Magnetic Resonance Imaging Breast Scan Classification based on Texture Features and Long Short-Term Memory Model

Suha Raheem Hilal[1*], Hussain S. Hasan[2], Ali M. Hasan[3]

Abstract

The aim of study is building new program for processing MRI images using MATLAB and to investigate different breast MRI detection algorithms that inform normal and abnormal scans of MRI. In this research an algorithm is proposed to extract texture feature and inform normal and abnormal scans of MRI. First, the MRI scans are pre- processed by image enhancement, intensity normalization, background segmentation and detection of mirror symmetry of breast. Second, the proposed gray level co- occurrence matrix (GLCM) and gray level run length matrix (GLRLM) methods are used to extract texture features from MRI T2-weighted and STIR images. Finally, these features are classified into normal and abnormal by using long short term memory (LSTM) model. The research will be validated using 326 datasets that downloaded from cancer imaging archive (TCIA). The achieved classification accuracy was 98.80%.

Key Words: Deep Learning, GLCM, GLRLM, MRI Breast Scans, Feature Extraction, LSTM Classifier.

DOI Number: 10.14704/nq.2021.19.7.NQ21082 **NeuroQuantology 2021; 19(7):41-47**

Introduction

Medical imaging is a branch of medicine that employs technology to create photographs of the human body's internal organs. For a number of conditions, these pictures are used in diagnostics, as educational aids, and in routine healthcare. Medical imaging is often referred to as diagnostic imaging because it is widely used to assist physicians in making precise diagnoses (Birry, et al. 2013). Magnetic resonance imaging (MRI) is a well-known technique that helps in the diagnosing and monitoring of breast related conditions. For the breast cancer in particular, the use of MRI has been developed for screening, stage-detecting, and the therapeutic-response based tracking. MRI has a much higher sensitivity than more widely used clinical modalities like mammography and ultrasound (Antropova, et al. 2018).

Breast cancer is considered as one of the highly occurred conditions in women and a major cause of death in the world. Since dynamic contrast-enhanced (DCE) magetic resonance imaging (MRI) has the highest sensitivity, it is commonly used for the detection and diagnosis of breast cancer (Adachi, et al. 2020).

As related work including several studies, Zhang in 2019 among of researcher propose a completely convolutional network-based mask-guided hierarchical learning (MHL) architecture for breast tumor segmentation (FCN). They create a 3D breast mask as the region of interest (ROI) for each image by developing a model of FCN as an initiation step. Then, to conduct a coarse-to-fine segmentation process of breast tumor lesions, create a two-stage FCN model.

Corresponding author: Suha Raheem Hilal
Address: [1*]Department of Physiology and Medical Physics, College of Medicine, Al-Nahrain University, Kadhimiya, Baghdad, Iraq; [2]Department of Physiology and Medical Physics, College of Medicine, Al-Nahrain University, Kadhimiya, Baghdad, Iraq; [3]Department of Physiology and Medical Physics, College of Medicine, Al-Nahrain University, Kadhimiya, Baghdad, Iraq.
E-mail: [1*]hasufau@gmail.com, [2]altai1965@yahoo.com, [3]a.hasan4@colmed-alnahrain.edu.iq
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 11 May 2021 **Accepted:** 14 June 2021

