Measurement 146 (2019) 800–805

Contents lists available at ScienceDirect

Measurement

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

Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform

R. Vijayarajeswari [a], P. Parthasarathy [b,*], S. Vivekanandan [b], A. Alavudeen Basha [b]

[a] Department of Computer Science Engineering, Mahendra Engineering College (Autonomous), Mallasamudram, Namakkal, India
[b] School of Electrical Engineering, VIT University, India

ARTICLE INFO

Article history:
Received 20 March 2019
Received in revised form 17 May 2019
Accepted 23 May 2019
Available online 24 June 2019

Keywords:
Support vector machine
Hough transform
Formatting
Styling
Breast cancer

ABSTRACT

Breast cancer is one of the significant health problems in the world. If these abnormalities in breast cancer are detected early there is a maximum chance for recovery. For this early prediction we can go for mammography. It is one of the most effective and commonly used method for detecting and screening breast cancer. This paper presents classification of mammograms using feature extracted using Hough transform. Hough transform is a two dimensional transform. It is used to isolate feature of particular shape in an image. Miniaturized scale characterization and masses are the two most vital markers of threat, and their mechanized identification is exceptionally important for early breast cancer diagnosis. Since masses are regularly undefined from the encompassing parenchymal, computerized mass location and arrangement is significantly additionally difficult. This paper talks about the strategies for classification and feature extraction. Here, Hough transform is used to detect features of mammograms image and it is classified using SVM. The classification accuracy is more by the use of SVM classifier. This method is tested on 95 mammograms images collected and classified using SVM. From the result it shows that the proposed method is effectively classify the abnormal classes of mammograms.
© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Tumor is a kind of malady and the principle attributes of breast cancer disease is the wild development of cells in a particular area of a body. This cell development is likewise refereed as tumor. Breast malignancy is shaped when disease creates from breast tissue [1]. It is one the significant medical issues in the present world. As indicated by the statics of world social insurance association in 1960's and 1970's a quick increment of breast cancer disease was enrolled in the episode rates in the few nations [2]. Early recognition of tumor can expand the recuperation rate, all things considered, and it can keep from kicking the bucket. Mammography can be utilized for early forecast, location and treatment of breast cancer. Mammography can without much of a stretch identify the tumor cells that are little and extremely hard to feel and it is the standout amongst the most widely recognized strategy used to distinguish breast cancer. Breast imaging in the mammography is chiefly finished with the assistance of low-measurements X-beams with high goals and high differentiation [3–5]. It is additionally utilized for both screening and diagnosing breast disease. At times Full Field advanced mammography (FFDM) is utilized to stay away from superfluous biopsies. Because of this in a last decade numerous inquiries about have been done in this specific field to construct a computational framework to encourage the radiologist. They are chiefly a Computer supported outline and conclusion frameworks. This frameworks give extra wellspring of data and can expand the right recognition rates for infection like bosom tumor Table 1.

The assessed affectability of radiologist in breast disease screening is just around 70%, however the execution would be enhanced on the off chance that they were incited with the conceivable area of variations from the norm. Breast malignancy CAD framework can give such help and they are imperative and important for breast growth control. Mammography gives a philosophy to help radiologist for recognizing the majority in the mammogram pictures and to arrange them as ordinary or unusual [6]. It can likewise effectively distinguish cell development that are little. Presently multi day's classifiers assume a noteworthy job in therapeutic conclusion. It gives the likelihood of mistakes and results inside a brief timeframe. Framework execution likewise rely upon methods utilized for division of the mammogram picture and highlight extraction. Standard improvement methods like histogram are utilized to hone the intrigued area of the mammogram picture limits. Complexity upgrade is done between the district of intrigue and close-by typical tissue.

* Corresponding author.
E-mail address: arjunsarathii@gmail.com (P. Parthasarathy).

https://doi.org/10.1016/j.measurement.2019.05.083
0263-2241/© 2019 Elsevier Ltd. All rights reserved.

Table 1
Some related for mammogram classification [7–14].

