Materials Today: Proceedings 37 (2021) 2738–2743

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Survey of machine learning algorithms for breast cancer detection using mammogram images

G. Meenalochini *, S. Ramkumar

School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 12 August 2020
Accepted 20 August 2020
Available online 9 October 2020

Keywords:
Breast cancer
Mammogram
Classification
Diagnosis
Screening Techniques

ABSTRACT

Breast cancer is the primary cause of death in most cancer affected women. Mammography is one of the most dependable strategies for early detection and diagnosis of breast cancer and reduces the death rate. Mammograms are radiographic images of the breast which are utilized to identify the early symptoms of breast cancer. These radiographic images reduce human errors in detecting cysts and reduce the diagnosing time and also increase the diagnosis accuracy. An overview of the machine learning techniques for breast cancer detection and classification has been presented in this paper, which can be divided into three main stages: pre-processing, extraction of features, and classification. This article discusses about the effects of several Machine learning techniques for automation of mammogram image classification are investigated. This investigation assembles agent works that show how Machine learning technique is applied to the result of different issues identified with various analytic science examinations. This study portrays the impacts of pre-taken care of mammogram images before entering the classifier, which brings about higher effective classification. The detection stage is trailed by segmentation of the tumor region in a mammogram image. This study is an attempt to gather and compare the various screening techniques, classifiers, and their performance in terms of sensitivity, specificity and exactness for breast cancer diagnosis.

© 2020 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Newer Trends and Innovation in Mechanical Engineering: Materials Science.

1. Introduction

Breast cancer growth is the most critical disease which causes demise in the female populace, increasingly over 2.1 million ladies inconsistently. WHO assessed that in the year 2018 there will be 627,000 ladies died from breast cancer. Its development broke down and almost 15% of all malignant growth passing's among women [1]. To reduce the death rate caused by breast cancer, we need early steps to detect and classify the tumors. The early location builds the odds of successful treatment, improving the illness forecast. There are various screening techniques available to detect breast cancer. Table 1 shows the review of screening techniques (Screening examinations are tests performed to find disease before symptoms begin). Out of these strategies, the Mammography assumes a significant job in dependable early identification and determination of breast cancer. Even though mammography is of incentive in screening ladies for breast malignant growth radiologists can contrast, its adequacy relies upon radiologists' understandings. Endeavors to improve exactness and diminish inconstancy in translation may build the adequacy of mammography in identifying early breast cancers [2].

Digital mammograms (FFDM) are recorded as images on a computer. Fig. 1 shows the Mammogram machine. A machine intended for mammograms to envision just at breast tissue. The machine takes x-beams at lower dosages than common x-beams. Since these x-beams don't experience tissue effectively, the machine has 2 plates that pack or straighten the breast to spread the tissue separated.

Mammogram recommended for the following reasons.

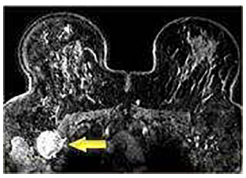
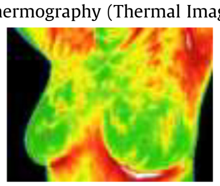
- Diagnosis of breast irregularities
- The follow-up to a previous abnormal mammogram
- Tracking the progress of lumps or irregularities

Types of Mammogram techniques are represented in Fig. 2.

* Corresponding author.
E-mail addresses: gmlochini@gmail.com (G. Meenalochini), ramkumar.drl2013@gmail.com (S. Ramkumar).

https://doi.org/10.1016/j.matpr.2020.08.543
2214-7853/© 2020 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Newer Trends and Innovation in Mechanical Engineering: Materials Science.

Table 1
Review of screening techniques.