They finally validate their MHL method on 272 patients (Zhang, J. et al. 2019). The estimation of breast density was the aim of Ivanovska and her associates in 2019 in evaluation of MRI data by presentation pipeline which include intensity inhomogeneity correction, breast volume segmentation, nipple extraction, and fibro glandular tissue segmentation (Ivanovska, et al. 2019). Antropova in 2018 proposed subtraction-MRI images of maximum intensity projection (MIP) by Antropova, N., Abe, H. and Giger, M. L. as a way to use convolutional neural networks for the utilization of four-dimensional (4-D) categorized images and eventually for the classification of lesions (CNN). A total of 690 cases were used in the research. On three MRI presentations, regions of interest were chosen around each lesion. Pretrained VGG Net was used to extract CNN features from the ROIs. Three support vector machine classifiers were used to classify malignant or benign lesions. In study of Yurttakal and his team in 2020, when they compared to feature-based approaches, CNNs have shown to be more accurate in image recognition and have shown promise in medical imaging. Using MRI images, CNN was used to classify lesions as malignant or benign tumors. The accuracy of the network is 98.33% (Yurttakal, A. H. et al. 2020). Extraction feature from PET /MRI images in 34 patients. The CAD pipeline detected and segmented suspicious regions and classified lesions as benign or malignant (Vogl, et al. 2019). Among of researcher proposed algorithm for preprocessing and classifying MRI image of brain using bi-directional modified gray level co-occurrence matrix method and long short term memory model respectively (Hasan, et al. 2020).

Materials and Methods

The core-aim of the present study was to build an automated algorithm that informs normal and abnormal scans of MRI. This will also improve the diagnosis procedures' quality as clinicians can spend more time on patients on pathological breast. It starts with the data downloading stage from cancer imaging archive. In the pre-processing level, a series of algorithms is introduced, followed by feature extraction algorithms and classification.

1) Collection of Data

A dataset comprising 89 breast MRI scans with confirmed diagnostic reports was downloaded from The Cancer Imaging Archive (TCIA), of which 11 scans were of healthy patients and 78 were of pathological patients with different abnormalities such as high-risk normal, ductal carcinoma in situ, fibroids and carcinomas. Each scan comprised T2W and STIR sequences that were acquired by "PHILIPS Achieva 1.5 Tesla scanner". The MRI cases were split into individual MRI images, and each image was separated into left and right breasts. Overall, 326 T2W-TSE images and 326 STIR images were obtained. Both MRI sequences (T2W-TSE and STIR) comprised 161 and 165 healthy and pathological images respectively. Figure 1 shows sample breast MRI images from the dataset.

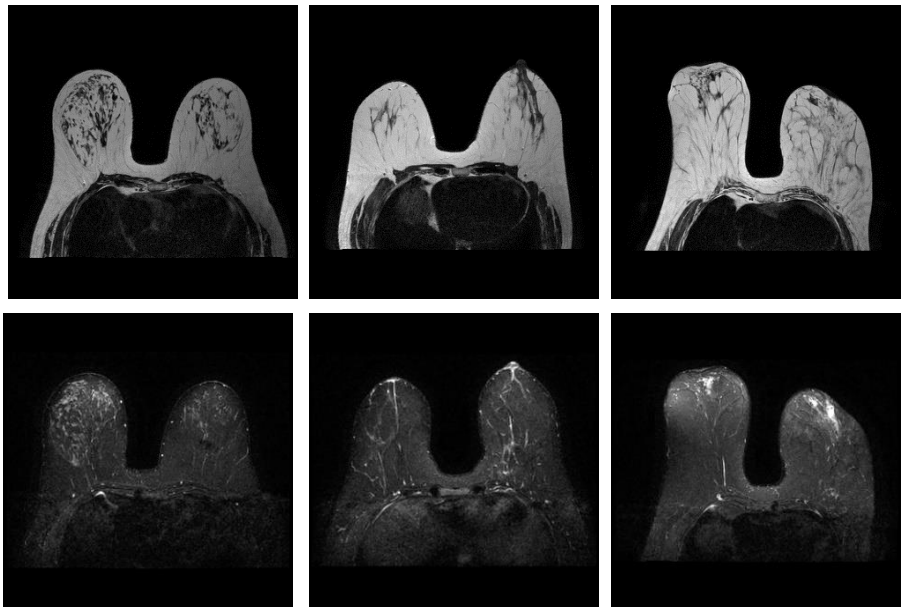

Figure 1. Examples from the used datasets: The first and second rows represent the T_2W and STIR MRI modalities of breast scans

2) Preprocessing of Images

The pre-processing step involves performing an algorithmic set on the breast MRI slices to prepare them for the feature extraction stage. This stage includes breast MRI enhancement of slices by Gaussian filter, intensity of MRI normalization of slices and mirror symmetry detection of breast.

