Mammography images classification system based texture analysis and multi class support vector machine

Cite as: AIP Conference Proceedings **2475**, 070003 (2023); https://doi.org/10.1063/5.0110733
Published Online: 31 March 2023

Ahmed Khalid Abdullah, Raghad Majeed Azawi, Ibrahim Tareq Ibrahim, et al.

Mammography Images Classification System Based Texture Analysis and Multi Class Support Vector Machine

Ahmed Khalid Abdullah[1, a)] Raghad Majeed Azawi[2, b)],Ibrahim Tareq Ibrahim[1, c)] and Asmaa Abbas Ajwad[1, d)]

[1]Department of Physiology, College of Medicine, University of Diyala, Diyala, Iraq.
[2]Department of Computing Science, College of Medicine, University of Diyala, Diyala, Iraq.

[a)] Corresponding Author: ahmadkh1979@yahoo.com
[b)] raghadm001@gmail.com
[c)] Ibr_33@yahoo.com
[d)] asmaamedeng@gmail.com

Abstract. Breast cancer is the greatest common reason of loss women in the world and the additional important cause of cancer losses world-wide. Classification and Detection of breast cancer are very significant since it offers body information of abnormal and normal soft tissue which supports in primary treatment planning and patient's situation follow-up, which is critical for woman's excellence in her life. X-ray mammography is the chief check used within quick diagnosis and screening, mammography is using in the medical imaging, and its exploration and processing are thesolutions for improving this tumor or cancer prognosis, several computer_aided finding structures have been advanced to provide support radiologists and internists for their diagnosis. In this article, an method is proposed to efficiently analyze digital mammograms based on texture segmentation to the detection for first stage tumors and there are a number of methods for medical image classification. The proposed algorithm was Multi Class Support Vector Machine and system accuracy of (98%).

Keywords. Breast Cancer, Digital Mammography, Texture Analysis, Gray Level Co_occurrence Matrex (GLCM).

INTRODUCTION

Breast cancer is an uncontrolled growth and an abnormal change of breast tissue due to abnormal cell divisions (proliferation). It is characterized via malignant growths for tumor that get up for the glandular tissue within the breast [1]. There are several types of breast cancer which can be classified by the physical characteristics of the tumour and location within the breast. Research illustrates that screening mammography using digital mammography images reduced the mortality rate from breast cancer by approximately 1,300 per year in the UK due to the early detection of cancer [2]. The ability of image readers to detect and diagnosis the small breast masses are reliant on upon the making for image high-quality. Digital mammography is the present regular imaging method of the early discovery of breast cancer. In medical image analysis methods have played a main part in various medical applications. In general, applications contain the extraction of features from medical image that are subsequently utilized for a variability of classification jobs, such as the difference between normal and abnormal regions.

Depending on a classification job, extracted features may be colour properties, shaped characteristics, or some formative image characteristics [3]. In mammography images several studied trainings have attempted to discover breast lesions in digital images such as Thamarai S. in 2009 [4], and Fatima E. et al., in 2011 [5]. However, no

1st International & 4th Local Conference For Pure Science (ICPS-2021)
AIP Conf. Proc. 2475, 070003-1–070003-10; https://doi.org/10.1063/5.0110733
Published by AIP Publishing. 978-0-7354-4327-3/$30.00

robust empirical study has been published to detect breast masses in digital mammography images mathematically. Thamarai S. in 2009 [4] designed a technique for breast tumour detection. This proposed technique consists of enhancement image method for removing noise and Gary Level Co-occurrence Matrix (GLCM) for segmentation technique and feature extraction stage.

The last stage is Support Vector Machine (SVM) is utilized for classification of the image into abnormal tumours and normal images. The methodology achieved a sensitivity of 88.75%. Fatima E. et al., in 2011 [5] exploited the capability of SVM classification based on Haralick Vector algorithm for classification breast lesion in mammography images to either malignant or benign. Fibro glandular region segmentation using the maxima method was utilized for segmentation stage. Then, co-occurrence matrix (Texture analysis) was utilized for feature extraction stage of segmented image. Finally, SVM classification based on Haralick vector is used for classification stage of breast tumours. The overall accuracy of the original mammograms is 77% in in average and 95% in case of classification using SVM algorithm. Prannoy G. and K. Sarava, in 2017 [6] designed a system for breast cancer detection utilizing image processing techniques. The suggested technique is consisting for several steps. Step one is Pre-processing utilizing removal noise algorithm.