Techniques	Procedure	Advantages	Disadvantages
Mammography 	X-ray examination for assessment and recognition of the breast variations from the norm.	Most Promising techniques used by radiologists frequently.	• Expose Radiation • Not suitable for dense breast (Women's age < 40)
Breast MRI 	Utilizations radio waves and a Magnetic field to change the modification of protons of hydrogen cores and produce extremely definite, cross-sectional images.	Best for the women's having high risk factor.	• Biopsies are recommended • Not suitable for patients with metallic devices
Breast Ultra-Sound(Sonography) 	Utilizes sound waves for the inner assessment of the body part. (Echo's are changed over into images)	• No Radiation • Suitable for women's having dense breasts. • Painless technology	• Difficult to cover the entire breast • Poor resolution
PET (Positron Emission Tomography) 	For a PET scan, Fluoro Deoxy Glucose (FDG) is the generally utilized radioactive substance that must infused into an arm vein to separate between tissues. The greatest measure of the radioactive sugar is consumed by the harmful tissues considered as the most active cells.	• Does not suffer from breast density, earlier surgery or radiation therapy.	• Cost expensive • Limited resolution • Low imaging speed
Thermography (Thermal Imaging) 	An infrared scan maps the variation in temperature over the breasts and afterward, the image is formed.	• noninvasive • having fast imaging time • suitable for dense breast	• Procedural errors • If the internal body temperature is asymmetric the outcomes of the measurement are FN and FP.

For the best classification outcome, a hybrid of classifiers will be more beneficial. This research work deals with the review of various classification techniques proposed in the literature review for breast mammogram images. Section II explains the review of various classifiers in terms of Data set, Classification techniques, Outcome, Features, Parameters, advantages, and disadvantages. Section III summarizes the conclusion.

In breast cancer diagnosis with classification techniques, the pre-processing stage is a step used to emphasize the anomalies that improve the features in mammogram images which has low contrast. Preprocessing achieves to increase the contrast, emphasizing suspicious region. The Enhancement operation is performed to displays detailed information from the invisible image. Enhancement is a supportive tool for diagnosis and yields better visual quality [3]. The extraction of ROI or the malignant from the original

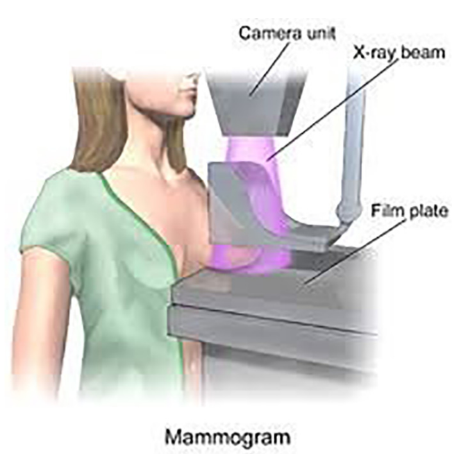


Fig. 1. Mammogram machine.

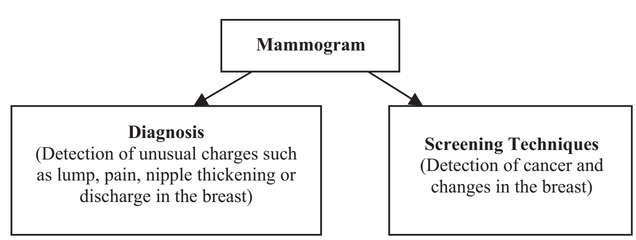


Fig. 2. Types of mammogram techniques.

image to make the diagnosis more accurate is known as segmentation. The segmentation algorithm can be classified as various types such as Region-based, Edge-based, Clustering, Thresholding. The required specific features (intensity, size, shape, texture) extracted and selected from the mammogram images. The process of classifications performed to classify the category of normal and abnormal (benign/malignant) from the mammogram image. These steps of classification are represented in Fig. 3.

