziz Mohammed

College of Medicine

University of Fallujah

ALanbar, Iraq

Email: yosra_mohammed@uofallujah.edu.iq

1. INTRODUCTION

Currently, breast cancer is one of the most common cancers and a main reason of women death worldwide particularly in developing countries like Iraq [1]. It is accounting for 25% of all malignancy, with about new 1.57 million cases in 2012 [1]. It is the first – ranked cause of woman cancer-related deaths. Although fundamental improvement in the survival for this disease has been recorded in countries with high-resource, the risk go on to rise, recording high mortality rates in developing countries [1]. The Iraqi cancer Registry disclosed the breast cancer is the highest rates of cancer cases (19.1%) and the highest annual estimate of cancer in women (25.8 per hundred thousand of the female population). The second –ranked incidence of cancer mortality was breast cancer (2.7/100,000 populations) [2]. According to a study, the rate of continuing to live is 88% post five years of detection and 80% post ten years of detection in early stage which means that about 88% of women diagnosed with breast cancer will survive for at least 5 years after their early diagnosis; that's why it is serious to detect breast cancer as earlier as possible [3].

Nowadays, diagnosing cancer using machine learning ML is influential, since ML has the ability of deductions which the traditional statistical methods do not have [4]. A classification model or a "classifier" can be powerful in assisting oncologists to correctly diagnose breast tumor biopsy. Classification means guessing the category of new incidence from a set of predictive features from the sample dataset. The

existence or non-existence of the malignancy can be assured by the outcome of biopsy therefore it is counted as a binary output.

The application of Logistic Regression (LR) in several areas is growing especially in the medical domain. It is a statistical algorithm that can be used perfectly for binary classification to find the association between one or more continuous or categorical parameter and a bilateral dependent output [5]. The technique of logistic regression is able to designate different datasets to previously defined classes, this is carried out by establishing the discrimination rules which are set in the training stage and then used to allocate the new incidences to classes determined in advance [6].

Artificial neural networks (ANN) are commonly used as a robust decision making systems especially for medical diagnosing after being trained using historical data set. ANNs advantages can be summarized in that tuning neural weights is done online with no need to any pre-training phase, and persistence and performance systems is ensured. ANN is a powerful classifier that represents a nonlinear relationship between input and output. Basically, a simple ANN consists of three layers, an input layer, hidden layer/s and an output layer. At the input layer the inputs are weighted, i.e each input value is multiplied by certain weight. At the hidden layer, all weighted inputs along with a bias are summed. Finally at the output layer the summed value obtained is converted to activation signal using transform function. The ANN is trained with a learning algorithm according to the type of the given problem. Generally the learning algorithms are either supervised learning, unsupervised learning or reinforcement learning [7].

The main goal of this work is to compare the performance of two techniques, logistic regression and neural networks to define the more powerful technique in classifying the type of breast tumor into benign or malignant classes. The other sections of this paper are: section two is a literature review part and work done in the area of Breast Cancer. Third section presents data collection and preprocessing and the used methodology. The experimental results are presented in Section four, and finally Sections five and six discuss results, conclusion and future work.