After that, they are adjusted and filtered pixel for increase intensity. Step two, segmentation by the best ROI algorithm. Step three, feature extraction by using edge detection and texture features. Step four, Neural Network (ANN) to classification the dataset. The current method segments the breast mass utilizing (ROI) and classification of the breast masses by utilizing the ANN classifier. V. Hariraj et al., in 2018 [7] designed a system for define and detect lesion kind in mammography breast images. The proposed method contains of several stages. The pre-processing stage, the X-ray image to noise removal (filtering). Stage two, segmentation using K-means clustering. Step three, feature extraction by (ROI). The final Step, a proposed algorithm to classification, Fuzzy Multi-Layer SVM Classifier. The accuracy of testing in FMLSVM is 98%. Youcef Gh. et al., in 2019 [8] this study, Authors describe a novel mammogram classification framework for classifying breast tissues as normal, benign or malignant. First, the images obtained by this technique are processed with an anisotropic diffusion filter to reduce the noise and preserve edges. Then, a feature matrix is generated using: on the one hand, textural features from both a Gray Level Co-occurrence Matrix (GLCM) and a Gray Level Run Lengths Matrix (GLRLM) applied to all the filtered regions of interest (ROI) of a mammogram. Then The key point of the proposal is the modeling of Back Propagation Neural Network (BPNN) by a network of queues for representing the textural and morphological properties of the masses. The relevant and ordered features are injected in BPNN classifier using queuing network. The accuracy of testing system is 98.1% for normal cases and 95.2% for abnormal cases.

THE PROPOSED SYSTEM

Our work is presenting a proposed automated method which can be used to classify the mammography images into two group's abnormal and normal images. The proposed method has four stages. As follows, first stage image pre-processing techniques such as image enhancement and removal of noise. Second stage, feature extraction by texture Analysis. Third stage, SVM training algorithm. In the last stage, classification of mammography images by (SVM classifier).

1. Preprocessing stage using a median filter and Histogram equalization.
2. Feature analysis stage using Haar Wavelet and GLCM.
3. SVM training stage.
4. SVM testing stage and classification stage utilizing algorithm for MCSVM. Block diagram for proposedsystem is demonstrated in Fig. (1).

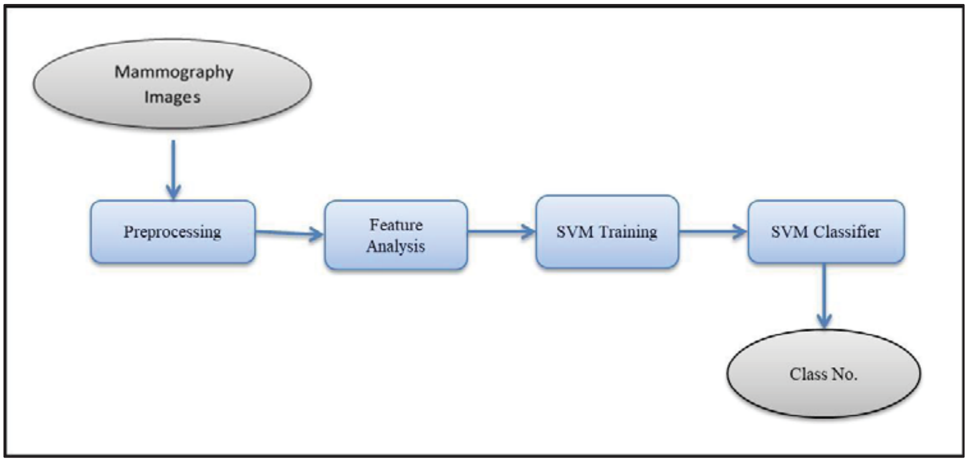
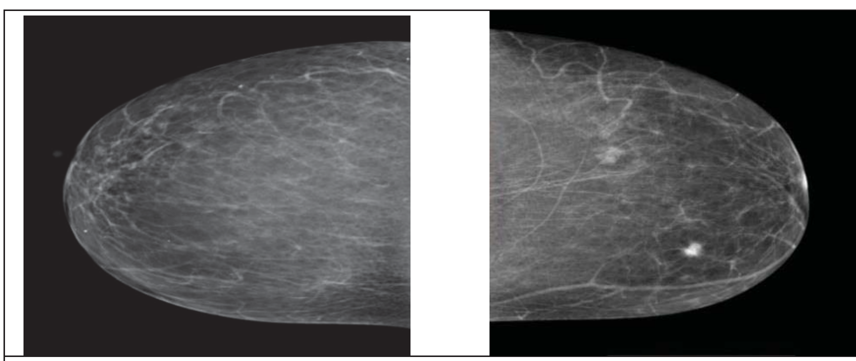