- *Symmetry Detection of Breast*

Along the mid-sagittal plane, which divides the body into two equal portions, the human body is symmetrically bilateral. In order to diagnose diseases and detect pathologic abnormality, researchers have looked at bilateral symmetry from the perspective of morphologic characterization (Derado, et al. 2008). The most important factors that determine breast symmetry are differences in breast contour, form, position, or volume, which can be used to diagnose the breasts using MRI scans in axial view. Any asymmetrical difference may suggest the presence of earlier potential abnormalities in one or both breasts (Chan, et al. 2017). In this study, the symmetry plane which used to find the reflection line of breast MRI scans is determined by registration of original and reflected breast MRI scans and calculation of the eigenvector of the first eigenvalue for the transformation matrix that represents the reflection and registration mappings. Additionally, an ensemble of normalized cross-correlation matches was randomly sampled for a consensus is used to support determination of symmetry plane recognition (Cicconet, et al. 2016). Then, an ellipse model is used to eliminate the chest and abdominal area roughly.

3) Feature Extraction

- *Preparing MRI Breast Slices for Feature Extraction*

This step involves implementing a group of image pre- processing algorithms to ready MRI breast slides and making them more appropriate to implement the GLCM and GLRLM methods. In this study, the textual characteristics were taken such that there is no color information or an ordinary tumor that can be extracted. This can be inferred by showing the density of different images, so that tissue properties can be extracted from breast slices that were found in the following images from MRI with the aim of indicating important information from a clinical point of view and through use of GLCM and GLRLM. Besides, tissue characteristics were employed to statistically understand the two-breast-symmetrical situation.

- *Gray-Level and Run-Length Co-Occurrence Matrices Features*

Gray-level co-occurrence matrix (GLCM) is a second-order statistical method that is used to generate texture features and provide information about the MRI breast scan texture patterning. The texture-feature-producing GLCM uses the joint frequencies (spatial relationship) of all grey-level based pairwise combinations for each distance-detected pixels in an identified orientation of an image under four offsets θ = 0°, 45°, 90°, and 135°, and two distances d=1 and 2, so (d) refer to number of pixels.

For an image I of size ($N \times N$), the GLCM is calculated using equation 1.

$$GLCM(i,j) = \sum_{x=1}^{N}\sum_{y=1}^{N}\begin{cases}1, if\ I(x,y) = i\ and\ I(x+\Delta_x, y+\Delta_y) = j \\ 0, otherwise\end{cases} \quad (1)$$

Where, Δx and Δy are the offset distances between the reference pixel of coordinates x, y, and its neighbors. i and are the coordinates of GLCM.

The 21 co-occurred statistics used in this study, are derived from each GLCM.

While GLRLM relies on the fact that coarse textures are characterized by many neighboring pixels having the same grey level. Where, Run-length features measure the coarseness of an image in specified linear directions. The 7 run-length statistical features that are used in this study are derived from each GLRLM.

For each slice, the grey level maximum number was scaled down to 256 (8 bits/pixel), without utilizing a full dynamic range at 65,536 (16 bits/pixel) prior to the GLCM computation. This critical processing stage was to minimize the entries of zero values in the co-occurrence matrix (Ali M. Hasan, and Farid Meziane, 2016).

4) Feature Aggregation

The GLCM method determines 8 matrices (4 orientations and 2 distances). Twenty-one statistical descriptors were identified to produce 168 descriptors for each slice of an MRI within each matrix. Furthermore, 7 descriptors are determined by GLRLM from each MRI breast slice. Consequently, 175 descriptors are determined from each MRI breast slice. Because of utilizing two MRI modalities of breast scans (T2W and STIR), that's mean there are 350 descriptors are

determined from both MRI breast modalities by GLCM and GLRLM methods respectively. The subsequent classification uses these descriptors to distinguish between normal and pathological MRI breast scans.