Author	Image classes	SBT/ROI	No. of Images	Features	Classifiers	Acc %
Karssemeijer et al.	4-class	SBT	615	Hough Transform	KNN	65
Petroudi et al.	2-class	SBT	132	Statistical & grey level based Features	KNN	75
Oliver et al.	4-class	SBT	300	Relative areas, center of mass	Decision tree	47
Bovis et al.	4-class	SBT	377	SGLD features	ANN	71
Mustra et al.	4-class	ROI	144	GLCM features	KNN	79
Kumar et al.	4-class	ROI 128 * 128	480	Wavelet packed texture description	SVM	73
Qu et al.	2-class	SBT	322	–	E-FELM	72
Z. Chen et al.	4-class	SBT	322	Texton features	KNN, Bayesian	75
Present Study	3-class	265 * 8256	95	Hough Transform	SVM	94

Region of interest (ROI) was considered as the enhanced sharp edge or boundaries of mammogram image. Now segmentation is done on ROI using common statistical based morphological approaches. In this work three main steps was done. First method is to remove the unwanted markings and label in the image. Then intensity based segmentation is done for removing pectoral muscle. After segmentation of pectoral muscle Hough transform is done on ROI to extract features. This is an effective method to recognize a pattern. It is an image transform method in which feature extraction of particular image with particular shape within an image can be achieved. It also transforms an original image in to a 2D function. Feature extraction play an important role in the classification. The feature extracted from Hough transform is utilized to justify them as normal or abnormal. Many methods like Bricial neural networks, LDA-Linear discriminant analysis, nearest neighbor methods were used. In this work SVM classifier are used to classify the data obtained from feature extraction [7–15].

2. Related work

Breast cancer disease is considered as a most quickly expanded malignancy among ladies in western nations and all the created urban communities in India. The American Cancer Society [16–18] gauges that around 230,480 ladies in the US will be determined to have breast tumor, and 39,520 ladies will bite the dust from breast growth. An ongoing report by National Cancer Registry Programs tell the "Breast growth represents 28–35% of all diseases among ladies in significant urban communities (Delhi, Mumbai, Ahmedabad, Chennai and so on.) And it is expanding quickly in vast figures". Mammography, biopsy and biopsy needle, these three strategies for the most part used to recognize breast tumor. The initial step is mammography for recognition of breast tumor [19,20]. Ordinary speaks to mammogram with no harmful cell, favorable speaks to mammogram demonstrating a tumor however not created by dangerous cell and disease speaks to tumor delivered by carcinogenic cell. It is troublesome assignment to recognize among every one of the three classes. Ongoing utilization of textural models and machine learning classifiers has built up another examination heading to recognize breast malignancy. Numerous analyst in the past have utilized a particular ROI for surface examination [21].

ROI in mammogram image is segmented into maximum possible number of non-overlapping small squared shape region of fixed size to acquire a large dataset for the further studies. A typical mammogram classification system generally consists of three sequential steps: (1) Extraction of region of interest, (2) features extraction from selected ROI, and (3) classification of mammogram based on extracted features.

3. Proposed methodology

The mammograms are preprocessed in the early stage and this is increase the difference between needed objects and unwanted background noise. Intensity is the parameter measured here. Preprocessing is done because of the low contrast of mammographic images that is hard to interrupt masses in the mammograms.

Generally there is no much variation in the intensities of pectoral muscle when compared to the tumor intensity. So we should remove pectoral muscle region before feature extraction. Preprocessing stage is very much used to remove the labels and background noise in the digital mammograms. After removing all unwanted labels and noise in the mammogram Hough transform is applied to this processed image which is similar to random transform. It is mainly used to detect arbitrary shapes and straight lines. Hough transform is tolerant of gaps noise and occlusion in the mammograms. The feature should be extracted effectively and it should be separated. The Fig. 1 shows the proposed block diagram Figs. 2–9.

3.1. Description

This section explains the techniques used to classify digital mammogram which has the process like,

- Image acquisition
- Pre-processing
- Feature extraction using Hough transform
- Classification using SVM

In the early stage mammogram images were collected and processed below.