FIGURE 1. Block diagram for proposed breast classification system.

IMAGE ACQUISITION STAGE

Medical image acquisition is very significant to the diagnosis of disease. For the application of the proposed classification and detection system of breast cancer using mammography images, a dataset was collected from some sources for different characteristics of breast masses. Dataset has been collected from the University of Salfard [9]. Two groups of digital mammography images were used in this study (the normal case and abnormal cases). Abnormal cases were biopsy proven. In total, 140 mammography images (70 containing malignant breast masses and 70 normal images) were used in this study. Figure 2 demonstrates the samples of these types of mammography images. The feature of have been processed to be 8 bit (pixel value from 0 to 255) and the type of them is BMP with different sizes of images and it's colored or gray level.



Class One: Normal.	Class Two: Abnormal.

FIGURE 2. Samples of Normal and Abnormal for digital mammography images.

PREPROCESSING STAGE

The images have been processed during the processing stage and analysis which performs reduction of noise and enhancement techniques to enhance the image quality. Median filter is utilized for decrease "the salt and pepper noise". It is done for smoothening for mammography images. Here we are utilizing in mammography image (3x3 filters) to eliminate noise [3]. Figure (3) shows after applying 3 x 3 median filter.

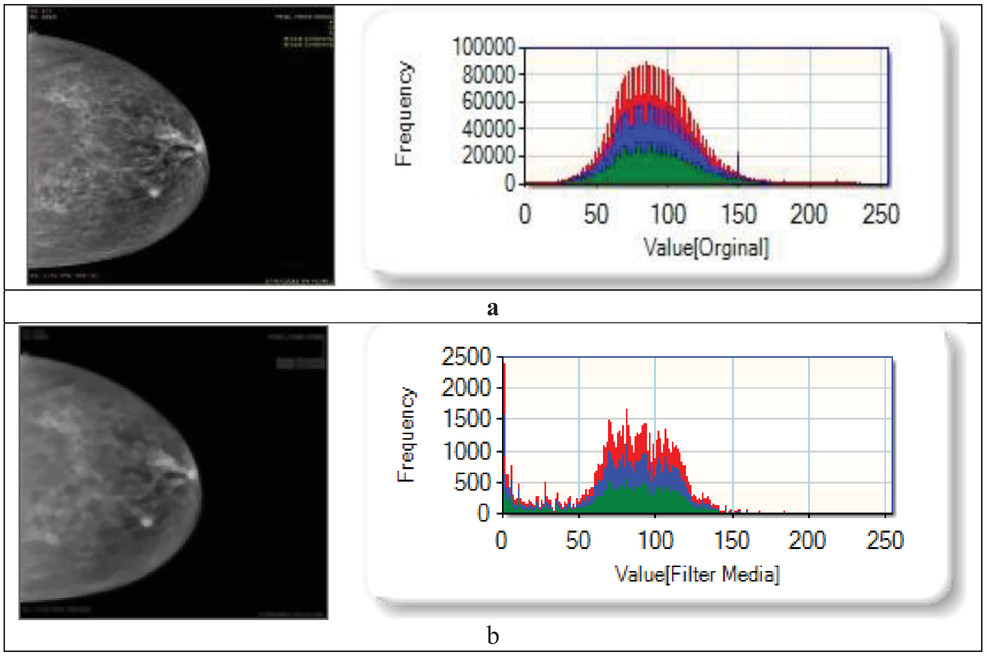


FIGURE 3. After Applying 3 x 3Median Filter (a) Input Mammography and (b) Median Filter.

