

# Breast invasive ductal carcinoma diagnosis using machine learning models and Gabor filter method of histology images

Rania R. Kadhim, Mohammed Y. Kamil

College of Science, Mustansiriyah University, Baghdad, Iraq

## Article Info

### Article history:

Received Jun 9, 2022

Revised Aug 7, 2022

Accepted Aug 26, 2022

### Keywords:

Breast cancer

Computer-aided diagnosis

Gabor filter

Invasive ductal carcinoma

Machine learning

## ABSTRACT

Breast cancer is the most common type of cancer in women and the leading cause of death from a malignant growth in the world. Machine learning methods have been created to help with cancer detection accuracy. There are several methods for detecting cancer. Histopathological images are more accurate. In this study, we employed the Gabor filter to extract statistical features from invasive ductal carcinoma histopathology images. From the histopathological images, we chose 100, 200, 400, 1000, and 2000 at random. These statistical features were used to train several models to classify these images as malignant or benign, including the decision tree, quadratic discriminant analysis, extra randomized trees, gradient boosting, Gaussian process, Naive Bayes, nearest centroid, multilayer perceptron, and support vector machine. The models' accuracy, sensitivity, specificity, precision, and F1\_score were examined. The models produced the highest results when there were 100 images and a wavenumber of 0.2. While as the number of images increased, the models' effectiveness reduced. The most obvious finding to emerge from this study is that we suggest using deep learning instead of machine learning models for large datasets.

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



## Corresponding Author:

Mohammed Y. Kamil

College of Science, Mustansiriyah University

St. Palastine, Baghdad, Iraq

Email: m80y98@uomustansiriyah.edu.iq

## 1. INTRODUCTION

Cancer is the main cause of death in humans worldwide. Only in 2018, 18.1 million patients were classified as having cancer; breast cancer is the most common type [1]. Breast cancer is a potentially deadly malignant tumor that affects women worldwide and passed lung and bronchus cancer as the second largest cause of death in women [2]. Each year, over 1.5 million women are diagnosed with breast cancer, accounting for about 15% of all female deaths, and the percentage is increasing [3]. This number is expected to reach 19.3 million by 2025, according to the world health organization [4]. Invasive ductal carcinoma (IDC), also known as infiltrating ductal carcinoma, is the most common subtype of breast cancer, affecting roughly 80% of patients [5]. IDC originates in a milk duct and then spreads to the breast's fibrous or fatty tissue. It may spread to lymph nodes and other body locations [6]. Primary detection and accurate diagnosis can minimize breast cancer mortality rates. To differentiate benign from malignant tumors, frequent testing and medical imaging, such as magnetic resonance imaging (MRI) [7], mammography [8]-[10], and ultrasound, are important.

However, pathological diagnosis using histological imaging is recommended since it provides more direct evidence for classification, evaluation, and experimental treatment, although it is a time-consuming and tiring process. Pathologists who are professionals can also make diagnostic errors [11]. Pathologists must

visually examine histology samples under a microscope to identify IDC, which is a complex and time-consuming process because of their diverse appearance, texture, and structure. Pathologists would profit from automating the detection of this type of cancer since it would accelerate the diagnosis and reduce the number of errors. Therefore, computer-aided diagnosis systems are the most effective method for identifying histopathological images as malignant or noncancerous.

This study aims to recognize the effect of the Gabor filter's wavenumber on the models' performance. And know the impact of the number of images on classification results and assess the efficiency of machine learning (ML) models for classifying IDC histopathology images of breast cancer. In addition, evaluation of the Gabor filter orientation values' effectiveness on models' behavior.

The rest of this paper is categorized as follows: section 2 shows related work, and in section 3, a method is presented. The proposed methodology is described in section 4, while the different results obtained are discussed in section 5. Finally, in section 6, we state the conclusions and future work.

## 2. RELATED WORK

Researchers have proposed various methods for detecting breast cancer using histopathology images in recent years. Models based on machine learning are being touted as revolutionary. Many studies have focused on detecting statistical and textural features and feature extraction utilizing the Gabor filter.

Kumar and Batra [12], developed a hybrid technique for extracting the orientation field from mammograms utilizing a Gabor filter and a Gradient-based algorithm. The image database contains 60 images collected from two different datasets mammogram image analysis society (MIAS). The orientation field of a mammogram is evaluated for subjective interpretation, exhibiting both normal and problematic areas. Entropy features are extracted from the normal and abnormal mammograms for objective analysis. The objective analysis yields a 100% accuracy rate with the combined dataset, while subjective analysis yields a 93.33% accuracy rate. The proposed technique was found to be successful in detecting architectural distortion.

Bolhasani *et al.* [13], provided 922 image data of histopathology microscope images from 124 patients with IDC. When compared to previous datasets, this one distinguishes since it includes the same sample from each of the three IDC classes, resulting in around 50 samples in each class. Each specimen has four magnification settings (4x, 10x, 20x, and 40x). All of the collected images were labeled with their diagnostic class, and machine learning models can be trained to recognize IDC classes.

Kumar *et al.* [14], implemented a successful stacked generalized ensemble strategy for classifying H&E histopathology breast cancer images into malignant and benign. For experimental evaluation, six models were used: VGG16, CNN, VGG19, Xception (Elu), Xception (ReLu), and DA. The results of the experiments show that the Stacked Generalized Ensemble method did a lot better than other methods at classifying histopathological breast cancer images. It was able to get a 97.53% accuracy rate compared to existing deep learning models, and the least accurate model, CNN, achieved 43.53%.