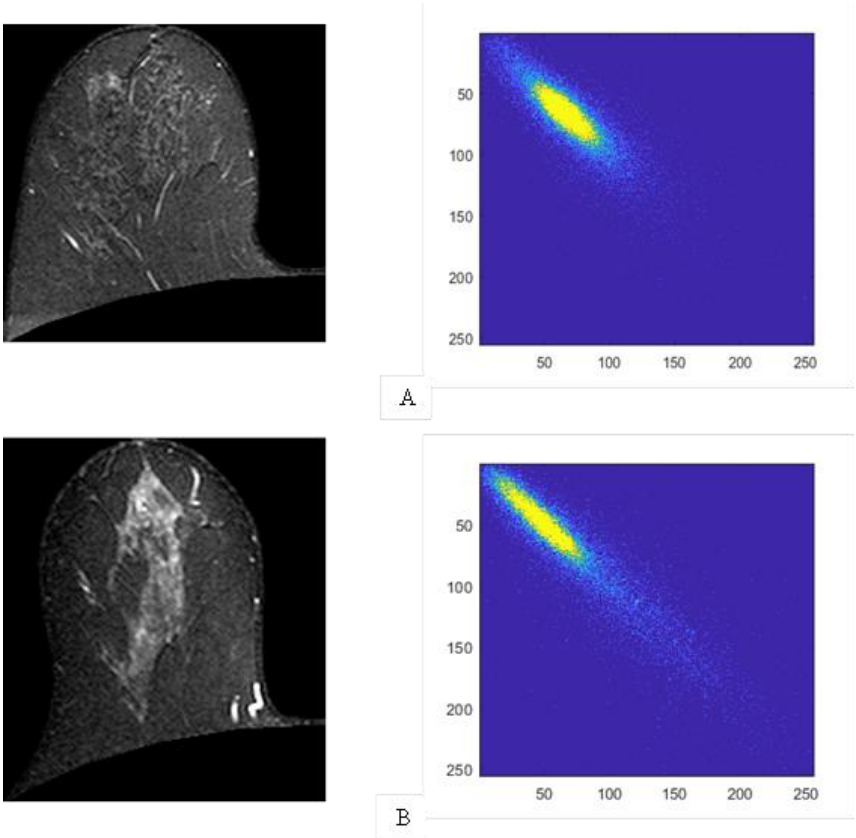

Figure 2. A) Normal MRI breast slice and associated GLCM matrix with θ = 0° and d=1 B) Abnormal MRI breast slice and associated GLCM matrix with θ = 0° and d=1

5) Feature Selection

It is possible for the high-bordered feature groups to detrimentally impact the outcomes of classification, as there are many advantages that can reduce the accuracy of the classification due to the presence of repetition or lack of relevance to some features. One of the main goals of feature selection technologies is to try to define a small subset of features that reduce the increase and increase the relevance. The selection of the feature is a significant step in uncovering the most useful features and in improving the efficiency of the classifier to categorize anonymous data reliably. It is considered a powerful statistical technique by which to analyze data and discover the importance of the predictive tool contained within the feature. It is worth noting that setting the critical value α to 0.001 aims to get very large missions. Forecasters' evaluation is based on the statistical value P-value and F-statistic due to the value of P-value lower than 0.001 is not enough to measure the predictor's significance. Instead of that, the predictors must have a high F statistic value. The rise statistical value of F shows that classes are largely separate from one another. This means that ANOVA works to reduce the number of descriptors from 350 to 198 for each MRI breast slice. However, there is a difference in the ranges that exist for the extracted descriptors, and this is what works to find many of the descriptors that have large values and of course it has a greater effect than the effect that descriptors with small values do on the classifier method. The differences between the extracted features in F-statistic and P-value of the GLCM at θ=0° and d=1 and GLRLM, is shown in Table 1. Where, the autocorrelation, cluster shade, cluster prominence, entropy, homogeneity, sum of square variance, sum variance, difference entropy, RP, RLN, LGRE, and HGRE are selected by ANOVA as the most significant features. Therefore, the feature vector is reduced from 350 to 198 descriptors. The ANOVA was implemented by using IBM SPSS Statistics software Version 20. It helps to improve the performance of the classifier by transforming the given raw descriptors into better form and

more suitable for the training process. Min-max normalization method was recruited, to generate linear transformation on the resulted descriptors with maintaining links with their ancestor sources.

