Classification of breast cancer images using completed local ternary pattern and support vector machine

M. Kusuma Sri*

Department of Electronics and Communication Engineering,
Anurag University,
Hyderabad – 500088, India
Email: kusumasri.personal2@gmail.com
*Corresponding author

E. Gomathi

Department of Electronics and Communication Engineering,
KIT-Kalaignarkarunanidhi Institute of Technology,
Kannampalayam, Coimbatore – 641402, India
Email: egomathimanian@gmail.com

Abstract: Breast cancer is the major cause of deaths in women compared to other cancers. Though early detection of breast cancer reduces cancer deaths, it is a challenging task for physicians. Local binary pattern (LBP) and local ternary pattern (LTP) techniques are widely applied in texture classification applications. Since LBP is more sensitive to noise in texture classification, it needs to be improved for achieving better results. Though LTP is more robust to noise, there are few drawbacks. Completed LBP and completed local binary count techniques achieve good accuracy for texture classification, but they inherit few drawbacks of LBP. In this paper, completed LTP operator is applied on breast cancer images for better classification accuracy than LBP and completed LBP operators, by extracting sign and magnitude components. Experimental results based on breast cancer database show that the proposed technique achieved better classification accuracy than existing similar approaches.

Keywords: breast cancer image; image segmentation; texture classification; machine learning; decision tree; logistic regression; SVM.; support vector machine.

Reference to this paper should be made as follows: Kusuma Sri, M. and Gomathi, E. (2022) 'Classification of breast cancer images using completed local ternary pattern and support vector machine', *Int. J. Bioinformatics Research and Applications*, Vol. 18, Nos. 1/2, pp.130–140.

Biographical notes: M. Kusuma Sri received her BTech in Electronics and Communication Engineering in the year 2003 from Swami Ramananda Tirtha Institute of Science and Technology, Nalgonda and MTech in Systems and Signal Processing from Adam's Engineering College, Paloncha in 2011. Currently, she is working as an Assistant Professor in the Department of

Electronics and Communication Engineering at Anurag University, Hyderabad, India. Her areas of interest include image segmentation, biomedical image analysis and artificial intelligence. She published few papers in international conferences and reputed journals.

E. Gomathi received her BE in Electronics and Communication Engineering in 1997 from Manonmanium Sundaranar University and the ME in Control and Instrumentation in 2005 from Karunya University. She received her PhD from Anna University. Currently she is working as an Associate Professor at Kalaignar Karunanidhi Institute of Technology, Coimbatore, India. She published several papers in refereed national and international journals. Her research interests include image processing, face recognition and genetic algorithm.

1 Introduction

Breast cancer is mainly responsible for large number of cancer related deaths among women. It is highly essential to diagnose breast cancer at earlier stages to improve the survival rate. Though breast cancer begins in the breast cells with a sign of a new lump or some sort of discomfort, it travels to lymph nodes under the arms and other parts of the body. Since early detection is required to pay immediate attention, mammograms are found to be useful in identifying abnormalities. Based World Health Organization (WHO) projections, it is estimated that cancer deaths could increase to 15 million by 2030 more number of cancers are expected to occur in developing nations.

Mammography is one of the solutions for early identification, but mammographic images are sensitive to the noise. Hence, noise removal stage is highly essential for radiologists to perform better diagnosis. In recent years, texture analysis and classification gained huge popularity in biomedical image analysis. Few applications are breast cancer detection, face recognition and movement detection. Numerous feature extraction algorithms are proposed by many authors to attain better accuracy in the above applications.

Mammography images are classified using decision tree classifier and Naïve Bayes algorithm for supporting oncologist to improve the overall treatment process. Improved classification accuracy will be of a great help to enhance the survival rate. Local binary pattern (LBP) framework is combined with the above classifiers for achieving maximum accuracy in detecting breast cancer cases (Padhi and Kumar, 2019). Gray level cooccurrence matrix (GLCM) extracts the many features of the image to improve the segmentation accuracy. Generally, segmentation methods are applied over the region of interest that can be identified by performing rigorous dataset examination. Segmentation process separates cancer cells from the background for focusing more on the affected part. In addition to segmentation, background removal algorithms are found to be useful in breast cancer diagnosis. Feature selection algorithm chooses optimal features to classify the mammogram images into benign and malignant (Sarosa et al., 2018).

k-Nearest Neighbour (k-NN) classifier is one of the improved classifier that finds applications in detection and recognition. k-NN algorithm is capable of providing highly accurate predictions while comparing with many other classification algorithms (Htay and Maung, 2018). Morphology segmentation techniques with region of interest extraction has been used in breast cancer segmentation (Salih and Kamil, 2018).