Hameed *et al.* [15], used pre-trained VGG19 and VGG16 architectures, and researchers trained four distinct models. Initially, all the individual models performed fivefold cross-validation, including fully-trained VGG16, fine-tuned VGG16, fully-trained VGG19, and fine-tuned VGG19. Then, using an ensemble technique, the ensemble of fine-tuned VGG16 and fine-tuned VGG19 performed competitively in the cancer class. Carcinoma classification sensitivity was 97%, and overall accuracy was 95% for the VGG19 and VGG16 models combined. In addition, the F1\_score was 95.29%. The results demonstrated the deep learning approach presented is effective for classifying complex-natured histopathology images of breast cancer.

Kumar *et al.* [16], introduced the stacked generalized ensemble (SGE) classification scheme for breast cancer into IDC exists or not. SGE method is compared against several machine learning models, including linear discriminant analysis, logistic regression, K-nearest neighbour (KNN), Naive Bayes (NB), classification and regression trees (CART), and support vector machine (SVM), and an ensemble of SVM, CART, and Naive Bayes. Additionally, they compared the proposed SGE's performance against an ensemble of three machine learning methods CART, Naive Bayes, and SVM. Machine learning techniques achieved an accuracy of 81%, while SGE reached a maximum of 87.80%. The studies and findings demonstrate that the proposed methodology outperformed previous methods.

## 3. METHOD

### 3.1. Dataset description

A histopathological image with IDC was used for classification. It is the most prevalent subtype of all breast cancer specimens [17]. It includes 162 full-mount side images of breast cancer caused by IDC.

There are 277,524, 50x50 image patches, 198,738 of which are IDC(-) and 78,786 of which are IDC(+) [18]. Sample of images from the current dataset is illustrated in Figure 1. Figure 1(a) and (b) presents samples of histopathology images taken from the existing dataset for both classes. The dataset is freely available through the Kaggle website.

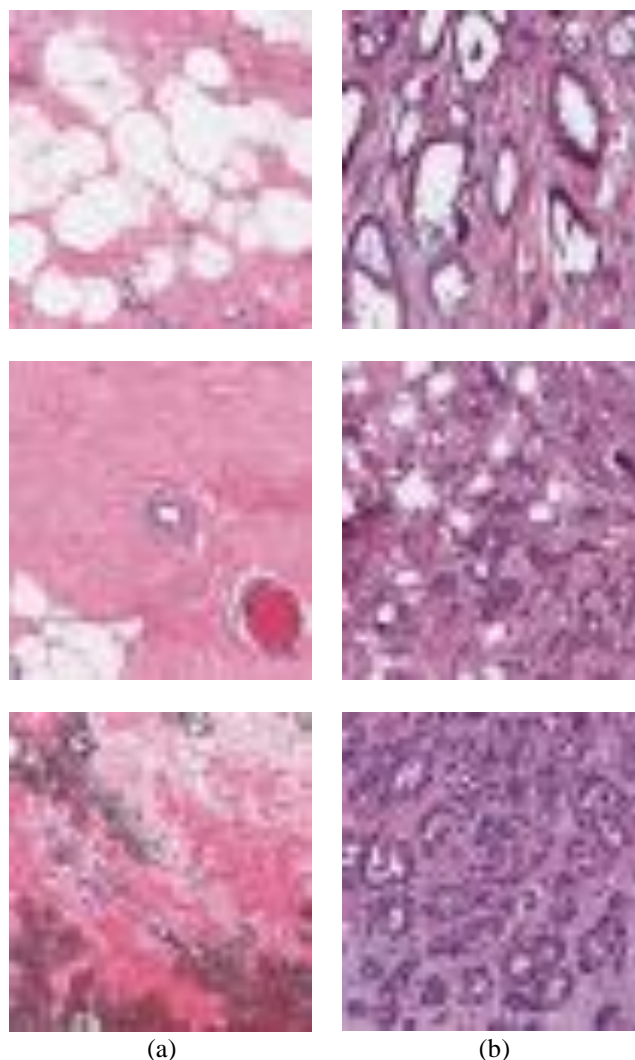


Figure 1. A sample of images from the current dataset (a) IDC (+) and (b) IDC (-)

### 3.2. Gabor filter

The Gabor Filters garnered serious interest because of their ability to approximate particular cells' features in certain species' optical brains. Additionally, it was demonstrated that these filters exhibit good localization characteristics in both the spatial and frequency domains, making them ideally suited for pattern segmentation challenges [19]. A Gabor filter is a sinusoidal plane with a specific frequency and direction. Gabor filters are orientation-sensitive filters that are used to analyze texture. If a Gabor filter configured in a specific direction creates a robust response, its direction matches one of the gratings in a given image. The Gabor filter is employed in two-dimensional image processing applications [20]. The Gabor filter records the local structure of the image in terms of image space, scale, and orientation [21].