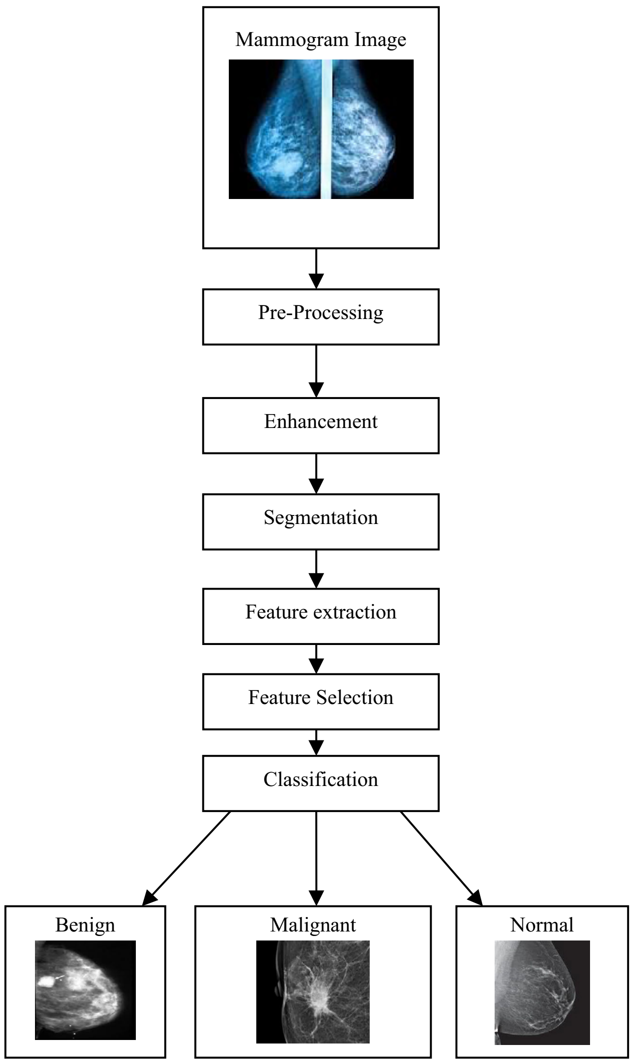


Fig. 3. Classification of mammogram image.

The classifier improves the classification performance and execution of mammography images for benign and malignant classification, yet in addition, gives radiologists accurate diagnosis results. Mammogram samples are taken the structure the predetermined database for all classes like harmful, generous, and typical. The element ought to be removed adequately and it ought to be isolated. After the component extraction from the mammogram image, the resultant parameters are given as a contribution to the classifier. There are innumerous classification methods for the automated classification of samples. Summary of various classifiers cons and pros is discussed in table 2.

2. Literature review

Several techniques have been a proposal to classify the mammogram images. To find the optimal technique for classification, various classification techniques are represented in Table 3 depending upon the methods to be used, data sets, features, techniques, outcomes, results, parameters, advantages and disadvantages. Most of the classification techniques use the Mini-MIAS database (Mammographic Image Analysis Society) to classify breast lesions. The fusion of classification techniques increases classification accuracy. To distinguish the breast lesions by the CAD system can be assisted by a colossal number of a few highlights and its perception causes the radiologists to abstain from missing breast anomalies [5]. It is very hard to differentiate the benign from malignant due to various variability associated with the mammographic appearance.

Kulkarni [6] (2019) has attempted to classify the mammographic lesions with Pixel N-gram features by using different classifiers (MLP, SVM and KNN) the performance is observed to discuss the result of increasing N. The final outcome was analyzed that the high performance is achieved by using MLP classifier was better than the performance using SVM or KNN classifiers. Ibrahim A.O. [7] et al. (2020) have developed a CAD system for breast diagnosis using Radial Basis Function Network (RBF) technique. The decision-making system classifies the tumors by RBF network classifiers. The proposed work compared RBF neural network with MLP algorithm. The overall performance of RBF neural network yields 79.166% accuracy and MLP algorithm was 54.1667%, which proved that the RBF neural network can successfully classify the mammo-

Table 2
Comparison of classifiers.