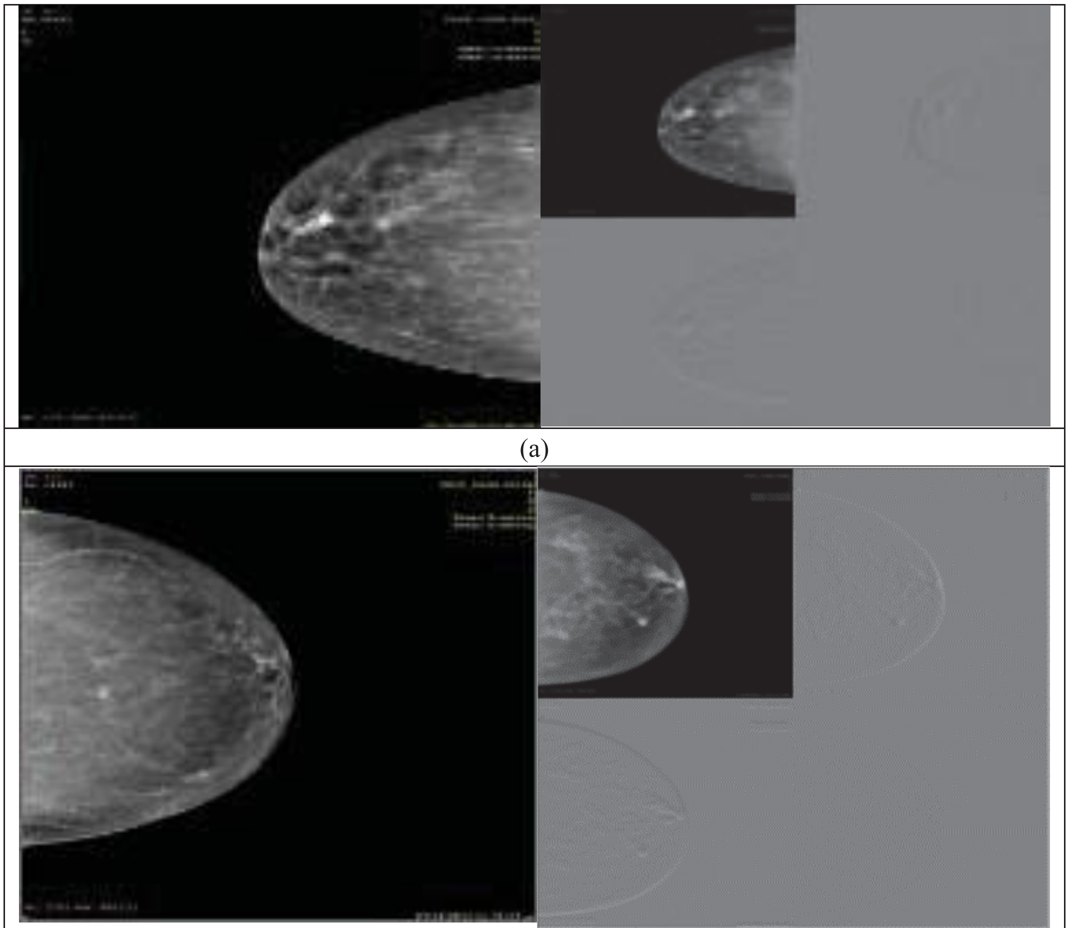
ADAPTIVE HISTOGRAM EQUALIZATION (HE)

HE method can be realized within various techniques. Within single perception the grey-levels (GL's) histogram in a window round every pixel is produced first. Increasing supply for grey levels, that is the sum above the histogram equalization, is utilized for plot input-pixel grey levels for GL's output . When the pixel in the gray levelis lower than all other elements within near window, this result is blacked and when median value within its window the output is 50% grey.

This part continues to a brief mathematical description for adaptive histogram equalization which can becirculated freely, and after that reflects the two chief kinds for adjustment. The relationship amid the equations and different (conceptual) perspectives on adaptive histogram equalization, such as GLs evaluation, might not to be instantly clear, but generalities can be articulated far more readily in this outline or framework [10].