Since the morphology based approaches are ineffective in cancer detection, fuzzy c-means (FCM) algorithm is applied on the region of interest to extract necessary features. FCM algorithm with GLCM feature extraction is greatly helpful for performing better segmentation (Saleck et al., 2017).

In machine learning, artificial neural networks (ANN), k-NN and support vector machine (SVM) learning models are widely deployed for classification (Reis et al., 2017). Local binary pattern is used to compute feature vectors (Král et al., 2016). In biomedical image classification problems, SVM is preferred over other similar classifiers. SVM classifier is utilised in this work for classifying mammogram images into cancer and normal. In SVM classifier, support vectors are the data points that can be maximised which helps to build while performing the data grouping.

Mammography based breast cancer diagnosis is a safest method for breast cancer detection, but abnormalities in the images may not be visible after performing segmentation (Wang et al., 2014). Hence preprocessing and feature extraction are required before performing segmentation to avoid human errors in cancer diagnosis (Bandyopadhyay, 2010). Level set segmentation, spatial fuzzy clustering and convolutional neural network are used in segmentation and classification of breast cancer images. However, the computational complexity is the major issue in developing automatic classification procedure (Yousefikamal, 2019). Histogram statistics and spatial relation based GLCM have been utilised in many research works to reduce the number of computations in breast cancer risk assessment (Manduca et al., 2009; Heine et al., 2012). To improve the classification accuracy and to reduce computations, multiple optimisation and unsupervised sparse auto encoder have been used in mammographic image segmentation and classification approaches (Kallenberg et al., 2016).

Based on the above works, it is observed that classification suffers mainly due to the noise components and there is a need for better classification technique for accurate detection of cancer in mammographic images. However, the algorithms focuses on extracting distinctive texture features that have robustness to noise. LBP considers only magnitude component whereas local ternary pattern (LTP) considers magnitude component as well as sign component. However, LBP and LTP operators are suffering from noise and it may lead to incorrect classification of different patterns.

In this paper, completed LTP (CLTP) operator is used to improve the classification accuracy that is more robust to noise. This paper comprises of five sections. Section 2 provides brief review of various descriptors like LBP, LTP. Section 3 presents the CLTP operator based segmentation. Section 4 discusses the results obtained using mini mammography image analysis society (MIAS) database. Section 5 presents the conclusion of the paper.

2 Texture classification using various descriptors

In this section, LBP and LTP operators are reviewed which are widely utilised in texture classification applications.

2.1 Local binary pattern

LBP is basic methodology for developing many advanced texture classification algorithm in biomedical imaging applications. LBP operator is computed by comparing the intensity value of centre pixel with the intensity values of its neighbours. The LBP code generation is shown in Figure 1. Mathematically expression for LBP is given by

$$LBP_{P,R} = \sum_{p=0}^{p-1} 2^p s\left(i_p - i_C\right),$$

$$s\,(x) = \{1, x \geq 0,$$

$$0, x < 0, \tag{1}$$

where i_C and $i(p = 0, P - 1)$ denote the grey value of the centre pixel and grey value of the neighbour pixel respectively.

Figure 1 Process of LBP code generation (see online version for colours)

62	80	104
50	34	15
9	60	3

(3x3 neighbourhood of the image)

→

1	1	1
1		0
0	1	0

(Binary neighbourhood)

→

2^0	2^1	2^2
2^7		2^3
2^6	2^5	2^4

(Corresponding weights)

→

11100101
decimal=167

(LBP Code)

2.2 *Local ternary pattern*

LTP is one of the important texture classification descriptor and widely recognised as an important descriptor similar LBP. Figure 2 describes the process of LTP code generation. LTP operator is expressed as

$$LTP_{P,R} = \sum_{p=0}^{p-1} 2^p s\left(i_p - i_C\right),$$

$$S(x) = \begin{cases} 1, & x \geq t, \\ 0, & -t < x < t, \\ -1\,, & x < -t, \end{cases} \tag{2}$$

where t is the threshold value. Upper and lower patterns are constructed after performing thresholding operation. The concatenation of both upper and lower patterns is used to represent the LTP operator.