Texture analysis is required for automated distribution analysis. Tumors exhibit a variety of micro-patterns at varying frequencies and orientations. These patterns are critical for a CAD system to identify damaging regions. Gabor filters are an effective method for identifying these patterns. Gabor filters are linear filters utilized in a wide range of computer vision applications, including texture classification, face recognition, and cancer detection. Gabor filters excel in coupled localization in frequency and spatial

domains [22]. The optimal value of the Gabor filter parameters is then determined, and the output is the normalized image. The following complex mathematics is used to define these filters [12]:

$$g = \exp \left[ -\frac{(xcos\theta + ysin\theta)^2 + \gamma^2(ycos\theta - xsin\theta)^2}{2\sigma^2} \right] \cdot \exp[i[2\pi(xcos\theta + ysin\theta)\omega + \Phi]] \quad (1)$$

where  $\theta$  is the filter orientation,  $w$  is the sinusoidal wavenumber,  $\gamma$  is the spatial aspect ratio,  $\sigma$  is the Gaussian function's standard deviation, and  $\Phi$  is the phase offset.

### 3.3. Models

A decision tree (DT) is an approximation of a piecewise constant. A decision is a recursive split of the instance space used to classify. DT is easy to understand and interpret. The cost of utilizing the tree is proportional to the number of data points required to train it. DT can deal with both numerical and categorical data. DT is capable of dealing with problems with many possible outcomes [23].

Quadratic discriminant analysis (QDA) is an individual covariance matrix estimated for each class of observations. A problem with QDA is that it can't be used to make things smaller in size. QDA is more flexible than linear discriminant analysis (LDA) because it does not require equal variance and covariance. To put it another way, the covariance matrix for each class in QDA might be different. When you have a short training set, LDA is preferable to QDA. On the other hand, QDA has suggested if the training set is massive and the classifier's variance isn't a key concern [24].

The extra randomized trees (ERT) algorithm is one bagging method type. The basic point of the average method is to build a massive number of estimators exclusively and then average their results. The combined estimator is typically superior to any single-base estimator since its variance is reduced [25].

Gradient boosting (GB) is one of the methods of boosting. GB refers to an extension of boosting to arbitrary differentiable loss functions. GB is a ML approach to improving a model's predictive value by continuously improving [26]. The gradient is the gradual change that occurs during the procedure. Boosting is a means of speeding up the increase in prediction accuracy to a very high level. Both binary and multi-class classification is supported by GB [27].

The Gaussian process classifier (GPC) takes a probabilistic and practical approach to learning in kernel machines, giving it a competitive advantage in model architecture interpretation, integrated learning, and model selection treatment [28]. In comparison to other popular classifiers, GPC offers three significant advantages. First, GPC can deal with high-dimensional and non-linear difficulties during travel mode detection. Second, GPC generates probabilistic outputs rather than deterministic classification findings, accounting for the inherent model uncertainty in travel mode identification. Third, GPC is a non-parameterized model, which means that it may tune hyperparameters directly using training data [29].

Logistic regression (LR) is a method for predicting a classified independent variable from a set of dependent variables. LR is a technique for forecasting the outcome of a categorical dependent variable. As a result, the output must be discrete or categorical [30]. LR can quickly discover the most efficient categorization factors and categorize observations based on various data sources [31].

Naive Bayes (NB) is a Bayes' theory-based classification approach and the condition of predictor independence, where a feature's presence in a class is unrelated to the incidence of any other characteristic [32]. The NB classifier is simple to build and highly successful when dealing large amounts of data. This method is commonly used for text categorization and difficulties involving several classes [33]. The nearest centroid classifier (NC) is a classifier that identifies a group of training samples according to their centroid (mean) distance from the observed item or data. The empirical data and multiple class centroids' distances are ranked, and the closest distance is chosen [34].

A multilayer perceptron (MLP) has a hidden layer or layers (except for one input and one layer of output). In comparison to a single layer perceptron, MLPs can learn non-linear functions [35]. Weights are linked to all connections. However, only three weights are used ( $w_0$ ,  $w_1$ , and  $w_2$ ). Three nodes comprise the input layer. The bias node is set to 1. The other two nodes use both  $X_1$  and  $X_2$  as external inputs (quantities depending upon the given data) [36].

Support vector machine (SVM) is the most important task to find a hyperplane that can distinguish between similar and different data classes. The algorithm is still effective even when the numeral of dimensions exceeds the numeral of samples because it only uses a small portion of the decision function's training points (known as support vectors), it consumes less memory, and has a high degree of stability because of the reliance on support vectors rather than data points. SVM can handle the numerical prediction problem [37].

#### 4. PROPOSED METHOD

Gabor filters are used to construct ten models for the dataset classification. Three wavenumber values of the Gabor filter (0.2, 0.05, and 0.025) and three orientation values of the Gabor filter ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ) were applied to the randomly selected images and assessed for their effect on the classifiers' behavior. The histopathological images are picked at random from the IDC database: 100 images including 50 normal, 50 abnormal, 200 images including 100 normal, 100 abnormal, 400 images including 200 normal, 200 abnormal, 1,000 images including 500 normal, 500 abnormal, and 2,000 images including 1,000 normal, 1,000 abnormal.

Extracted features are inputs to the classifying models; the features are given to DT, QDA, ERT, GB, GPC, LR, NB, NC, MLP, and SVM individually. The Python programming language environment and the Anaconda distribution's supporting libraries are utilized for the experiments. Scikit-Image library is employed for Gabor filter feature extraction and DT, QDA, ERT, GB, GPC, LR, NB, NC, MLP, and SVM classification models.