Classifier	Procedure	Advantages	Disadvantages
SVM	A Support Vector Machine is a binary linear classification whose decision boundaries explicitly constructed to minimize generalization error.	• Effective in high dimensional space • Effective in Situations Where the number of measurements exceeds the number of samples • Decision function • Versatile	• Control is important in cases where the number of features is much greater than the number of samples. • Do not provide projections of probability straight away • Not perform well when the data has more noise.
KNN	(K-Nearest Neighbors) KNN measures the distance to each sample from an unknown sequence and selects the nearest K samples as the classification basis.	• KNN yielded higher prediction accuracy than SVM [4] • Lazy learner (Instance-based earning) • New data can be added seamlessly. • easy to implement	• High risk with 1. large data set 2. high dimension • Need feature scaling • Sensitive to noise data, missing values and outliers
CNN	Convolutional Neural is a selected kind **of** artificial neural network that utilizations perceptrons, a machine learning unit algorithm for supervised learning to analyze the data.	Accuracy in Image recognition problems	• High computational cost • slow training rate at complex tasks • need a lot of training data

Table 3
Review of classification techniques.

Authors & year	Data set	Features used	Classification technique	Outcome	Parameter	Advantages	Dis advantages	Result
Prabh Kaur et al. (2019) [11]	Mini - MIAS data set of 322 images	Grey level Morphological features.	Multi class Support Vector Machine (MSVM) and K-means clustering / decision tree	To characterize typical, benevolent and dangerous classes of Mammographic masses.	Accuracies: SVM - 96.9% LDA - 93.8% KNN - 94.8% Decision tree - 89.7% 10-fold cross validation	Consistency of classification accuracy	Randomized image selection with limited data set.	The classification algorithm result gives that MSVM is superior to anything the decision tree.
M.ParisaBeham et al. (2019) [12]	149 Mammographic images from the pixel scan hub, Trichy.	Discriminant features such as LSP (Local Binary Pattern), geometric / texture features	K - nearest Classifier - K-NN	Highlights extricated from the wovlet normalized images have created better classification.	Classification accuracy: KNN - 71.14% (LBP - block size 12)	Distinguish variation from the norm of mammogram images in basic and effective way.	Limited prominent feature extraction techniques.	Well correlation with different standardization methods by changing K esteems and LBP block size
HomayoonYekaei et al. (2019) [13]	MIAS - 1024 samples of 322 Patients.	Both global and local features	MCNN (Multi Convolutional Neural Network)	Better classification rates than the classical state-of-the-art method. Detect the strong and accurate diagnosis of benign and malignant lumps.	Mean Accuracy 97 ±0.3% Sensitivity of detection 95.9% Specificity 94.8% Detection error 3% AUC - 0.88	Classical hand-crafted feature extraction step is avoided	Limited data set variety	Increases the layer thickness in a Multi scale CNN, the accuracy also increases.
Juan wang et al. (2018) [14]	521 SFM (Screen Film Mammogram) images, 188 FFDM (Full Field Digital Mammogram	Local image features	DNN Classifier	Classify a recognized cluster as being favorable or harmful (benign or malignant).	TPF - 80% FP - 1.03 FPS/images TPF - 85% FP - 0.40 clusters/ images	This Classifier considers both local image features and the surrounding image background.	Less features to be introduced.	Miniaturized scale arrangement recognition can be advantageous for decreasing the FPs in discovery.
S.AkilaAgens et al. (2019) [15]	322 digital images from Mini - MIAS Data set	Multi scale features	MACNN	Automatically categorizes the mammographic image in to normal, Malignant and benign classes.	AUC - 0.99 Overall sensitivity - 96%	Improves the precision of the classification framework by melding the more extensive setting of data utilizing Multi scale filters without arranging the calculation speed.	Probability of correct prediction	Exhibit the essentialness and viability of Multiple dilation Convolution CNN model with every single learnable layer.
Sura jasim Mohammed et al. (2018) [16]	100 images from MIAS	Statistical features	Multi Class SVM	Classifies the segmented, suspected regions and gives good accuracy rate with small misclassification ratios.	**Hold Out Method:** Accuracy - 0.9571% Sensitivity - 0.80% **One-leave out Method:** Accuracy - 0.9333% Sensitivity - 0.909%	Test Phase for evaluating the performance of the classifier.(Hold Out Method &One-leave out Method)	limited instances (examples) Set. Need high training rate.	Giving better estimation accuracy in less time consumed.
Arpana M.A, et al. (2014) [17]	MIAS data set of 322 images.	Statistical, texture, gradient features are extricated and clinical highlights are gotten straightforwardly from the dataset.	The neural network feed forward classifier	Isolating the tumor district as favorable, threatening or typical.	-	High performance of Feature extraction.	Lack of hybridized classifiers.	The breast locale removed by the introduced calculation roughly follows that separated by a specialist radiologist.
IoanBuciu et al. (2011) [18]	MIAS data set of 322 images.	The features / directional features of the neural network foGabPCA are extracted at	Proximal Support Vector Machine	GabPCAfeatures end up being progressively hearty against commotion contrasted with unadulterated PCA highlights andseem to have increasingly	Recognition rate - 84.37%, Sensitivity - 97.56% Specificity - 60.86% AUC - 0.79%	SVM was independently prepared (to get Its ideal isolating hyperplane)for each case.(GabPCA and PCA)	Performance depends on Low/high frequency image	The Robustness of Gabor highlights for computerized mammogram images contorted by Poisson