3 Proposed breast cancer image segmentation using CLTP

Breast cancer image segmentation needs to be improved with better segmentation algorithm that needs to be more robust to noise. Both the LBP and LTP techniques have the scope to modify its original code generation process for better classification of given images. Here LTP descriptor is modified to enhance the segmentation process. In this section, mammography image segmentation is proposed for breast cancer classification using CLTP.CLTP algorithm is developed to have more robustness to noise than similar algorithms. Here, associated completed LTP is constructed that helps to enhance and increase its discriminating property. In CLTP, image decomposition is performed to extract sign and magnitude complementary components. The above components are written as follows

$$S_p^{upper} = s\,(i_p - (i_c + t)),\ S_p^{lower} = s\,(i_p - (i_c - t))$$

$$m_p^{upper} = |i_p - (i_c + t)|,\ m_p^{upper} = |i_p - (i_c - t)|, \tag{3}$$

where the S_p^{upper} and S_p^{upper} are used to build the $CLTP\_S_{P,R}^{upper}$ and, $CLTP\_S_{P,R}^{lower}$ respectively.

$CLTP\_S_{P,R}^{lower}$ is expressed using the following equation

$$CLTP\_S_{P,R}^{upper} = \sum_{p=0}^{p-1} 2^p s\left(i_p - \left(i_c + t\right)\right), \tag{4}$$

$$S_p^{upper} = \begin{cases} 1\,, & i_p \geq i_c + t \\ 0, & \text{otherwise} \end{cases}$$

Similar to the upper component, it is possible to express $CLTP\_S_{P,R}^{lower}$ as

$$CLTP\_S_{P,R}^{lower} = \sum_{p=0}^{p-1} 2^p s\left(i_p - \left(i_c - t\right)\right), \tag{5}$$

$$S_p^{lower} = \begin{cases} 1\,, & i_p < i_c - t \\ 0, & otherwise \end{cases}$$

The $CLTP\_S_{P,R}$ is formed by the concatenation of $CLTP\_S_{P,R}^{upper}$ and $CLTP\_S_{P,R}^{lower}$, which is written as

$$CLTP\_S_{P,R} = \left[CLTP\_S_{P,R}^{upper}, CLTP\_S_{P,R}^{lower}\right], \tag{6}$$

Similar to $CLTP\_S_{P,R}$, the $CLTP\_M_{P,R}^{upper}$ is determined using the two magnitude complementary components m_p^{upper} and m_p^{lower} as below:

$$CLTP\_M_{P,R}^{upper} = \sum_{p=0}^{p-1} 2^p t\left(m_p^{upper}, c\right), \tag{7}$$

$$t\left(m_p^{upper}, c\right) = \begin{cases} 1\,, \left|i_p - \left(i_c + t\right)\right| \geq c, \\ 0, \left|i_p - \left(i_c + t\right)\right| < c \end{cases}$$

$$CLTP\_M_{P,R}^{lower} = \sum_{p=0}^{p-1} 2^p t\left(m_p^{lower}, c\right), \tag{8}$$

$$t\left(m_p^{lower}, c\right) = \begin{cases} 1, \left|i_p - \left(i_c - t\right)\right| \geq c, \\ 0, \left|i_p - \left(i_c - t\right)\right| < c, \end{cases}$$

$$CLTP\_MP = \left[CLTP\_M_{P,R}^{upper}, CLTP\_M_{P,R}^{lower}\right], \tag{9}$$

where c is the mean value of mp. $CLTP\_C_{P,R}^{upper}$, and $CLTP\_C_{P,R}^{lower}$ can be described as

$$CLTP\_C_{P,R}^{upper} = t\left(i_c^{upper}, cI\right), CLTP\_C_{P,R}^{lower} = t\left(i_c^{lower}, cI\right), \tag{10}$$

where $i_c^{upper} = ic + t$, $i_c^{lower} = ic - t$ and cI is the average grey level of the image.

Breast cancer image classification is focused in this work using CLTP, GLCM and SVM classification techniques. CLTP descriptor is used for the given breast cancer image segmentation. GLCM approach is utilised for feature extraction process that is significant for better classification. SVM classifier is finally applied on the feature extracted images to categorise into normal image and cancer image. Figure 3 shows the important stages involved in the proposed classification approach. Input images are collected from mini MIAS database, the images are of size 1024×1024 divided into non overlapping smaller size (64×64 size) images which helps for better feature extraction by segmentation and improves the overall performance. Now the CLTP algorithm is applied to find lower and upper pattern values for each image then features are extracted by using GLCM (Spanhol et al., 2015). In this paper, thirteen features of GLCM are used which are stored with labelling. Finally, these features are applied into SVM for classification of images. SVM model is built in machine learning environment to classify the images as Benign and Malignant cancer images.