#### 5. RESULTS AND DISCUSSION

In this study, ten machine learning models were used to classify the data of histopathology images of breast cancer into benign or malignant tumor types. In the beginning, we used the Gabor filter to extract features from the images and use these features as inputs to the models, and then evaluated the performance of these models by evaluating five criteria (accuracy, sensitivity, specificity, precision, and F1\_score).

Part 1: we applied values of the Gabor filter's wavenumber 0.2, 0.05, and 0.025 to 100 randomly chosen images at three orientations of the Gabor filter  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ , as shown in Figure 2. We investigated the impact of the wavenumber on the accuracy, sensitivity, specificity, precision, and F1\_score. In most cases, the accuracy is highest at the wavenumber 0.2, when the orientation values are  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ , and the accuracy drops as the wavenumber decreases. From Figure 2(a), (b), and (c), the ERT model accuracy was higher at wavenumber 0.05 and with the orientation values  $0^\circ$  and  $45^\circ$ . In most instances, the highest sensitivity was reached with a wavenumber of 0.2. At the same time, decreasing the wavenumber decreased the performance of the models. From Figure 2(d), (e), and (f), the ERT model sensitivity increased by decreasing the wavenumber when orientations  $0^\circ$  and  $90^\circ$ . We found the specificity values were greater at wavenumber 0.2 when the orientation was  $0^\circ$ , except for the NC algorithm, which broke this result and attained a high value at wavenumber 0.025 and the ERT algorithm at wavenumber 0.05, as shown in Figure 2(g). Figure 2(h) illustrates that when the orientation is  $45^\circ$ , the specificity remains constant with an increase or decrease rate of 10. In contrast, when the orientation is  $90^\circ$ , the majority of the highest values of specificity occurred when the wavenumber was 0.2. Figure 2(i) shows that the SVC, LR, and GPC models achieved the highest specificity value at wavenumber 0.05. Also, the NB model achieved the highest value at wavenumber 0.025. The effect of wavenumber on value precision was studied. In the majority of instances, the performance of the models improves as the wavenumber at orientations  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ . As demonstrated in Figure 2 (j), (k), and (l), the highest precision value at wavenumber 0.05 and orientation  $0^\circ$  for the ERT model. Also, the highest value at wavenumber 0.05 and orientation  $90^\circ$  for GPC model.

The influence of the wavenumber on the F1\_score criteria was analyzed. F1\_score measures the harmonic mean of sensitivity and precision. The classifiers' behavior was nearly stable at the orientations  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ . Although there are a few exceptions, such as the value of the F1\_score decreases in the NB algorithm when the orientation is  $90^\circ$ , and the wavenumber is 0.025, as seen in Figure 2 (m), (n), and (o).

Part 2: using five criteria, the impact of the number of images in the data on the machine learning algorithm's performance was investigated. We randomly selected 100, 200, 400, 1000, and 2000 images from the data at wavenumber 0.2 and orientation values  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ . The models' effectiveness decreased as the number of images increased, ranked highest with 100 images as shown in Figure 3. As displayed in Figure 3(a), (b), (c), (d), (e), (f), (g), (h), (i), (j), (k), (l), (m), (n) and (o), this is most likely due to the inefficiency of machine learning models with large amounts of data. Therefore, we recommend employing deep learning models with a high number of images.

Part 3: in this section, 100 images were used with a Gabor filter wavenumber of 0.2, and the effect of changing the Gabor filter orientation ( $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ ) on models' performance was analyzed as shown in Figure 4. The majority of the highest accuracy results were obtained at orientation  $90^\circ$ , as seen in Figure 4(a). The sensitivity yielded the highest results at  $45^\circ$ , as shown in Figure 4(b), while the maximum results for specificity and precision were obtained at angle  $0^\circ$ , as shown in Figure 4(c) and (d). According to Figure 4(e), most F1\_scores were higher at  $45^\circ$ .