FEATURES EXTRACTION STAGE

It is a challenging mission to extract fine features fixed for classification. The purpose of feature extraction is to decrease the original dataset by computing the features or the positive properties, which distinguish one input pattern from another. There are several feature extraction techniques but in this section, the texture-based ones can be most effective for classifying the medical images. There are several texture-based feature extraction techniques but Gray Level Co-occurrence Matrix (GLCM) is highly common and effective. In this method one Level Discrete Wavelet transform (Haar Wavelet) is firstly utilized to decompose input image into 4 images and then GLCM process is applied on each image. As a result, there are four sub-band (Low Low, Low High, High High, and High Low) images by haar wavelet at each scale. For feature extraction, only the four sub-bands are used for (DWT) decomposition at this scale then feature extraction based on (GLCM) as shown in Figure (4) images decomposition by Haar Wavelet [11].



(a)

(b)

FIGURE 4. Images Decomposition by Haar Wavelet (a) Original image and (b) 1 level of Haar Wavelet.

GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM)

It was utilized for feature extraction from digital mammography image. A feature of the image based on pixels and its neighboring pixels are extracted from image GLCM matrix is build comprises the textural based on two-pixel intensity values in the matrix. Feature-based on pixel and its neighboring pixel is extracted by GLCM (i, j) matrix. GLCM is a two-dimensional function, composed of vertical direction pixels m and of horizontal direction pixels n. The vertical and horizontal coordinates of the image is given by i, j. $0 \leq i \leq n \leq j \leq m$ where total pixel number is $m \times n$. First, pixel intensity and its neighboring pixel are computed for the whole parts of the image [12]. This technique decreases the calculation complexity. After computation for GLCMs of 4 sub-bands images, it is utilized to compute the image features which distinctively describe the images as demonstrated in figure (5) the overview of SVM classifier system. Table (1) shows an example of feature extraction stage for digital mammography image which contains breast masses.

$$\mathrm{Energy} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P^2(i,j) \tag{1}$$

$$\text{Entropy} = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j)\, log_2\, (P_d\,(i,j)) \tag{2}$$

$$\text{Contrast} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j) * (i-j)^2 \tag{3}$$

$$\text{Homogeneity} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{P(i,\ j)}{1+(i-j)^2} \tag{4}$$

$$\text{Variance (v)} = \sum_{i=0}^{Ng-1} \sum_{j}^{Ng-1} (i - \mu x)^2 p(i,j) \tag{5}$$

$$\text{Dissimilarity} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} |i-j| \times P\ (i,j) \tag{6}$$

$$\text{Maximum Probability} = max\ \{\ p(i,j)\} \tag{7}$$

$$\text{Correlation} = \frac{\sum_{i,j}[(ij)\ p(ij)] - \mu x\ \mu y}{\sigma x\ \sigma y} \tag{8}$$

Where "μ_x, μ_y, σ_x and σ_y are the means and standard deviations" in p_x and p_y

$$\mu x = \sum_{i=0}^{ng-1} i \sum_{j=0}^{ng-1} p(i,j) \tag{9}$$

$$\mu x = \sum_{j=0}^{ng-1} j \sum_{j=0}^{ng-1} p(i,j) \tag{10}$$

$$\sigma_x = \sum_a (a-\mu x)^2 \sum_b p(a-b) \tag{11}$$

$$\sigma_y = \sum_b (b-\mu y)^2 \sum_a p(a-b) \tag{12}$$

TABLE 1. Example of feature extraction stage for digital mammography image which contains breast masses.