Figure 2 The process of LTP code generation

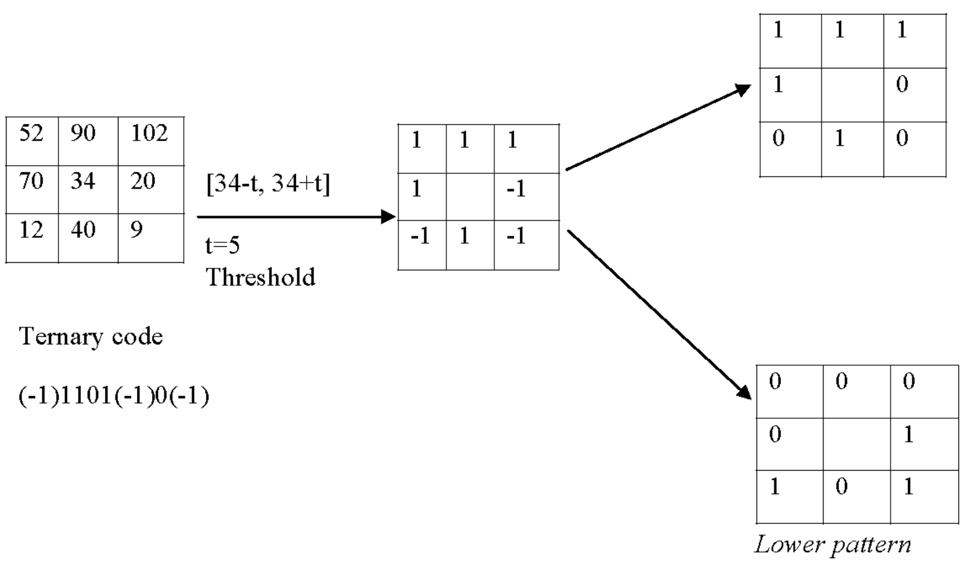
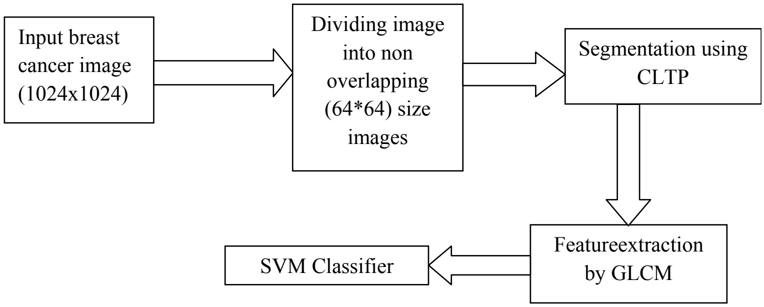


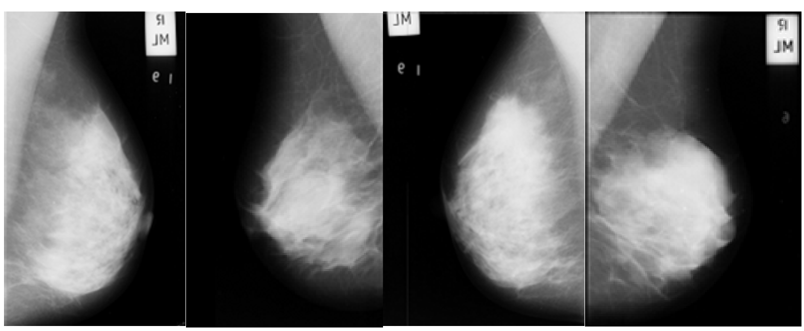
Figure 3 Proposed image segmentation using CLTP and SVM classification