(*continued on next page*)

Table 3 (*continued*)

Authors & year	Data set	Features used	Classification technique	Outcome	Parameter	Advantages	Dis advantages	Result
		different orientations and rward classifier f requencies		iscriminative force.			components	clamor with various intensity levels
V Harvind Viswanath (2019) [19]	Raw Sample Images	ANOVA/ Statistical features	SVM, K-NN and RF supervised classifiers	To overcome the problem of low accuracy for breast cancer diagnosis produced by low contrast on acquired mammogram images	SVM: Accuracy - 84.84% Precision - 73% Specificity - 68% KNN: Accuracy - 86.99% Precision - 80% Specificity - 80% RF: Accuracy - 84.84% Precision - 90% Specificity - 89%	supports that the choice of a correct classifier influences enormously on the accurate diagnosis of mammograms.	Partial Automated system. Lack of Data set.	RF classifier reaches higher performance for classifying two and three mammogram conditions
Mohamed MeselhyEltoukhy et al. (2010) [20]	MIAS - Data set composed of 322 mammograms	Texture Features	nearest neighbor classifier	ready to locate a suitable list of capabilities that lead to huge improvement in characterization exactness.	5-fold cross approval. Effective characterization pace of mammogram pictures for typical and anomalous utilizing closest neighbor classifier came to 99.03% in fold 4.	Use of seven texture features in curvelet transform for every wedge.	More parameters to be calculated	utilizing curvelet based texture feature can improve the classification of mammogram.
M. Arfan Jaffar et al. (2009) [21]	MIAS of 322 images	Gray level - Texture highlights, for example, (mean, fluctuation, skewness, kurtosis, entropy, vitality, differentiation and homogeneity)	Support Vector Machine (SVM) and Multilayer Perceptrons (MLP)	Reduce the mammogram image successfully and yield high grading results	SVM: Accuracy - 96.781% Sensitivity - 8.662% Specificity -8.213% MLP: Accuracy - 94.119% Sensitivity -96.112% Specificity-95.413%	Increase interpretative performance through a number of computer-aided diagnoses (CAD)	Lack of Hybrid features.	The proposed calculation is predominant regarding affectability, explicitness, and precision.
YijieJin (2019)[22]	MIAS of 322 images	rudimentary features	binary classifier with (CNNI-BCC),	trained and evaluated several convolutional neural networks for mammogram classification and tumor detection.	The accuracy on the test set is 73.24% for 10 epochs	Automatically detect the tumorous lesion without prior information of the presence of a cancerous lesion.	Considered the accuracy but not the sensitivity and specificity.	VGG16 and SSD algorithms to fit for the dataset, and obtained mAP of 0.842 in tumor detection.
Aderonke Anthonia Kayode (2019) [23]	322 images from Mammographic Image Analysis Society database	Textural features (15 features)	SVM	A computerized classification system for mammograms utilizing altered SVM classification method.	Sensitivity-94.4% specificity - 91.3% positive predictive value - 89.5% negative predictive value system -95.5%	Help radiologists make the right and timely decision making, thereby improving their analytical abilities.	The significant variability leads to diagnostic errors.	The results of the performance evaluation suggested that the system could be used as a potential radiologist tool to support mammogram interpretation decision-making
Debelee, T. G., et al. (2019)[24]	The local image datasets are gathered from the Bethezata General Hospital (BGH) and the Mammographic Image Analysis Society (MIAS) open datasets	texture features (GLCM and Gabor)	SVM and MLP	The consistency of the features is assessed individually and fusing features to each other and using cross-validation five classifiers (SVM, KNN, MLP, RF, and NB)) are used to assess the expressive strength of the features.	For BGH and MIAS Database, SVM Accuracy - 99%, 97.46%, sensitivity - 99.48%, 96.26% specificity - 98.16%, 100% MLP Accuracy - 97%, 87.64% sensitivity - 97.40%, 96.65% specificity - 96.26%, 75.73%	Compared with the number of classifiers limit.	Lack of Fusion of texture features	Using the MLP classifier, maximum performance is achieved for function fusion between extracted features based on Gabor and CNN.