2. LITERATURE REVIEW

At present, physicians are making surgical biopsy to decide whether the breast tumors are benign or malignant. Since biopsy might be critical, then it must be halted as possible as we can. Thus, to detect the type of tumor and keep away from unnecessary surgical biopsy, a smart system or classifier implementing machine learning techniques can be useful for both patients and physicians. Recently, many different algorithms have been developed for diagnosing the breast tumors. Breast Cancer detection of medical images is a very important component of image processing technique. Yousif A. Hamad et al. classified mammogram images into three classes: normal, benign and malignant through an approach that includes some functions for noise removal, then features were improved to get better characteristics of medical images for a correct diagnosis using balance contrast enhancement techniques (BCET). The outcome of second stage was subjected to image segmentation using FCM clustering method (Fuzzy c-Means) and Thresholding technique to cut the out borders of the breast and to locate the Breast Tumor boundaries (shape, area, spatial sizes, etc.) in the images. The third stage was feature extraction using Discrete Wavelet Transform (DWT). At last, classification of the stage of Breast Tumor into benign, malignant or normal was done using PNN (Probabilistic Neural Network), the findings was a high classification rate of (90%) [8]. Mahfuzah Mustafa et al. have conducted a method to improve the Gradient Vector Flow (GVF) Snake Active Contour segmentation technique in mammography segmentation. Digital Database of Screening Mammograms (DDSM) was used for the purpose of screening. The result shows that Chan-Vese technique outperforms Localized Active Contour with 90% accuracy [9]. Muhammet Fatih Ak applied Data visualization and machine learning techniques including random forest, support vector machine, naïve Bayes, logistic regression, decision tree, k-nearest neighbors, and rotation forest to data of breast cancer tumors from Dr. William H. Walberg of the University of Wisconsin Hospital. The logistic regression model with all features outperformed and had scored 98.1% classification accuracy [10]. Jabeen sultana and Abdul Khader Jilani predicted the existence of Breast cancer by evaluating dataset on various classifiers like Multi-Layer Perceptron (MLP), Random Forest, Simple Logistic-regression method, IBK, K-star, Decision table, Decision Trees (DT), PART, Multi-Class Classifiers and REP Tree. Findings showed that Simple Logistic Regression was the best model followed by other methods [11]. Mohammed Abdulrazaq Kahya used the BreaKHis (The Breast Cancer Histopathological Images) datasets to develop a method to classify breast tumors into two classes benign and malignant. The method used was adaptive penalized logistic regression APLR by smoothing the features matrix to raise the overall classification accuracy of breast cancer histopathological images. The findings showed that APLR is a promising technique for classification of medical image and tumors diagnosis [12].

This work represents a comparison between two different machine learning techniques in the prognosis of breast tumors whether malignant or benign. The model developed in this study can assist and help oncologists in breast cancer detection.

3. MATERIAL AND METHODS

3.1. Data

Breast-cancer-Wisconsin data downloaded from machine learning repository of UCI is the dataset used in this work [13]. The dataset has 699 breast FNAs. It is composed of 11 columns, every row is observations of a patient's breast FNA that was obtained from medical analysis. Column no.1 is the identification code of each patient and was removed since it is not required for the analysis; the next nine columns represent the predictors utilized to analyze every FNA obtained from patient breast tumor; clump thickness (range to which cell aggregates, mono- or multilayered), uniformity of cell sizes, uniformity of cell forms, marginal adhesion (coherence of the marginal cells of the cell aggregates), size of the single epithelial cell (diameter of the inhabitation of the biggest cells comparative to erythrocytes), Bare nuclei (the ratio of single cell nuclei that were freed from encirclement cytoplasm), chromatin blandness, nucleolus normality, and mitosis [14]. Column no. 11 represents the outcome or the tumor class; malign=4 and benign=2.

"All malignant aspirates were histologically confirmed whereas FNAs diagnosed as benign masses were biopsied only at the patient's request. The remainder of benign cytologies were confirmed by clinical reexamination 3 and 12 months after the aspiration. Masses that produced unsatisfactory or suspicious FNAs were surgically biopsied" [15].

The nine features are numbers ranging from 1 to 10, which were recorded via lab tests or medical analysis. The proportions of the outcome or the "Class" is; Malignant: 241(34.5%).and Benign: 458 (65.5%) At first, the data was converted to Excel sheet for easiness, then data cleaning was done by replacing missing values by the mean of the nearby attribute values [16], after that the data has been imported to the statistical program SPSS, V19.0 for analysing to build the statistical models.