Image	Energy	Entropy	Contrast	Homogeneity	Variance	Dissimilarity	Max	Corr.
LL	3.852	0.288	43.612	0.6554	114.758	2.3669	25563	0.649
HL	1.059	0.372	90.548	0.2547	548.894	1.8871	10554	0.895
LH	1.564	0.477	116.77	0.7738	549.475	1.2668	15558	0.775
HH	0.998	0.5436	118.1257	0.8456	801.5776	1.9967	15896	0.843

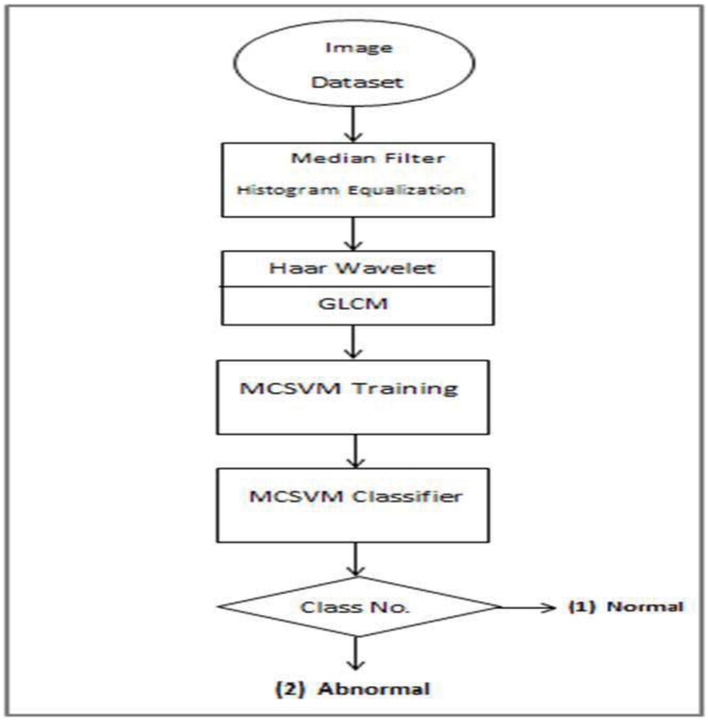


FIGURE 5. Overview of SVM Classifier System.

CLASSIFICATION STAGE BY MULTI CLASS SUPPORT VECTOR MACHINE

Information of texture is extracted from the Haralick vector computed from the co-occurrence matrix (GLCM). Haralick vector computation is based on the choice of the analysis the displacement size (d), window size, and the direction depending on which the two selected pixels are compared. It is recognized that SVM-based classification is highly dependent on the training phase, displacement orientation values and window size utilized to compute training vectors [5]. Support Vector Machines (SVM) has been advanced from Vapnik (1998). SVM is a statistical learning theory and supervised learning algorithm . In this paper, the error has been decreased from the training dataset. It constructs a classifier that splits training data samples and it maximizes the minimum margin or distance. In theory, SVM has capable to effectively classify every linearly discrete data. Consider this data through two classes [13, 14].

There are several techniques to expand dual class SVM problem into multi-class SVM problem, such as -one versus one- , -one versus rest- or -one versus all-, error corrected output coding (ECOC), and directed acyclic graph (DAG) [12]. This work, the one versus one technique is utilized. It converts multi-class problem into ½ (k (k-1)) binary classification problems. In another meaning, each probable pair's class is considered. On each binary classification problem, the training set is in the next step.

SVM Training: Train a SVM classifier with the SVM train function. The most common syntax is

SVM Struct= SVM train (data, groups, "kernel - function", RBF)

Where: data is Matrix of data points, where each column is one feature.

Groups: Column vector with each row corresponding to the value of the corresponding row in data. Groups must have only two kinds of entries. Groups can, therefore, have logical entries, or they can be a double vector or a two-valued cell array.

SVM Classification: Support vectors are the data points that located nearest to the decision surface. The classification of support vectors is considered the most difficult. They have high impact on the optimum location of the decision surface. We can demonstrate that the optimal hyper plane stems from the function class with the lowest capacity (VC dimension). The boundary about the separating hyper plane can be maximized by Support vector machines. The decision function is completely determined by a training samples subset, Quadratic programming problem and the support vectors [15].