gram images with better classification accuracy. R.Vijayarajeswari, et al., [8] (2019) used Hough transform to detect mammogram image features. Those features are used as SVM classifier inputs. The SVM classifier achieved an accuracy range of 94% which exceeds the accuracy range obtained the LDA classifier (86%). Mughal et al., [9] (2019) constructed the classification model with back propagation neural network (BPNN). The system significantly diagnoses the tumor at the initial stage on MIAS and DDSM datasets with 99.0% of classification accuracy. N. Santhanalakshmi et al., [10] (2019) have implemented the breast cancer diagnosis with KNN classifier which mainly provide a distinction between the malignant and breast masses. The portion of the cancerous cells in the breast region is being detected from the input given. The KNN (K Nearest Neighbor) classifier has been trained using their extracted features and corresponding labels with every image from the breast image dataset. The overall accuracy obtained by the proposed method is almost >90%.

3. conclusion

The proposed study is an attempt to examine the recent breast cancer classification techniques using digital image processing and the result of reliability, performance, affordability and outcomes are analyzed. This research study overviewed the methodologies and procedures proposed for the classification of breast tumors. The Performance comparison of various classification techniques is reviewed. The hybrid techniques can be used to get better accuracy of classification. The integration of classifiers to reduce both FPs and FNs. Maintaining the balance between accuracy and time complexity facilitates the development of the best CAD systems for breast cancer diagnosis. This paper proposes an idea of classification techniques for the detection of breast tumors by comparing the properties and pros and cons of various classifiers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] World Health Organization, available at https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/ (accessed January 2020).
[2] M. Karabatak, M.C. Ince, An expert system for detection of breast cancer based on association rules and neural network, Expert Systems Appl. 36 (2 PART 2) (2009) 3465–3469, https://doi.org/10.1016/j.eswa.2008.02.064.
[3] S. Anand, S. Gayathri, Mammogram image enhancement by two-stage adaptive histogram equalization, Optik 126 (21) (2015) 3150–3152, https://doi.org/10.1016/j.ijleo.2015.07.069.
[4] M. Heidari, A.Z. Khuzani, A.B. Hollingsworth, G. Danala, S. Mirniaharikandehei, Y. Qiu, B. Zheng, Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm, Phys. Med. Biol. 63 (3) (2018), https://doi.org/10.1088/1361-6560/aaa1ca 035020.
[5] A. Alkhateeb, Breast cancer computer-aided diagnosis system from digital mammograms, JAMMR 30 (5) (Aug. 2019) 1–15.
[6] P. Kulkarni, Fine grained classification of mammographic lesions using pixel N-GRAMS, AJCT (2019).
[7] A.O. Ibrahim, Classification of mammogram images using radial basis function neural network, in: F. Saeed, F. Mohammed, N. Gazem (Eds.), Emerging Trends in Intelligent Computing and Informatics. IRICT 2019. Advances in Intelligent Systems and Computing, Springer, Cham, 2020.
[8] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, A. Alavudeen Basha, Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform, Measurement 146 (2019) 800–805, https://doi.org/10.1016/j.measurement.2019.05.083, ISSN 0263-2241.