3.2. Logistic regression (LR)

The logistic function in (1) represents the conditional probability for occurrence of an event "dependant variable" [17],

$$\text{Probability}(Y = 1) = \frac{e^{-z}}{1+e^{-z}} \tag{1}$$

Where probability values are in the range of 0 to 1, z (or logit) is the linear multiple regression model of the predictors

$$z = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n \tag{2}$$

The coefficients of the independent variables are $\beta_0 \dots \beta_n$ which are computed by estimating the maximum likelihood, $X_1 \dots X_n$ are explanatory variables and n is their number While reference probability is defined as,

$$p(y = 0) = 1/(1 + e^{-z}) \tag{3}$$

the log(odds), or log-odds ratio, is defined by,

$$\log(\text{odds}) = \ln \left[\frac{p}{1-p} \right] = z \tag{4}$$

It represents the log of the ratio of the chance of an event to happen, $p(Y=1)$, to the chance it will not $p(Y=0)$, it is computed from the probability of every incidence. The odds ratio is defined as

$$\frac{p}{(1-p)} = e^{\beta x} \tag{5}$$

Two models were developed using LR, a full model with all 9 predictors and a reduced model with only 5 attributes which showed statistical significance i.e p-value was less than 0.05.

Validation of a model is an essential stage in model building [18]. The validation is using different data set pertaining the values of the coefficient as for the training data to calculate the percentage of correct

classifications. The percentage of correctly predicted samples from the training samples must be \geq to the validated samples [19].

There are a lot of statistical methods of validation of a model in binary logistic regression such as bootstrapping, jackknife technique, repeated data-splitting, and data splitting [20]. We implemented data-splitting method in this study. Random division of the data into two sets; the first set with of 80% (550) sample used for building the LR model having 177 malignant and 373 benign, while the second set was composed of 20% (149) sample with 64 malignant and 85 benign was used for the purpose of validation of the two models.

Both full and reduced models were trained in the start with the training data set, after that the validation data set was applied to the fitted models to assess the performance of the two models. The value of the obtained posterior probability for malignancy was then classified into two divisions; where values ranging from 0 to 0.5 were assigned to benign class, and values ranging from > 0.5 to 1 were assigned to malignant class. Evaluation of the two models was done then in terms of metrics mentioned in section 3.4.

3.3. Neural networks

In our study, two types of ANN were used. The first one is Multilayer Perceptron (MLP) network as shown in Figure 1 which is a well known network architecture has been used in medical, engineering, mathematical modeling research. In MLP, a fixed value (bias) along with weighted sum of inputs are propagated to the hidden layer via a transfer function to generate the output, and the topology of feed-forward layers arrangement of units is called Feed Forward Neural Network (FFNN) [21]. The learning ability of the MLP is highly increased by the hidden layer. The input is modified by the activation function of the network so as to give a required output. Model building is strongly affected by the hidden nodes number, hidden layers number, and the type of activation function selection [22]. The output of a MLP NN is;

$$y(j) = T \left(\sum_{k=0}^n w_k(j) * x_k(j) + c \right) \quad (6)$$

Where $y(j)$ is the output value, x_k is the input vector, T is the transform function, c is a constant, w_k is the vector of weights, n is the size of input vector. The equation is in discrete time j [7].

The second type of NN used is radial basis function neural network RBF which is based on supervised learning. RBF NN are efficient in modeling nonlinear data and training this type of NN can be done in one stage counter to MLP. In the hidden layer RBFNN uses nonlinear Gaussian transfer function whereas in the output layer it uses a linear summation transfer function. The real values of the n -dimensional input vector X is fed to all units in the hidden layer at the same time as shown in Figure 2. The Gaussian RBF is given by;

$$\phi(\|x - x^{(i)}\|) = \exp\left(-\frac{\|x - x^{(i)}\|^2}{2\sigma_i^2}\right) \quad (7)$$

Where the functions $\phi(\|x - x^{(i)}\|)$, $i=1,2,\dots,N$ are called the RBFs, where a p -norm (often the Euclidean 2-norm) denotes $\|\cdot\|$, $x^{(i)}$ is the basis function centre and σ_i is its radius. A linear combination of basis functions can be used for approximation of a nonlinear function. The output: $R^n \rightarrow R$, of the network is thus

$$F(x) = \sum_{i=1}^N w_i \phi(\|x - x^{(i)}\|) \quad (8)$$

Where N represents the neurons number in the hidden layer and the real parameters w_i , $i = 1, 2, \dots, N$ are the linear output neurons weights [23].