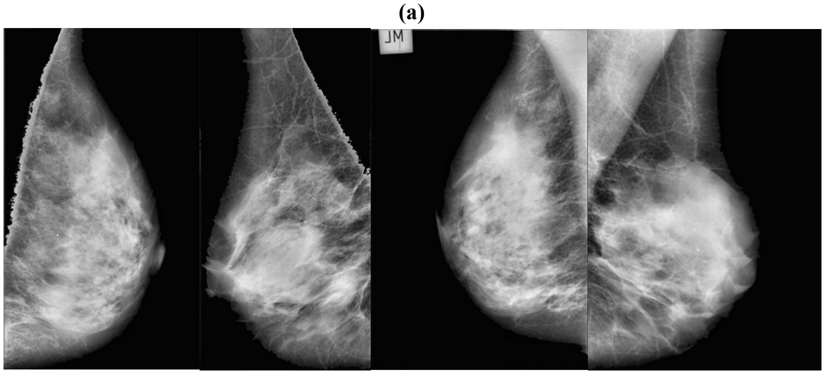
4 Results and discussion

Breast cancer Images are collected from mini MIAS database, which consist 322 images of normal and diseased images. Few sample input images of MIAS database and its segmented images are depicted in Figure 4. Database consists of different types of diseased and normal images that are provided in Table 1. Different types of MIAS database images are considered. Few abnormalities present in MIAS database images are calcification, ill-defined masses and speculated masses. Table 1 provides the complete list of abnormalities and different types of MIAS data based images. Severity of abnormality is classified as benign and malignant cancer images.

Figure 4 (a) Input MIAS database images and (b) segmented cancer images



(a)



(b)

Table 1 Different types of MIAS database images

Types of images	
F	Fatty
G	Glandular
D	Dense
Abnormality present in the class	
CALC	Calcification
CIRC	Well-defined/circumscribed masses
SPIC	Speculated masses
MISC	Other, ill-defined masses
ARCH	Architectural distortion
ASYM	Asymmetry
NORM	Normal
Severity of abnormality	
B	Benign
M	Malignant

All the images in the database are first divided into non-overlapping smaller size images then Segmented by using CLTP operator. Later GLCM features were extracted and classified using SVM model, k-NN model, Decision tree and logistic regression by considering training phase(0.8) and testing phase(0.2) for each class. Segmentation and feature extraction is implemented in MATLAB and classification is implemented in python. Results are based on confusion matrix with true positive, true negative, false positive and false negative.

Table 2 describes the confusion matrix with predicted class and actual class for finding number of true positives, number of true negatives, number of false positives, number of false negatives. Classification accuracy is obtained from the confusion matrix. Accuracy of different classifiers is displayed in Figure 5 as chart. Accuracy is defined as the measure of correct prediction. It is calculated by

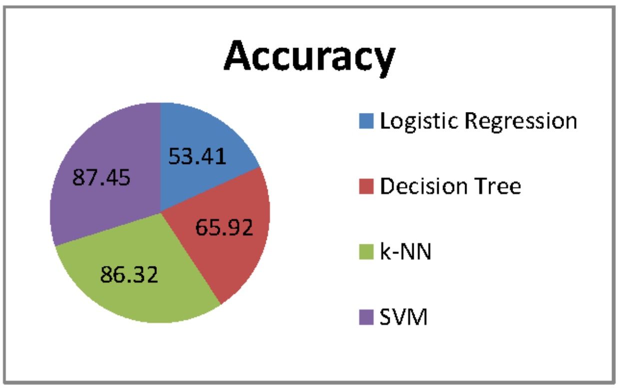
$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{11}$$

Here the results are compared by considering logistic regression, decision tree, k-NN and SVM classifiers. Accuracy of these classifiers is calculated using equation (11). Classification accuracy of better classification accuracy is achieved for SVM model while comparing with all the above considered models. The accuracy obtained using logistic regression, decision tree, k-NN and SVM classifiers are 53.41%, 65.92%, 86.32% and 87.45% respectively. Accuracy of SVM classifier is found to be better than other classifiers. It has to be noted that accuracy is improved after performing preprocessing and feature extraction using CLTP and GLCM operators.

Table 2 Confusion matrix

Actual class	*Predicted class* 1	2	3	4	5	*Total*
1	6550	742	622	210	140	8264
2	764	5865	624	212	0	7465
3	530	704	5325	320	0	6879
4	450	423	430	4150	0	5453
5	461	240	320	860	3850	5731
Total	8755	7974	7321	5752	3990	33,792

Figure 5 Accuracy of classifiers (see online version for colours)



In addition to classification accuracy, the performance of the work is assessed using precision, recall and F1 score. They are defined as follows:

Precision:

Precision is proportion of predicted positive which is actually positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

Recall:

Recall is the proportion of positive which is predicted as positive

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

F1-Score:

F1 score is the harmonic mean of the precision and recall, where the best F1 score value is1 and worse value is 0. The correctness of classifier (F1) is given by

$$\text{F1 Score} = \frac{2*\text{precision}*\text{recall}}{\text{precision} + \text{recall}} \tag{14}$$

Table 3 describes the results obtained for different supervised machine learning models by applying breast cancer images. Accuracy gives the correct prediction of breast cancer using these machine learning models. Precision, F1-score and recall are better in k-NN and SVM classifiers than logistic regression and decision tree learning models. SVM approach is slightly better than k-NN learning model; hence SVM model is suggested for breast cancer prediction in this work. All these performance measures are calculated after performing preprocessing using CLTP and GLCM operators. However, there is a scope for improvement by deploying other preprocessing techniques to improve accuracy and other performance measures.