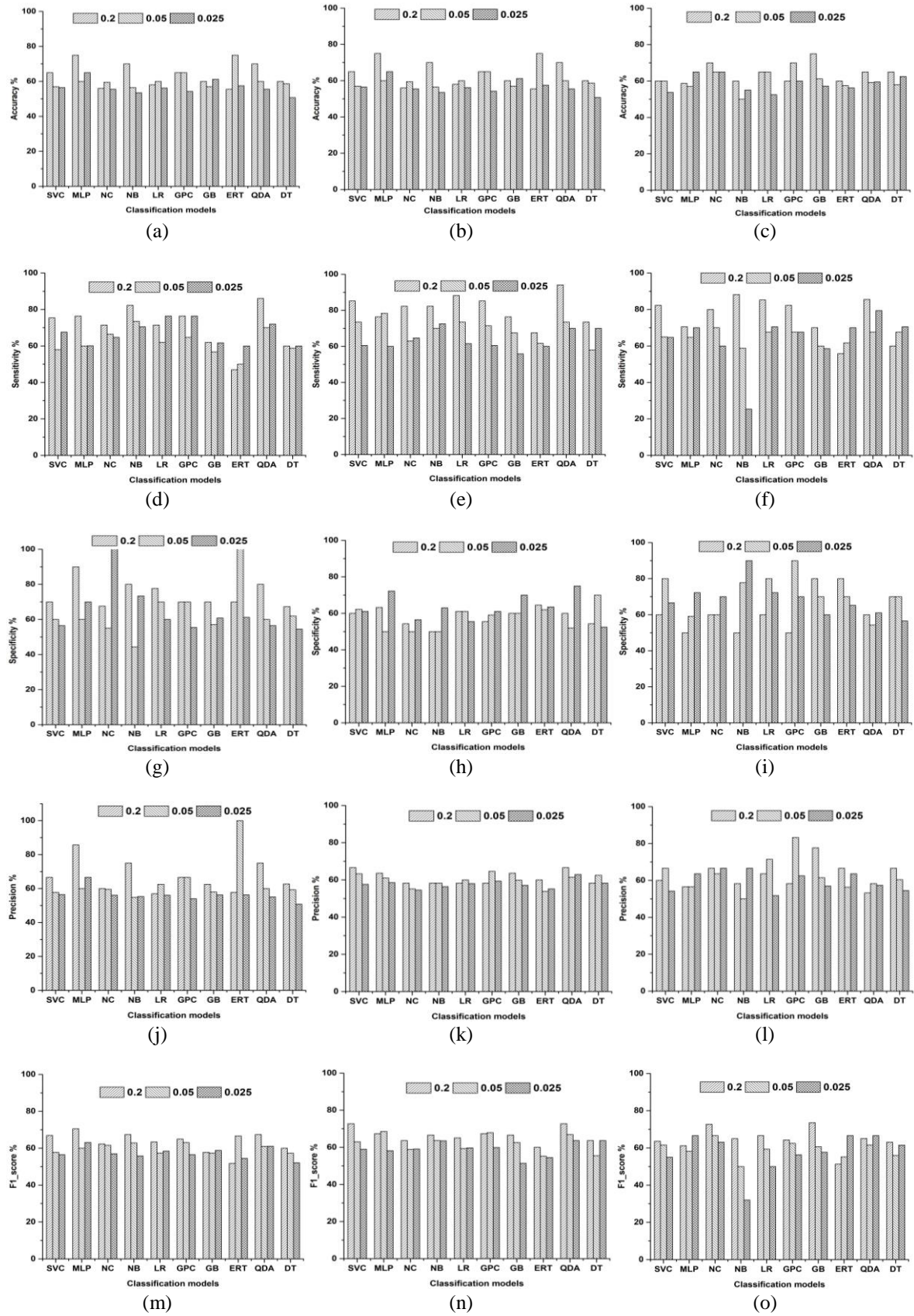


Figure 2. The performance of algorithms using 100 images of different Gabor wave numbers at (a), (d), (g), (j), (m) 0° orientation, (b), (e), (h), (k), (n) 45° orientation, and (c), (f), (I), (l), (o) orientation 90°