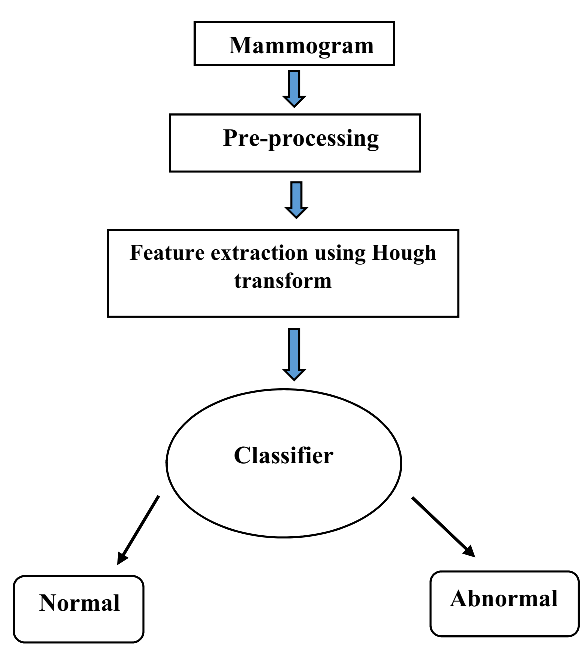

Fig. 1. Proposed method.

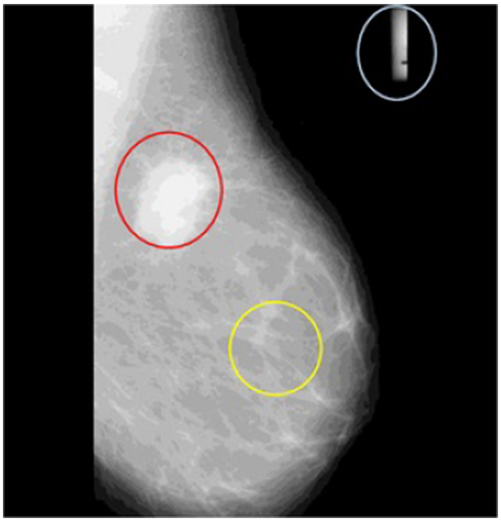

Fig. 2. Elements of mammogram image.

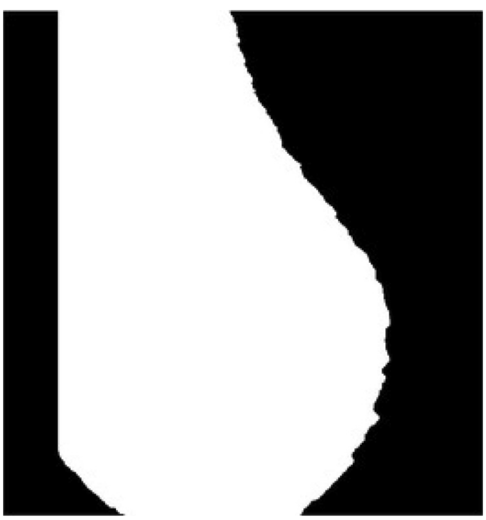

Fig. 3. Label removed binary image – using gradient based threshold.

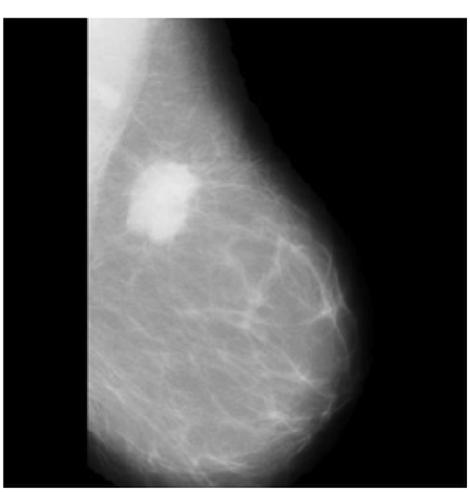

Fig. 4. Label removed mammogram image.