Table 3 Comparison of different machine learning classifiers

Model	*Accuracy*	*Precision*	*F1-Score*	*Recall*
Logistic regression	53.41%	48.32%	49.83%	51.45%
Decision tree	65.92%	52.78%	54.70%	56.78%
k-NN	86.32%	62.62%	68.46%	75.52%
SVM	87.45%	72.53%	74.46%	76.63%

5 Conclusion

Completed LTP operator is built in this work to overcome the drawbacks of LBP in breast cancer image segmentation. Better feature extraction of breast cancer images is possible in this work by dividing the given image into non overlapping smaller size images. Thirteen features of GLCM have been used for feature extraction and the various machine-learning algorithms are applied for classification. From the obtained results it is clear that CLTP is more robust to noise and achieved better accuracy while using SVM compared to other machine learning algorithms. Future work may be focused on deep learning neural network models that can learn discriminating features on its own**.**

References

Bandyopadhyay, S.K. (2010) 'Pre-processing of mammogram images', *International Journal of Engineering Science and Technology*, Vol. 2, No. 11, pp.6753–6758.

Heine, J.J., Scott, C.G., Sellers, T.A., Brandt, K.R., Serie, D.J., Wu, F.F., Morton, M.J., Schueler, B.A., Couch, F.J., Olson, J.E. and Pankratz, V.S. (2012) (2012) 'A novel automated mammographic density measure and breast cancer risk', *J. Nat. Cancer Inst.*, Vol. 104, pp.1028–1037.

Htay, T.T. and Maung, S.S. (2018) 'Early stage breast cancer detection system using GLCM feature extraction and K-nearest neighbor (k-NN) 'on mammography image', *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, Bangkok, pp.171–175.

Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C. and Lillholm, M. (2016) 'Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring', *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp.1322–1331.

Král, P. and Lenc, L. (2016) 'LBP features for breast cancer detection', *International Conference on Image Processing (ICIP)*, Phoenix, Arizona, USA, pp.2643–2647.

Manduca, A., Carston, M.J., Heine, J.J., Scott, C.G., Pankratz, V.S., Brandt, K.R., Sellers, T.A., Vachon, C.M. and Cerhan, J.R. (2009) 'Texture features from mammographic images and risk of breast cancer', *Cancer Epidemiol. Biomarkers Prev.*, Vol. 18, pp.837–845.

Padhi, T. and Kumar, P. (2019) 'Breast cancer analysis using WEKA', *2019 IEEE International Conference on Cloud Computing, DataScience and Engineering (Confluence)*, Noida, India, pp.978–1–5386-5933.

Reis, S., Gazinska, P., Hipwell, J.H., Mertzanidou, T., Naidoo, K., Williams, N., Pinder, S. and Hawkes, D.J. (2017) 'Automated classification of breast cancer stroma maturity from histological images', *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 10, pp.2344–2352.

Saleck, M.M., ElMoutaouakkil, A. and Mouçouf, M. (2017) 'Tumor detection in mammography images using fuzzy C-means and GLCM texture features', *2017 14th International Conference on Computer Graphics, Imaging and Visualization*, Marrakesh, pp.122–125.

Salih, M. and Kamil, M.Y. (2018) 'Mammography image segmentation based on fuzzy morphological operations', *2018 1st Annual International Conference on Information and Sciences (AiCIS),* Fallujah, Iraq, pp.40–44.

Sarosa, S.J.A., Utaminingrum, F. and Bachtiar, F.A. (2018) 'Mammogram breast cancer classification using gray-level co-occurrence matrix and support vector machine', *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, IEEE, Malang, Indonesia, pp.54–59.

Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L. (2015) 'A dataset for breast cancer histopathological image classification', *IEEE Transactions on Biomedical Engineering*, Vol. 63, No. 7, pp.1455–1462.

Wang, Z., Yu, G., Kang, Y., Zhao, Y. and Qu, Q. (2014) 'Breast tumor detection in digital mammography based on extreme learning machine', *Neurocomput.*, Vol. 128, pp.175–184.

Yousefikamal, P. (2019) Breast Tumor Classification and Segmentation using Convolutional Neural Networks, arXiv preprint arXiv: 1905.04247.