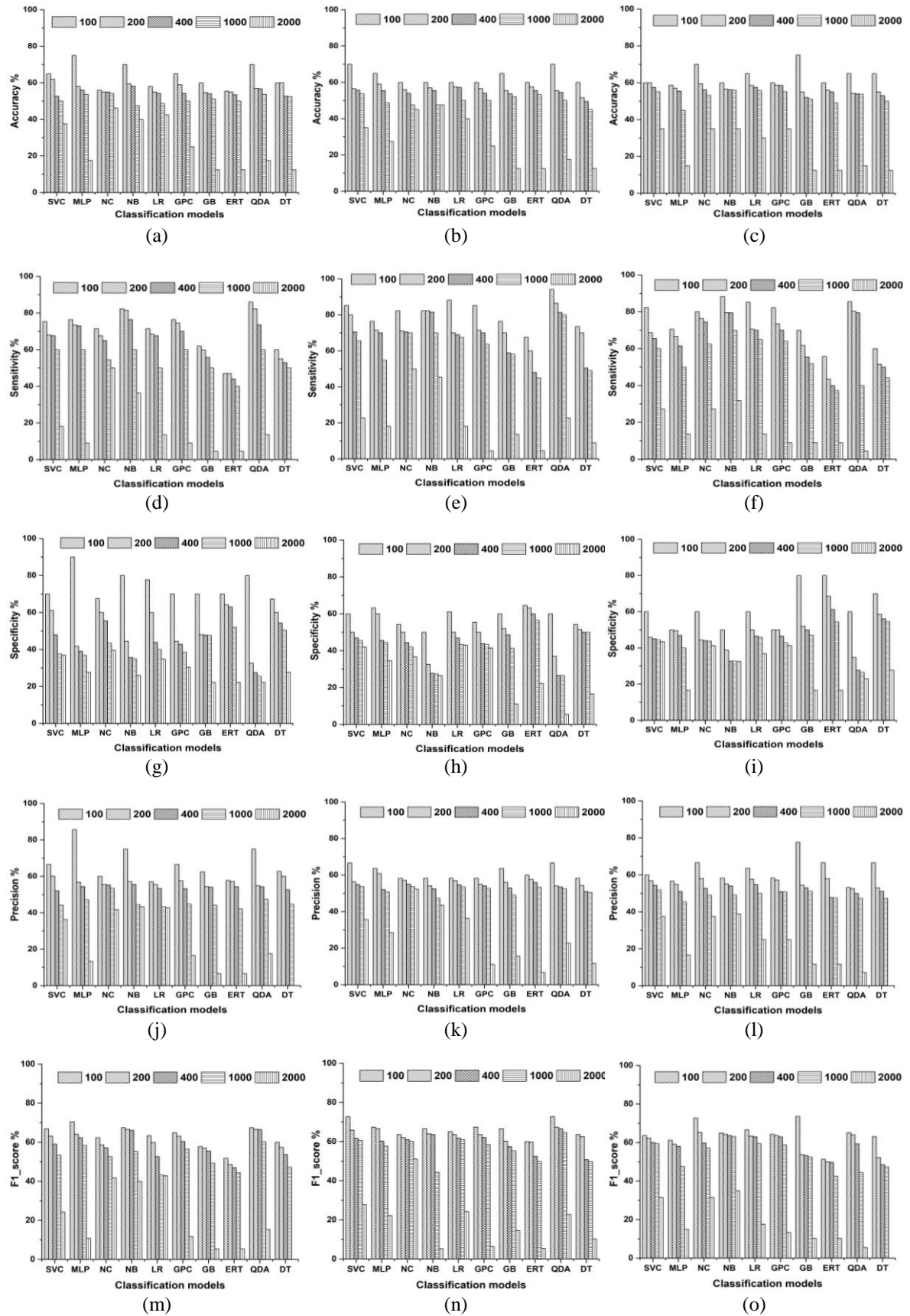


Figure 3. The performance of the algorithms with an increasing number of images at a Gabor wave number of 0.2 at (a), (d), (g), (j), (m) 0° orientation, (b), (e), (h), (k), (n) 45° orientation, and (c), (f), (i), (l), (o) 90° orientation



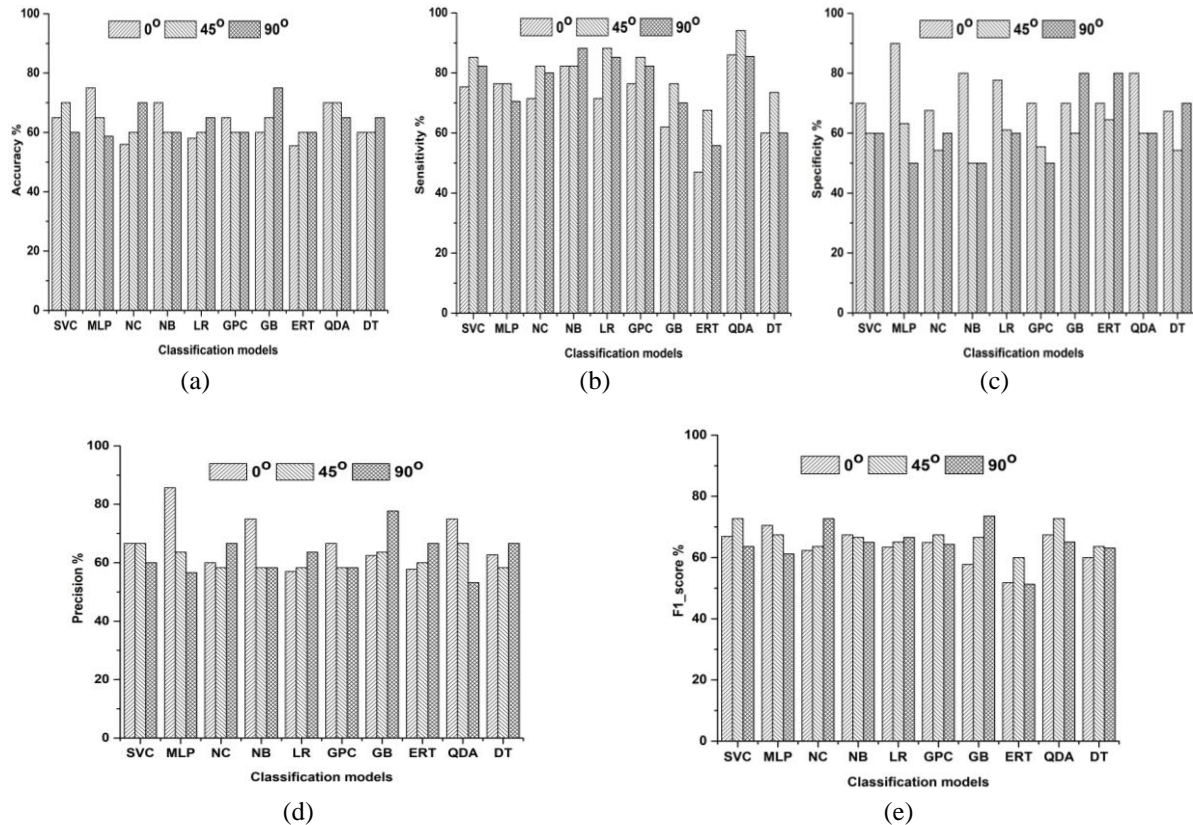


Figure 4. Algorithms performance utilizing 100 images at 0.2 wavenumber with various Gabor orientations (a) accuracy, (b) sensitivity, (c) specificity, (d) precision, and (e) F1\_score

## 6. CONCLUSIONS

Detection of breast cancer is critical to improving survival rates; CAD can increase prediction accuracy. Because of this, researchers developed methods to help doctors make accurate diagnoses and reduce human errors. This study evaluates ML systems for classifying IDC Breast Cancer Histopathology Images using five criteria; the models employ the Gabor filter to extract features. This study aims to examine the Gabor filter wavenumbers' effect on ML models' behavior. In most situations, the highest results were obtained with a wavenumber of 0.2. The second objective is to examine the effect of the number of images on the effectiveness of the models. As the number of images increased, their effectiveness decreased, leading to inefficiency. We suggest using deep learning instead of machine learning models for large datasets. The third objective was to evaluate the effects of different Gabor filter angle values on the behavior of the models, and by measuring the F1\_score, which indicates the harmonic rate of sensitivity and precision, we observed that the majority of the highest results occur at an angle of 45°.