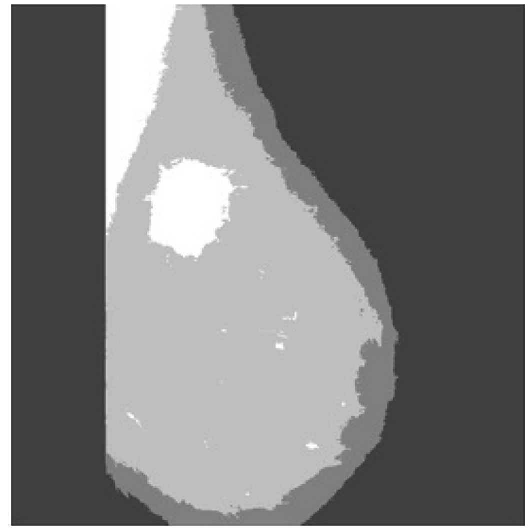

Fig. 5. Mammogram image – estimation maximization applied.

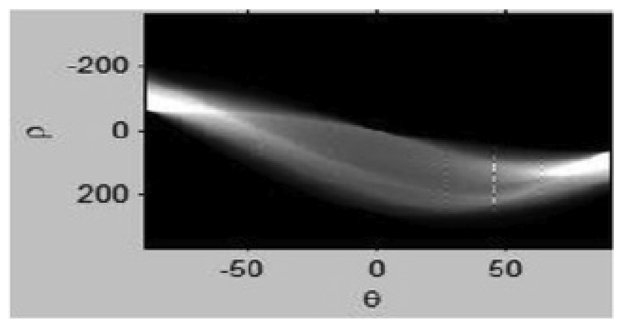

Fig. 6. Pre-processed mammogram image without edge detection.

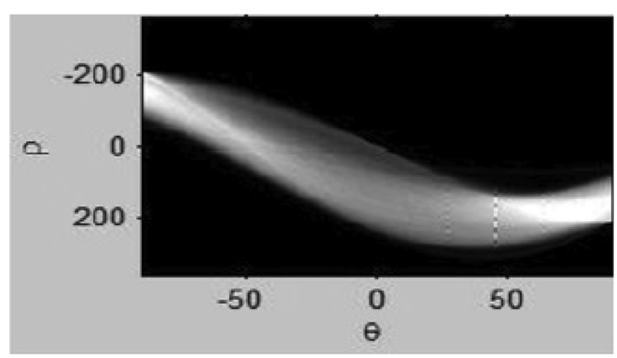

Fig. 7. Pre-processed mammogram image with canny edge detection.

3.1.1. Image acquisition

Mammogram images of type normal, benign and malignant for fatty-glandular breast types are collected form the MIAS database-mammographic image analysis database society. It has mammogram images digitized at 50 μm pixel edge. It has been reduced and clipped so that every image is of 1024 * 1024 pixels. There are 322 mammograms images from this 95 images taken for this work. Cancer detection in the dense breast cancer is difficult to find thorough mammogram. So dense images was not taken in this.

3.1.2. Pre-processing

The images are collected from database will have unwanted information and background noise. Pre-processing stage is mainly

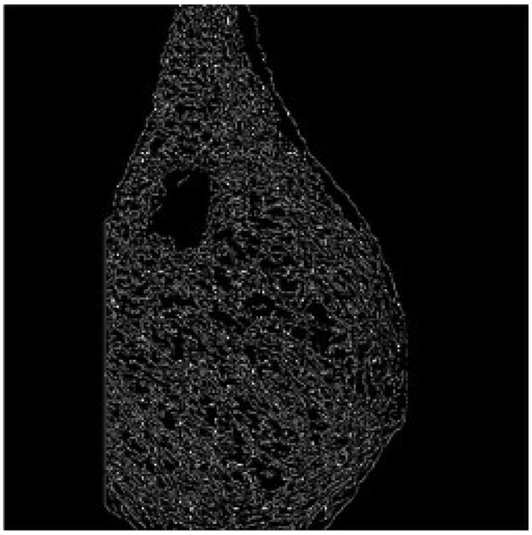

Fig. 8. Mammogram image with edge detection.