RESULTS AND DISCUSSION

The results that obtained from the classification system of digital mammography images have been discussed in this section. The technique was implemented on mammography images data set (these are class one: normal (70 images) and class two: abnormal (70 images). Experiments are carried out in Visual Basic.Net.2013 and an Intel Core i7, 2.20-GHz processor , 64-bit OS and RAM in 4-GB in this paper,an algorithm described is developed and successfully trained for utilizing a mixture from image processing and SVM tool-box with 8 texture features. The remaining 50% of mammography image for different kinds utilized such as analysis data stage. The end result signifies that 50% images from total are classified correctly. The classification system is a mammography images with (8) GLCM features and the direction = (0˚) and system accuracy of (98 %). A proposed classification scheme is valued by utilizing classification classified level which is the rate of correct classified image to the total classified breast surgery image, Sensitivity (TPR), Specificity (FPR). Table (2) represents the classification results for the system and figure (6) illustrates the classification rate of the system and computation of Specificity and Sensitivity.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{13}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{14}$$

$$\text{Classification Rate} = \frac{TP+TN}{TP+FP+TN+FN} \tag{15}$$

TABLE 2. Classification Results of System.

Classifier	Rate
Sensitivity	1
Specificty	0.98
Classification Rate	0.99

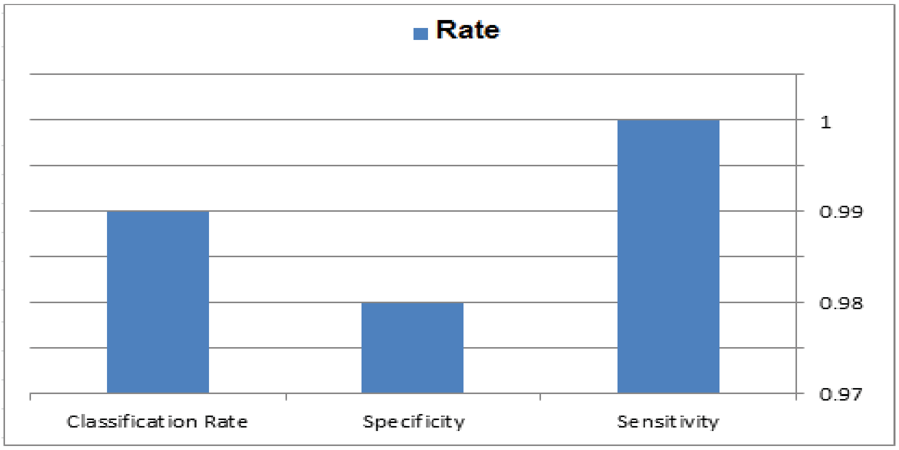


FIGURE 6. Classification Rate, Specificity and Sensitivity of the System.

Generally, digital mammography is the process of converting the analogue data to a digital signal is called digitisation. This process includes two stages: sampling and quantization. The analogue data is produced by a detector initially, and then the detector converts it to digital data using special electronic circuit named analogue-to-digital convertors (ADCs). Then, the digital signal transfers to a computer in order to digitize and store it digitally. In digital mammography, the last step in the process of image formation is the image processing application. Image processing is utilised to improve the diagnostic information within the image. This can be achieved by improving the contrast and edge definition or through decreasing image noise. Every manufacturer utilises their own image processing algorithms, and the appearances of the image after this processing can vary between manufacturers. Many studies have investigated the impact of these variations in image appearance. Some studies concluded that the difference between several image-processing algorithms was significant and others found that the difference was not significant. The standard database is used to validate the proposed scheme. However, we have noticed that the use of features improves performance by a great percentage and high accuracy. However, the classification method by SVM algorithm outperforms relief and is one from the point of view of accuracy in the performance of the system. These measures are sensitivity 1 , specificity 0.98 for normal versus abnormal classification and classification rate 0.99 for all data in the system.

When comparing results with previous studies and this paper as shown in the table (3) this proposed system proved its effectiveness to classify mamoghraphy breast tumors type by applying 8 features of (GLCM) and (MCSVM) which also proved to be effective and the classification rate of the testing phase is 98 % . This proposed system will give fast and more accurate and help physician for proper treatment, and use MCSVM gives the system more power because it needs a very short time for training and representing a good accuracy classifier.

TABLE 3. Comparing with preview stadies.