## ACKNOWLEDGEMENTS

The authors would like to thank Mustansiriyah university for their valuable support and for providing the essential facilities for this research.

## REFERENCES

- [1] A. Kumar *et al.*, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, pp. 405–421, Jan. 2020, doi: 10.1016/j.ins.2019.08.072.
- [2] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 605–617, 2019, doi: 10.1007/s10278-019-00182-7.
- [3] R. R. Kadhim and M. Y. Kamil, "Comparison of breast cancer classification models on Wisconsin dataset," *International Journal of Reconfigurable and Embedded Systems (IJRES)*, vol. 11, no. 2, p. 166, Jul. 2022, doi: 10.11591/ijres.v11.i2.pp166-174.
- [4] Shalu and R. Mehra, "Breast cancer histology images classification: training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, Dec. 2018, doi: 10.1016/j.ict.2018.10.007.






- [5] B. N. Narayanan, V. Krishnaraja, and R. Ali, "Convolutional neural network for classification of histopathology images for breast cancer detection," in *2019 IEEE National Aerospace and Electronics Conference (NAECON)*, Jul. 2019, vol. 2019-July, pp. 291–295, doi: 10.1109/NAECON46414.2019.9058279.
- [6] F. T. Johra and M. M. H. Shuvo, "Detection of breast cancer from histopathology image and classifying benign and malignant state using fuzzy logic," *2016 3rd International Conference on Electrical Engineering and Information and Communication Technology, iCEEICT 2016*, 2017, doi: 10.1109/CEEICT.2016.7873137.
- [7] M. Y. Kamil, "Morphological gradient in brain magnetic resonance imaging based on intuitionistic fuzzy approach," in *2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*, May 2016, pp. 1–3, doi: 10.1109/AIC-MITCSA.2016.7759924.
- [8] A. H. Farhan and M. Y. Kamil, "Texture analysis of mammogram using local binary pattern method," *Journal of Physics: Conference Series*, vol. 1530, no. 1, p. 012091, May 2020, doi: 10.1088/1742-6596/1530/1/012091.
- [9] A. H. Farhan and M. Y. Kamil, "Texture analysis of breast cancer via LBP, HOG, and GLCM techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 7, p. 072098, Nov. 2020, doi: 10.1088/1757-899X/928/7/072098.
- [10] M. Y. Kamil and A.-L. A. Jassam, "Analysis of tissue abnormality in mammography images using gray level Co-occurrence matrix method," *Journal of Physics: Conference Series*, vol. 1530, no. 1, p. 012101, May 2020, doi: 10.1088/1742-6596/1530/1/012101.
- [11] Ragab, Sharkas, and Attallah, "Breast cancer diagnosis using an efficient CAD system based on multiple classifiers," *Diagnostics*, vol. 9, no. 4, p. 165, Oct. 2019, doi: 10.3390/diagnostics9040165.
- [12] A. Kamra, V. K. Jain, and S. Singh, "Extraction of orientation field using gabor filter and gradient based approach for the detection of subtle signs in mammograms," *Journal of Medical Imaging and Health Informatics*, vol. 4, no. 3, pp. 374–381, Jun. 2014, doi: 10.1166/jmihi.2014.1266.
- [13] H. Bolhasani, E. Amjadi, M. Tabatabaeian, and S. J. Jassbi, "A histopathological image dataset for grading breast invasive ductal carcinomas," *Informatics in Medicine Unlocked*, vol. 19, p. 100341, 2020, doi: 10.1016/j.imu.2020.100341.
- [14] D. Kumar and U. Batra, "An ensemble algorithm for breast cancer histopathology image classification," *Journal of Statistics and Management Systems*, vol. 23, no. 7, pp. 1187–1198, Oct. 2020, doi: 10.1080/09720510.2020.1818451.
- [15] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. J. Aguirre, and A. M. Vanegas, "Breast cancer histopathology image classification using an ensemble of deep learning models," *Sensors*, vol. 20, no. 16, p. 4373, Aug. 2020, doi: 10.3390/s20164373.
- [16] D. Kumar and U. Batra, "Classification of invasive ductal carcinoma from histopathology breast cancer images using stacked generalized ensemble," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 3, pp. 4919–4934, 2021, doi: 10.3233/JIFS-201702.
- [17] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of Pathology Informatics*, vol. 7, no. 1, p. 29, Jan. 2016, doi: 10.4103/2153-3539.186902.
- [18] R. Kadhim and M. Kamil, "Evaluation of machine learning models for breast cancer diagnosis via histogram of oriented gradients method and histopathology images," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 4, pp. 36-42, 2022, doi: 10.17762/ijritcc.v10i4.5532.
- [19] B. Ameer, S. Masmoudi, A. G. Derbel, and A. Ben Hamida, "Fusing Gabor and LBP feature sets for KNN and SRC-based face recognition," *2nd International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2016*, pp. 453–458, 2016, doi: 10.1109/ATSIP.2016.7523134.
- [20] S. Khan, M. Hussain, H. Aboalsamh, H. Mathkour, G. Bebis, and M. Zakariah, "Optimized Gabor features for mass classification in mammography," *Applied Soft Computing Journal*, vol. 44, pp. 267–280, 2016, doi: 10.1016/j.asoc.2016.04.012.