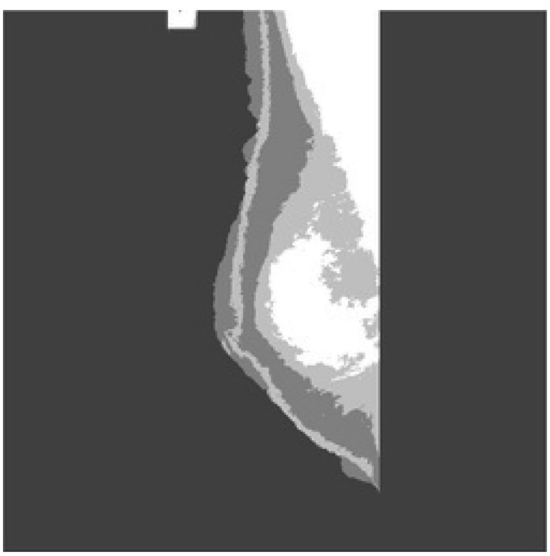

Fig. 9. Wrongly classified image after maximization estimation.

used to remove these from the mammogram and make image more suitable for further process. The figure below represents the unwanted region of breast cancer and tumor.

The region in yellow is the breast region, unwanted label and background is represented using the blue circle and red circle – which represents as tumor. The unwanted label should be removed first by using gradient based threshold method. Several morphological operations was carried out to generate a mask. Dilation and hole-filling are the main operations used here and by utilizing some reference we generated a binary edge map of the image using gradient based threshold method.

$$f(x,y) = \begin{cases} 1 \; if \; G(x,y) \geq GT \\ 0 \;\; else \end{cases} \tag{1}$$

Here GT is a gradient threshold which is found using Otsu's adaptive method. After this the binary image is dilated using diamond structuring element. Now this mask is multiplied with original image. The below two images shows mask generated through gradient method and other is label removed image. The main intention of the dilation operation in the binary image is to bring maximum breast region for processing it in the next stage. The main step in the upcoming process is to remove the pectoral image in the mammogram.

These muscles have same or closer intensity value compared with tumor intensity. So this muscle should be removed from the image for efficient feature extraction using maximization method which is also called as segmentation techniques. This will classify data values based on the maximum likelihood condition.

For example, if we want to classify the whole image in to two intensity classes maximization estimation is used to check whether the contextual pixels is classified in to class one or two based on the maximum likelihood condition. The main criteria for maximization estimation is given below.

$$\mathrm{L}\theta = Ln(\mathrm{P}(\mathrm{X}\theta)) \tag{2}$$

By using this equation maximization is possible through maximum likelihood function. For removing pectoral muscle and to retain the rest of the breast region four intensity class segmentation is done on mammogram image using maximum estimation. As mentioned before, the pectoral muscle and tumor region will come under the same class as shown below.

3.1.3. Feature extraction using Hough transform

It is feature extraction technique used in the digital signal processing to estimate the shape parameters from its boundary points. So Hough transform is used for the detection of arbitrary shapes. The normal parameterization is given by

$$x\cos\theta + y\sin\theta = \rho \tag{3}$$

Hough transform is a tolerant of gaps in edges and unaffected by noise and it is a derivation of random transform. It gives projections form different angle. Canny edge detection is used in the pre-processed image before applying the Hough transform. It also presents optimal edge detection filter to isolate step edges using the first derivative of the Gaussian.