Reference	Algorithm	System Classification Rate
Thamarai S. [4]	SVM	88.75%
Fatima E. et al. [5]	SVM	95%
Prannoy and Sarava [6]	ANN	90%
Proposed System	MCSVM	98%

CONCLUSION

The computer-based diagnosis is important for correctly classification and detection of breast cancer in digital mammography images. The new technique is a combination of texture analysis and Multi Class Support Vector Machine Classifier (MCSVM). The use of this algorithm has led to good results in the classification and detection of the disease. By utilizing this algorithm, an efficient breast lesions classification technique has been constructed with a maximum system accuracy of 98 % in direction 0˚ and the incorrect rating is about 2% from the total dataset. This method could serve accurate classification of beast lesions detection. The algorithm of MCSVM used for validation of the proposed scheme. However, the classfication method outperformes relief and is one form the point of view of accuracy in the performance of the system.

REFERENCES

1. Teresa, A. Guilherme, C. Eduardo, R. José , A. Paulo , E. Catarina, P. António and C. Aurélio, Classification of breast cancer histology images using Convolutional Neural Networks, PLOS ONE.. 1, June, 2017.
2. Cancer research UK, "Breast cancer mortality statistics, 2016. [Online]. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics.
3. R. Majeed, A. Dhahir, M. Jamal and I. Tariq, Brain Tumors Classification by Using Gray Level Co- occurrence Matrix, Genetic Algorithm and Probabilistic Neural Network, Diyala Journal of Medicine, Vol. 14, No. 2, June, 2018.
4. S. Thamarai, Early Detection Of Breast Cancer Using SVM Classifier Technique, Journal on Computer Science and Engineering, Vol. 1, No. 3, pp. 127-130, 2009.
5. E. Fatima, R. Fakhita, M. Abdelhak and L. Najib, Masses Detection Using SVM Classifier Based on Textures Analysis, Applied Mathematical Sciences, Vol. 5, No. 8, pp. 367 – 379, 2011.
6. G. Prannoy and K. Sarava, Breast Cancer Detection using Image Processing Techniques, Oriental Journal Of Computer Science & Technology, Vol. 10, 2017.
7. V. Hariraj, Z. Ibrahim, A. Shahriman, M. Zuradzman, T. Rajendran and R. Sathiyasheelan, Fuzzy Multi- Layer SVM Classification Of Breast Cancer Mammogram Images, International Journal of Mechanical Engineering and Technology (IJMET), Vol. 9, No. 8, 2018.
8. Y. Gherghout, Y. Tlili and L. Souici, Classification of breast mass in mammography using anisotropic diffusion filter by selecting and aggregating morphological and textural features. Evolving Systems, Vol. 12, No. 2, pp. 273-302, 2021.
9. K. Abdullah, The impact of simulated motion blur on breast cancer detection performance in full field digital mammography (FFDM), University of Salford, Vol. 21, 2018.
10. Fawzi, H., Mostafa, S. A., Ahmed, D., Alduais, N., Mohammed, M. A., & Elhoseny, M. (2021). TOQO: A new Tillage Operations Quality Optimization model based on parallel and dynamic Decision Support System. *Journal of Cleaner Production*, 128263.
11. R. Majeed, A. Dhahir, M. Jamal and I. Tariq, A Hybrid Approach for Classification of MRI Brain Tumors Using Genetic Algorithm, K-Nearest Neighbor and Probabilistic Neural Network, International Journal of Computer Science and Information Security (IJCSIS), Vol. 16, No. 5, pp. 1947-5500, 2018.
12. N. Dipanshu, N. Masalkar, and A. S. Shitole, Advance Method for Brain Tumor Classification, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 2, 2014.
13. Z. Rustam, I. Primasari, and D. Widya, Classification of cancer data based on support vectors machines with feature selection using genetic algorithm and Laplacian score, the American Institute of Physics, 2017.
14. Tariq, M. Sabah, Z. Anmar and E. AL-Mashat, The Evaluation of Calcium Score Validity in the Diagnosis of Patients with Coronary Artery Disease by Using CT Angiography", Diyala Journal of Medicine (DJM), Vol. 9, No. 1, 2015.
15. N. Vani, A. Sowmya and N. Jayamma, Brain Tumor Classification using Support Vector Machine", International Research Journal (IRJET), Vol. 04, No. 7, pp. 2395-0072, 2017.