To train RBF networks, once the type of radial basis function is selected, all needed to do is choosing the functions' dimensions and centers and estimating the output neuron weights. For the ANN, two models were developed using two different types of NN, namely MLP and RBF. The architecture of the MLP neural network had four layers; the input layer consisted of 9 input elements, corresponded to the data taken from cytology, then two hidden layers with sigmoid activation function, the first one had 7 nodes while the second hidden layer consisted of 5 nodes and the output layer with 2 neurons, representing 0 for benign and 1 for malignant lesions. A back propagation algorithm based on scaled conjugate optimization technique was used to model MLP for our dataset. To get the optimum neural network structure, a considerable number of neural networks have been simulated by changing the number of hidden layers, hidden nodes, iterations and learning rates. Whereas the feed forward topology of RBF network developed for this work was composed of 3 layers, input layer with the 9 input elements, a single hidden layer having a nonlinear RBF activation function with 9 neurons fully interconnected to the output layer units and a linear output layer with 2

elements. The error function E used to index the learning efficiency of both neural networks was the Sum of Squared Error (SSE) criterion function which had to be minimized over the given training set. The performance of the NN models was determined by dividing the dataset into two separate sets 70% of samples for the training and 30% for the validation.

After the networks had been trained perfectly using the training data, each network was tested by presenting the testing set to the trained network and a diagnostic output vector of 0's and 1's was generated.

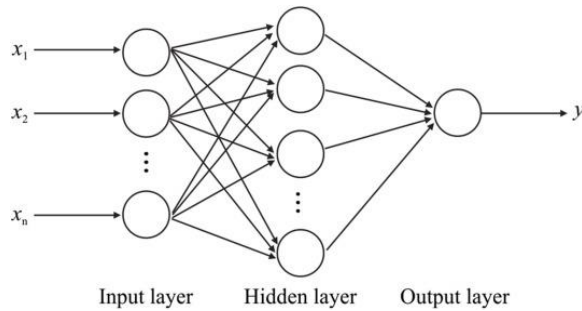


Figure 1. Multilayer feedforward MLP

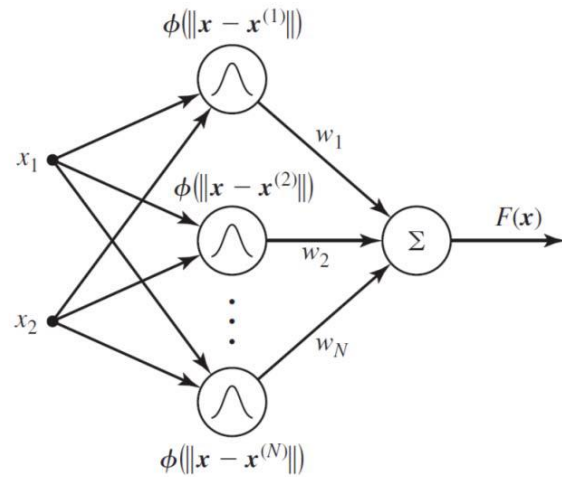


Figure 2. Radial basis function NN

3.4. Performance evaluation

The widely used metric in classification process is accuracy [21, 24, 25]. It is the rate of correctly classified test samples [21, 25]. Sensitivity is the ability of a model to correctly identify true positive instances and is called true positive rate whereas specificity is the ability of the model or test to correctly identify the true negative instances and is called true negative rate. The AUC (area under Receiver Operating Curve) was also used in the evaluation of models performance [26]. Two models were built and evaluated using metrics indicated by (9)-(13) and confusion matrix shown below,

$$CCA = (TP + TN)/(TP + FP + TN + TP) \tag{9}$$

TP represents the number of positive samples classified correctly and TN is the number of negative samples correctly classified. FN is the number of positive samples incorrectly classified as negative. FP is the number of negative samples incorrectly classified as positive [27].

$$Sensitivity = \frac{TP}{TP+FN} \tag{10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

$$AUC = \frac{1+TP_{rate}-FP_{rate}}{2} \tag{12}$$

Where, TPrate is sensitivity and

$$FP_{rate} = \frac{FP}{FP+TN} \tag{13}$$

Confusion matrix			
		predicted class →	
		Benign (0)	Malignant (1)
Actual class ↓	Benign (0)	TN	FP
	Malignant (1)	FN	TP

4. RESULTS

In this study, different models were set using IBM SPSS statistics 19 software and the performance of the classifiers was compared. The breast cancer dataset used was obtained from UCI-Machine Learning repository, and was fed to our Logistic-regression models, MLP, and RBF neural networks. Each classifier was well trained with the dataset and a Model is set and validated with test samples, then results were obtained.