6) Feature Classification

The LSTM type utilized for sequential and time data is also utilized for the classification because its ability to time-wise-identify features using memory block connections via its layers. The length of the input sequence is indicated by the time series in the LSTM mathematical expression. The highest precision of classification with the best results was obtained using the LSTM network at 98.80%.

Table 1. Comparison between the extracted features in F-statistic and P-value of the GLCM at θ =0° and d=1 and GLRLM

Features	F-statistic	P-value
Auto correlation (×10³)	4.019	0.046
Contrast (×10³)	0.063	0.801
Correlation (÷10)	0.142	0.533
Cluster Prominence (×10⁸)	51.470	<0.0001
Cluster Shade (×10⁵)	35.971	<0.0001
Dissimilarity (×10)	3.480	0.063
Energy (÷10)	5.118	0.024
Entropy	9.043	0.003
Homogeneity (÷10)	21.025	<0.0001
Max. Probability (÷10)	2.025	0.156
Sum of Square Variance (×10³)	3.944	0.048
Sum Average (×10²)	0.140	0.709
Sum Variance (×10⁴)	4.361	0.038
Sum Entropy	2.708	0.101
Difference Entropy	14.863	<0.0001
Information Measure of Correlation I (÷10)	0.226	0.635
Information Measure of Correlation II (÷10)	1.334	0.224
Inverse Difference Normalized (÷10)	1.785	0.188
Inverse Difference Moment Normalized (÷10)	0.099	0.631
Weighted Mean (÷10)	4.019	0.046
Weighted Distance	0.063	0.801
SRE	2.045	0.143
LRE	1.371	0.242
GLN	0.026	0.871
RP	5.137	0.024
RLN	22.388	<0.001
LGRE	33.207	<0.001
HGRE	0.026	0.871

The LSTM network's performance is compared to the accuracy of other classifiers such as SVM, LDA and KNN, as demonstrated in table 2. These results demonstrate the superiority of LSTM network to classify MRI breast scans precisely.

The LSTM requires to be configured by setting the parameters of network optimally. These configuration parameters are configured experimentally by trial and observe the performance and the optimal parameters that gave high classification accuracy. The LSTM network of the present work had seven layers of sequence inputs with 198 dimensions derived from the combined resulted features of each slice of MRI, 200 and 20% of hidden units and a drop out, respectively. Besides, the Adam optimization method was used to train the LSTM network, with the highest epoch and gradient threshold value sets of 500 and 1, respectively.

In this study, the randomly collected data set was divided into groups D_1, D_2,.... D_k, who were almost equally sized. However, the process of repeated training and testing, which was indicated by the symbol k, so that the classification result was determined as the overall classification accuracies. For example, in the first iteration, part D_1 is reserved for testing and the remaining parts D_2, D_3 and... D_k are collectively reserved for model training.

The cross-validation was implemented by using cross valid function with k fold in MATLAB R2020a image processing toolkit (Matlab, 2020). The training model (Figure 3) shows the performance of LSTM classifier when training on the combined extracted features (Dr. Padmaja et al. 2020).