- [21] A. Kafuo, S. Dliaf, A. Gonifeda, Z. Alshatouri, and A. Baba, "A Literature survey of Gabor filter and its application," *Technical report*, University of Turkish Aeronautical Association, Institute of Science And Technology Electrical and Electronics Engineering Master Program (Without Thesis), 2017, doi: 10.13140/RG.2.2.11079.50085.
- [22] Z. Lei, S. Z. Li, R. Chu, and X. Zhu, "Face recognition with local Gabor textures," in *Advances in Biometrics*, vol. 4642 LNCS, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 49–57.
- [23] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Computing and Applications*, vol. 23, no. 7–8, pp. 2387–2403, Dec. 2013, doi: 10.1007/s00521-012-1196-7.
- [24] A. Tharwat, "Linear vs. quadratic discriminant analysis classifier: a tutorial," *International Journal of Applied Pattern Recognition*, vol. 3, no. 2, p. 145, 2016, doi: 10.1504/IJAPR.2016.079050.
- [25] G. Louppe, L. Wehenkel, A. Suter, and P. Geurts, "Understanding variable importances in Forests of randomized trees," *Advances in Neural Information Processing Systems*, vol. 26, 2013.
- [26] C. Bowd *et al.*, "Gradient-boosting classifiers combining vessel density and tissue thickness measurements for classifying early to moderate glaucoma," *American Journal of Ophthalmology*, vol. 217, pp. 131–139, Sep. 2020, doi: 10.1016/j.ajo.2020.03.024.
- [27] Y. Liu *et al.*, "Symptom severity classification with gradient tree boosting," *Journal of Biomedical Informatics*, vol. 75, pp. S105–S111, Nov. 2017, doi: 10.1016/j.jbi.2017.05.015.
- [28] Y. Li, S. Liu, and L. Shu, "Wind turbine fault diagnosis based on Gaussian process classifiers applied to operational data," *Renewable Energy*, vol. 134, pp. 357–366, Apr. 2019, doi: 10.1016/j.renene.2018.10.088.
- [29] J. Wang and C. Zhao, "A probabilistic framework with concurrent analytics of Gaussian process regression and classification for multivariate control performance assessment," *Journal of Process Control*, vol. 101, pp. 78–92, May 2021, doi: 10.1016/j.jprocont.2021.03.007.
- [30] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic regression in clinical studies," *International Journal of Radiation Oncology\*Biophysics\*Physics*, vol. 112, no. 2, pp. 271–277, Feb. 2022, doi: 10.1016/j.ijrobp.2021.08.007.
- [31] Z. Khandezamin, M. Naderan, and M. J. Rashti, "Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier," *Journal of Biomedical Informatics*, vol. 111, p. 103591, Nov. 2020, doi: 10.1016/j.jbi.2020.103591.
- [32] J. Kolluri and S. Razia, "Text classification using Naïve Bayes classifier," *Materials Today: Proceedings*, Nov. 2020, doi: 10.1016/j.matpr.2020.10.058.
- [33] M. Karabatak, "A new classifier for breast cancer detection based on Naïve Bayesian," *Measurement*, vol. 72, pp. 32–36, Aug. 2015, doi: 10.1016/j.measurement.2015.04.028.
- [34] E. N. Tamatjita and A. W. Mahastama, "Comparison of music genre classification using nearest centroid classifier and k-Nearest Neighbours," in *2016 International Conference on Information Management and Technology (ICIMTech)*, Nov. 2016, pp. 118–123, doi: 10.1109/ICIMTech.2016.7930314.
- [35] G. Daqi and J. Yan, "Classification methodologies of multilayer perceptrons with sigmoid activation functions," *Pattern Recognition*, vol. 38, no. 10, pp. 1469–1482, Oct. 2005, doi: 10.1016/j.patcog.2005.03.024.
- [36] H. Taud and J. F. Mas, "Multilayer perceptron (MLP)," in *Geomatic approaches for modeling land change scenarios*, 2018, pp. 451–455.




- [37] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement*, vol. 146, pp. 800–805, Nov. 2019, doi: 10.1016/j.measurement.2019.05.083.

## BIOGRAPHIES OF AUTHORS



**Rania R. Kadhim**    holds a Master of Science (M.Sc.) in physics from Mustansiriyah university, college of science, in 2016. She is currently a Ph.D. student in digital image processing at Mustansiriyah university in Iraq. Her research areas of interest include medical image processing, machine learning, computer vision, and image classification. She can be contacted at email: r90r61@gmail.com.



**Mohammed Y. Kamil**    holds a Master of Science (M.Sc.) in optics from Mustansiriyah university, in 2005. and Ph.D. in digital image processing from Mustansiriyah university, Iraq, in 2011, besides several professional certificates and skills. He is currently a professor with the physics department at Mustansiriyah university, Baghdad, Iraq. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and an editor in the Iraqi Journal of Physics. His research areas of interest include medical image processing, computer vision, and Artificial Intelligence. He can be contacted at email: m80y98@uomustansiriyah.edu.iq.