[9] Bushra Mughal, Early Detection and Classification of Breast Tumor From Mammography, COMSATS Institute of Information Technology, Islamabad, 2019.
[10] N.S. Lakshmi, M. Chitra, E. Christina, Breast cancer detection using KNN classifier, IRE J. 2 (10) (2019) 76–79.
[11] P. Kaur, G. Singh, P. Kaur, Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification, Inf. Med. Unlocked 100239 (2019), https://doi.org/10.1016/j.imu.2019.100239.
[12] Beham M. Parisa, R. Tamilselvi, S.M. Mansoor Roomi, A. Nagaraj, Accurate classification of cancer in mammogram image, in: H. Saini, R. Singh, G. Kumar, G. Rather, K. Santhi (Eds.), Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems, Springer, Singapore, 2019.
[13] Homayoon Yektaei, Mohammad Manthouri, Faezeh Farivar, Diagnosis of Breast Cancer using Multi scale Convolutional Neural Network. https://doi.org/10.4015/S1016237219500340
[14] J. Wang, Y. Yang, A context-sensitive deep learning approach for microcalcification detection in mammograms, Pattern Recogn. 78 (2018) 12–22, https://doi.org/10.1016/j.patcog.2018.01.009.
[15] S.A. Agnes, J. Anitha, S.I.A. Pandian, et al., Classification of mammogram images using multiscale all convolutional neural network (MA-CNN), J. Med. Syst. 44 (2020) 30, https://doi.org/10.1007/s10916-019-1494-z.
[16] S. Jasim Mohammed, T. Hayder Abbas, Comparison of a classifier performance testing methods: support vector machine classifier on mammogram images classification, J. Kufa Math. Comput. 6 (1) (2019) 8–12, https://doi.org/10.31642/jokmc/2018/060102.
[17] M.A. Arpana, Prathiba Kiran, Feature extraction values for digital mammograms, Int. J. Soft Comput. Eng. (IJSCE) 4 (2) (2014).
[18] Ioan Buciu, Alexandru Gacsadi, Directional features for automatic tumor classification of mammogram images, Biomed. Signal Process. Control 6 (2011) 370–378, https://doi.org/10.1016/j.bspc.2010.10.003.
[19] H. Viswanath, L. Guachi-Guachi, and S. P. Thirumuruganandham, "EasyChair Preprint Breast Cancer Detection Using Image Processing Techniques and Classification Algorithms Breast Cancer Detection Using Image Processing Techniques and Classification Algorithms," 2019.
[20] Eltoukhy, Mohamed Meselhy & Faye, Ibrahima & Brahim, Samir. (2010). Curvelet Based Feature Extraction Method for Breast Cancer Diagnosis in Digital Mammogram. 10.1109/ICIAS.2010.5716125.
[21] Arfan Jaffar, Bilal Lodhi, Ayyaz Hussain, Nawazish Naveed, Fauzia Jabeen, Anwar Mirza, Multi domain features based classification of mammogram images using SVM and MLP, in: Innovative Computing, Information and Control, International Conference on, 2009, pp. 1301–1304, 10.1109/ICICIC.2009.270.
[22] Y. Jin, Medical Image Processing with Deep Learning : Mammogram Classification and Automatic Lesion Detection, pp. 1–19, 2019.
[23] A.A. Kayode, N.O. Akande, A.A. Adegun, M.O. Adebiyi, An automated mammogram classification system using modified support vector machine, Med Devices (Auckl) 12 (2019) 275–284, https://doi.org/10.2147/MDER.S206973.
[24] T.G. Debelee, A. Gebreselasie, F. Schwenker, M. Amirian, D. Yohannes, Classification of mammograms using texture and CNN based extracted features, J. Biomimetics Biomater. Biomed. Eng. 42 (2019) 79–97, https://doi.org/10.4028/www.scientific.net/jbbbe.42.79.