From Table 1, we can see that clump thickness, marginal adhesion, bare nuclei, and bland chromatin have p-values <0.05 which means that these are significant predictors of malignancy in the set model. On the other hand, Table 2 shows the reduced model coefficients which were calculated from model training using the training samples by the stepwise method.

For the neural network, both MLP and RBF models has shown a considerable improvement in all performance metrics than those scored by logistic regression models, and RBF NN has preceded all the other models developed in this work with highest correct classification rate of 95.4%, sensitivity of 98.5%, AUC of 96.125%, except for specificity which was 93% the same for both MLP and RBF. The performance of the developed logistic regression models and neural networks is compared in Table 3 and Figure 3 on testing samples in terms of percentage of correct classification rate (accuracy), sensitivity, specificity, and areas under ROC curve AUC.

Table 1. Parameter estimation of the full LR model using training sample

Variable	Coefficient estimate	Standard error	Wald	Sig.
clump_thickness	.586	.180	10.648	.001*
uniformity_of_cell_size	.218	.238	.838	.360
uniformity_of_cell_shape	.159	.260	.374	.541
marginal_adhesion	.304	.130	5.476	.019*
single_epithelial_cell_size	-.110	.243	.205	.650
bare_nuclei	.394	.105	13.963	.000*
bland_chromatin	.483	.204	5.589	.018*
normal_nucleoli	.115	.137	.708	.400
metosis	.487	.477	1.042	.307
Constant	-10.112	1.460	47.973	.000

*significant at level of 0.05

Table 2. Parameter estimation of the reduced LR model using the training samples.

Variable	Coefficient estimate	Standard error	Wald	Sig.
clump_thickness	.627	.172	13.323	.000
uniformity_of_cell_size	.376	.149	6.409	.011
marginal_adhesion	.282	.122	5.320	.021
bare_nuclei	.396	.099	16.021	.000
bland_chromatin	.566	.193	8.643	.003
Constant	-9.911	1.312	57.064	.000

*significant at level of 0.05

Table 3. Comparative performance of the four models on validation samples

Model	Full LR model	Reduced LR model	MLP NN	RBF NN
Sensitivity %	68.75	64.1	95.1	98.5
Specificity %	83.53	85.88	93.8	93.8
Accuracy %	74.7	74.03	94.2	95.4
AUC %	76.1	74.97	94.45	96.125

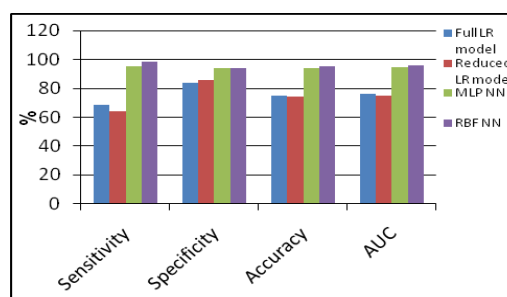


Figure 3. Comparison of models performance

5. DISCUSSION

ANN and LR are widely used for tasks of prediction or classification. In this work, comparison of the four models developed was based on the validation dataset after the models had been sufficiently trained with the training data to assure whether the output of these models will predict future samples precisely. The first model built was logistic regression full model including all the 9 features and the second model was the reduced model using the stepwise method where the variables with the largest p-value has been removed keeping the significant variables only, however none of the metrics used was improved by the reduced model except the specificity as shown in Table 3 and Figure 3.