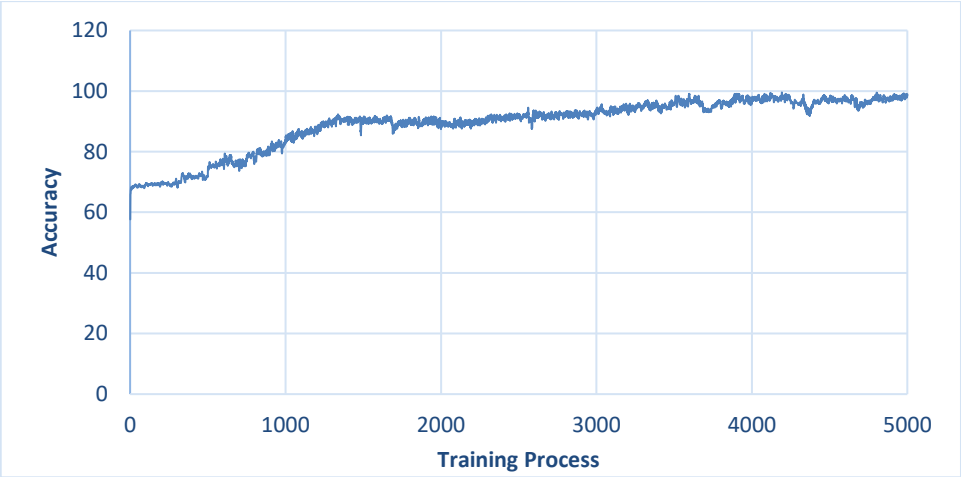

Figure 3. The training process of LSTM network with GLCM and GLRLM features

Result

Results were obtained from 11 normal and 78 pathological breasts MRI scans and classification based on GLCM and GLRLM was 99.37, 98.18, 98.15, 99.38 and 98.80 for TP, TN, Sensitivity, Specificity and Accuracy respectively.

Table 2. The results of classification obtained from KNN, SVM, LDA and LSTM

Methods	Accuracy 100%	TP 100 %	TN 100 %	Sensitivity	Specificity
KNN	96.30	96.89	95.75	95.7	96.93
SVM	95.70	96.89	94.54	94.54	96.89
LDA	94.20	95.03	94	93.86	95.1
LSTM	98.80	99.37	98.18	98.15	99.38

To demonstrate the efficacy of the hypothesized study, MRI-scan based datasets of between 2016 to 2020 were used for comparisons with our findings (Table 3). The motivation behind using these methods was to classify normal and abnormal breast MRI scans employing different techniques with different datasets of MRI.

Table 3. Comparison for previous suggested methods

Accuracy 100%	Classifier	No. of Patients	Feature Methods	References
93.3	SVM	54	GLCM GLRLM GLSZM	Marcon, M. *et al.* 2019
92.16	SVM	200	the fractional Fourier transform	Zhang et al., 2016
91	SVM	124	First order statistic GLCM	Parekh et al., 2017
80	SVM	92	Histogram DWT GLCM GLRM	Song, L., Lu, H. and Yin, J. 2020
67	LRA			
72	QDA			
83.69	NB	690	Gabor Wavelet	Raghavendra et al.,2016
98.80	LSTM	326	GLCM GLRLM	Proposed Method

Conclusion

For clarification, the visual diagnosis that is done through the use of MRI is considered a personal diagnosis and that it is based on the expertise of the radiologist, and for this was studied tissue analysis in order to develop the diagnostic process that is carried out through the use of MRI images. The current features aid to categorize normal and pathological textures. If a larger number of parameters are extracted for a larger number of patients with lesions. Development of a beneficial model can help specialist physicians for early diagnosis and treatment options in a shorter period of time. The combined GLCM and GLRLM features reach the highest accuracy of 99.80%, with the LSTM successfully classifying all MRI breast scans except one pathological breast MRI slice and three normal breast slices are failed to classify correctly.

References

Birry RAK. *Automated classification in digital images of osteogenic differentiated stem cells.* University of Salford 2013.

Antropova N, Abe H, Giger ML. Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *Journal of Medical Imaging*. *SPIE - The International Society for Optical Engineering* 2018; 5(1): 014503. http://doi.org/10.1117/1.jmi.5.1.014503.