The result analysis of our study showed that the ability of RBF NN to diagnose breast cancer is superior to Binary Logistic Regression models (both full and reduced) and to MLP achieving highest and most accurate results where the accuracy of the RBF NN model was 95.4% and a sensitivity of 98.5% followed by MLP. While logistic regression method showed much less accuracy (74.7% for full model and 74.03% for reduced), sensitivity (68.75% for full model and 64.1% for reduced), specificity (83.53% for full model and 85.33% for reduced), and AUC (76.1% for full model and 74.79% for reduced), in comparison with neural networks.

6. CONCLUSION

This paper represents a comparison of the diagnosing performance of two different machine learning techniques which are logistic regression and neural networks in the prognosis of breast tumors whether malignant or benign using the breast-cancer-Wisconsin. data file. The diagnosing performance of two types of artificial neural networks and binary logistic regression through the different models that were built, was compared based on sensitivity, specificity, AUC, and accuracy or CCR criteria. The results showed that using ANN in predictive analysis in oncology is more powerful than logistic regression algorithm, RBF outperforms MLP and logistic regression for all metrics. The sensitivity CCR, AUC values for RBF on testing data were the highest. The findings indicate that this study might be helpful in the diagnosis of breast tumors. For Future work other machine learning algorithm can be experimented or hybrid algorithm which is a combination of existing two or more algorithms can be applied to create a predictive model which can predict with higher accuracy. Other types of neural networks such as convolutional neural network networks CNN can be implemented, other classifiers such as multiple logistic regression can be tested.

REFERENCES

- [1] NA. Alwan, "breast cancer among iraqi women: Preliminary findings from a regional comparative breast cancer research project," *Journal of global oncology*, vol. 2, no. 5, pp. 255, 2016.
- [2] Iraqi Cancer Board. Results of the Iraqi Cancer Registry Baghdad, Iraq, Iraqi Cancer RegistryCenter, Ministry of Health, 2015. [Online]. Available: <https://moh.gov.iq/upload/upfile/ar/833>. [Accessed on Jan 27, 2020].
- [3] M. Rathi, A.K. Singh, "Breast cancer prediction using Naïve bayes classifier," *International Journal of Information Technology & Systems*, vol. 1, no. 2, pp. 77-80, 2012.
- [4] JA Cruz, DS Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*. vol. 2, pp. 2-21, 2006.
- [5] Ronny Gunnarsson, "Logistic regression," *Science Network TV*, First published June 22, 2014, Last revised August 30, 2019. <https://science-network.tv/logistic-regression/>, [Accessed on Sep 24, 2019].
- [6] DW Hosmer, S. Lemeshow, *Applied logistic regression*, New York: Wiley; 1989.
- [7] M.G Kanojia, S. Abraham, "Breast cancer detection using RBF neural network," in *2nd International Conference on Contemporary Computing and Informatics (IC3I) 2016*, pp. 363-368, 2016.
- [8] Y Hamad, K Simonov, & M. B. Naeem, "Breast cancer detection and classification using artificial neural networks," in *1st Annual International Conference on Information and Sciences (AiCIS)*, pp. 51-57, 2018.
- [9] M Mustafa, *et al.*, "Mammography image segmentation: Chan-Vese active contour and localised active contour approach", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 5, no. 3, pp. 577-588, 2017.
- [10] M. F. Ak, "A Comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications", *Healthcare*, vol. 8, no. 2, p. 111, 2020.
- [11] J Sultana, and A. K. Jilani, "Predicting breast cancer using logistic regression and multi-class classifiers," *International Journal of Engineering & Technology*, vol. 7, no. (4.20), pp. 22-26, 2018.
- [12] M. A Kahya, "Classification enhancement of breast cancer histopathological image using penalized logistic regression," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1, pp. 405-410, 2019.
- [13] D Dua, C Graff, "UCI Machine Learning Repository". Irvine, CA: University of California, School of Information and Computer Science, 2019. [Online]. Available: <http://archive.ics.uci.edu/ml>.
- [14] F Leisch, E. Dimitriadou, "Machine learning benchmark problems," *Cran.r-project*, 2015. [Online]. Available: <https://cran.r-project.org/web/packages/mlbench/mlbench.pdf>. [Accessed on Sep 24, 2019]
- [15] WH Wolberg, OL Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in *Proceedings of the national academy of sciences*, vol. 87, no. 23, pp. 9193-9196, 1990.