Adachi M, Fujioka T, Mori M, Kubota K, Kikuchi Y, Xiaotong W, Oyama J, Kimura K, Oda G, Nakagawa T, Uetake H. Detection and diagnosis of breast cancer using artificial intelligence based assessment of maximum intensity projection dynamic contrast enhanced magnetic resonance images. *Diagnostics* 2020; 10(5): 330.

Zhang J, Saha A, Zhu Z, Mazurowski MA. Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics. *IEEE transactions on medical imaging* 2018; 38(2): 435-447. http://doi.org/10.1109/TMI.2018.2865671.

*e*ISSN 1303-5150

Ivanovska T, Jentschke TG, Daboul A, Hegenscheid K, Völzke H, Wörgötter F. A deep learning framework for efficient analysis of breast volume and fibroglandular tissue using MR data with strong artifacts. *International journal of computer assisted radiology and surgery* 2019; 14(10): 1627-1633.

Yurttakal AH, Hasan E, Türkan İ, Seyhan K. Detection of breast cancer via deep convolution neural networks using MRI images. *Multimedia Tools and Applications* 2020; 79 (21-22): 15555-15573. http://doi.org/10.1007/s11042-019-7479-6.

Vogl WD, Pinker K, Helbich TH, Bickel H, Grabner G, Bogner W, Gruber S, Bago-Horvath Z, Dubsky P, Langs, G. Automatic segmentation and classification of breast lesions through identification of informative multiparametric PET/MRI features. *European radiology experimental* 2019; 3(1): 1-13.

Hasan MH, Hasan HS, Hasan AM. MRI Brain Scans Classification Using Bi-directional Modified Gray Level Co-occurrence Matrix and Long Short-Term Memory. *NeuroQuantology* 2020; 18(9): 54-63.

Derado G, Lee K, Nicolis O, Bowman FD, Newell M, Rugger FF, Vidakovic B. Wavelet-based 3-D multifractal spectrum with applications in breast MRI images. *In International Symposium on Bioinformatics Research and Applications* 2008; 281-292. Springer, Berlin, Heidelberg.

Chan S, Chen JH, Li S, Chang R, Yeh DC, Chang RF, Yeh LR, Kwong J, Su MY. Evaluation of the association between quantitative mammographic density and breast cancer occurred in different quadrants. *BMC cancer* 2017; 17(1): 1-11.

Cicconet M. Hildebrand DG, Elliott H. Finding mirror symmetry via registration. *arXiv preprint arXiv:1611.05971*, 2016.

Hasan AM, Meziane F. Automated screening of MRI brain scanning using grey level statistics. *Computers & Electrical Engineering* 2016; 53: 276-291.

Marcon M, Ciritsis A, Rossi C, Becker AS, Berger N, Wurnig MC, Boss A. Diagnostic performance of machine learning applied to texture analysis-derived features for breast lesion characterisation at automated breast ultrasound: a pilot study. *European radiology experimental* 2019; 3(1): 1-11.

Zhang YD, Wang SH, Liu G, Yang J. Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform. *Advances in Mechanical Engineering* 2016; 8(2): 1687814016634243.

Parekh VS, Jacobs MA. Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric MRI. *NPJ breast cancer* 2017; 3(1): 1-9.

Song L, Lu H, Yin J. Preliminary study on discriminating HER2 2+ amplification status of breast cancers based on texture features semi-automatically derived from pre-, post-contrast, and subtraction images of DCE-MRI. *PLoS ONE* 2020; 15(6): e0234800. http://doi.org/10.1371/journal.pone.0234800

Raghavendra U, Acharya UR, Fujita H, Gudigar A, Tan JH, Chokkadi S. Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images. *Applied Soft Computing* 2016; 46: 151-161.

Padmaja VK, Jayanth Srinivas S, Lokesh S. Image Enhancement Using Gaussian Filters. *International Research Journal of Engineering and Technology (IRJET)* 2020; 5(7).

Zhang H, Chen Y, Shi J. The influence of cranial nerves on the financial risk selection and trading behavior of investors. *NeuroQuantology* 2018; 16(6): 158-163.