- [16] H. Kang, "The prevention and handling of the missing data," *Korean journal of anesthesiology*, vol.64, no. 5, pp. 402-406, 2013. doi:10.4097/kjae.2013.64.5.402.
- [17] J Schwarz, p.Heidi Bruderer Enzler, "Research methodology: Tools; Applied data analysis (with SPSS)," Lucern University, 2014. [Online]. Available: <http://www.schwarzpartners.ch> [extracted on Sep 30, 2019].
- [18] J. Shao, "Linear Model Selection by Cross-Validation," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 486-494, 1993.
- [19] H. Yusuff, N. Mohamad, U.K. Ngah, A. Yahaya, "Breast cancer analysis using logistic regression," *International Journal of Research and Reviews in Applied Sciences*, vol. 10, no. 1, pp. 14-22, 2012.
- [20] S. Rana, H Midi, SK Sarkar, "Validation and performance analysis of binary logistic regression model," in *Proceedings of the WSEAS International Conference on Environment, medicine and health sciences*, Penang, Malaysia, pp. 51-55, 2010.
- [21] Y. Mohammed, "A comparative study between artificial intelligence techniques in an automatic infant's pain cry identification system," in *2018 1st Annual International Conference on Information and Sciences (AiCIS)*, Fallujah, Iraq, pp. 33-39, 2018. doi: 10.1109/AiCIS.2018.00019
- [22] J. Padmavati, "A comparative study on breast cancer prediction using RBF and MLP," *International Journal of Scientific & Engineering Research*, vol. 2, no. 1, pp. 1-5, 2011.
- [23] H. Jafarnejadsani, J. Pieper, J. Ehlers, "Adaptive control of a variable-speed variable-pitch wind turbine using radial-basis function neural network," *IEEE transactions on control systems technology*, vol. 21, no. 6, pp. 2264-2272, 2013. doi: 10.1109/TCST.2012.2237518.
- [24] H Sofian, J. T. C Ming, S Muhammad, and N. M Noor, "Calcification detection using convolutional neural network architectures in intravascular ultrasound image," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 3, pp. 1313-1321, 2020. doi: 10.11591/ijeecs.v17.i3.pp1313-1321.
- [25] Y. Abdulaziz and S. M. S. Ahmad, "Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients," *2010 International Conference on Information Retrieval & Knowledge Management (CAMP)*, pp. 260-263, 2010, doi: 10.1109/INFRKM.2010.5466907.
- [26] S Miri Rostami, M Ahmadzadeh, "Extracting predictor variables to construct breast cancer survivability model with class imbalance problem," *Journal of AI and Data Mining*, vol. 6, no. 2, pp. 263-276, 2018. doi: 10.22044/JADM.2017.5061.1609.
- [27] R Rosly, M Makhtar, M. K Awang, M. I Awang, M. N. A Rahman, and H Mahdin, "Comprehensive study on ensemble classification for medical applications," *International Journal of Engineering & Technology*, vol. 7, no. (2.14), pp. 186-190, 2018.

BIOGRAPHIES OF AUTHORS



Youstra A.A Mohammed, she is currently attached as Lecturer in University of Fallujah Anbar, Iraq. She received her B. Eng. in Computers and Control Engineering from University of Technology Baghdad, Iraq and her Master of Information Technology in Computer sciences from University Tenaga Nasional, Malaysia (UNITEN). Her research interest is in the machine learning, E-Learning, classification, Neural Networks, signal processing and image processing. Her current work concentrates on medical data analysis and classification in deep learning.



Eman Gadban Saleh, Received her Ph.D. degree in gynecology and obstetrics from Council of Arab Health Ministers, currently works as a lecturer in college of medicine, university of Fallujah, Iraq. Her research interests include oncology, obstetrics and gynecology.