This edge detection operation can reduce the pre-processing time and provides a relatively stable data source which resists geometrical and environmental changes for calculating Hough transform. It takes values for each edge point (X, Y) in the image calculated in the below equation. Also for non-analytical space is calculated in the Eq. (5) with the specified set of boundary points. For a shape ρ it is defined as P in the Eq. (6).

$$\rho = X_i \cos\theta + Y_i \sin\,\theta \tag{4}$$

$$B = \{X_B\} \tag{5}$$

$$p = \{X^o, s, \theta\} \tag{6}$$

For each value of X_B, r is calculated below and it is stored as a function φ. The value for r for each pixel X of gradient direction $\varphi(X)$ in an image is calculated in the Eq. (8) and it is stored in the accumulator.

$$r = X^o - X_B \tag{7}$$

$$A(X + r) \tag{8}$$

From the accumulator we will get the Hough transform output for the pre-processed mammogram. Now we have to select features from the transform image. The features should selected carefully because sometimes features may reduce the efficiency of the classifier. In this work intensity features are selected and the reason for using this features is because of the difficulty in interpreting the shape of the tumor. There may be well defined masses, speculated mass, ill-defined mass, architectural distortion, asymmetry and others. For here we going for intensity based features for better results. The intensity features used in this work are mean, variance, entropy and standard deviation.

4. Classification

Sample mammogram are taken form the specified database for all classes like malignant, benign and normal. After obtaining features from mammogram image the values are given to the classifier called SVM-Support vector machine. It is a classifier used here to attain better efficiency than other classifiers. It aims on reducing bounds on the generalization error (error made by the learning machine data unseen during training phase) rather than minimizing the mean square error over the data set. As a result SVM led to perform well when data applied outside the training set. It attains the accuracy of 94% which is higher when compared with all other classifiers. The SVM classifier for a mammogram is obtained below in the Eq. (9). Values obtained from the mammogram images of both cancerous and non-cancerous is used to determine a maximum margin hyper plane between the two classes.

$$f_{SVM}(X) = W_t \varphi(X) + b \tag{9}$$

Where

W_t – Weight vector

φ – Mapping function – this mapping function is used to map any input function to another dimension space. It is done for easy separation. The hyper plane is determined in such way that the distance from this hyper plane to nearest data points on each side called support vectors. SVM uses the values of train images and using these values it will easily classify the test images.

5. Results and discussion

This work provides the more accurate results with respect to the Hough transform applied to the mammogram image. The experimental results was discussed below. The mentioned method is applied to 95 mammogram image with normal and abnormal images. The Hough transformed output after applying the canny operator for edge detection is shown below. Tables 2 and 3 shows the parameters obtained with and without edge detection using canny method. Tables 4 and 5 shows the results after applying the classifier for with and without edge detection.

Table 2
Mammogram image without edge detection – Parameter table.

Sl. No.	Parameters	Image 1	Image 2
1	Mean	3.165 e^{+003}	5.4902 e^{+003}
2	Variance	5.4902 e^{+003}	1.0600 e^{+008}
3	Entropy	0.9239	0.9502
4	Standard deviation	6.6162 e^{+003}	1.0286 e^{+004}

Table 3
Mammogram image with edge detection – Parameter table.

Sl. No.	Parameters	Image 1	Image 2
1	Mean	4.4605	5.9196
2	Variance	73.88	116.06
3	Entropy	0.877	0.886
4	Standard deviation	8.595	10.77

Table 4
Result – without edge detection.

Sl. No.	Class	Trained	Tested	Detected
1	Normal	35	35	13
2	Abnormal	35	35	20

Table 5
Result – with edge detection.

Sl. No.	Class	Trained	Tested	Detected
1	Normal	35	35	23
2	Abnormal	35	35	25

6. Conclusion

In this work breast cancer detection using maximization estimation is done to acquire more accuracy. Increasing the intensity class in the estimation maximization produce better results. Usual shape features can't be used for this purpose because we are considering the entire image for feature extraction and classification. Also by using Hough transform normal and abnormal classes are effectively classified. Use of more intensity features like mean, variance and entropy can improve the results. By having SVM classifier we obtained the accuracy range of 94% which is higher when compared with other classifier like LDA, it has only 86% of accuracy.

References

[1] "Cancer facts and figure", http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancerkey-statistic (2011)-(2011).
[2] P. Parthasarathy, S. Vivekanandan, A comprehensive review on thin film-based nano-biosensor for uric acid determination: arthritis diagnosis, World Rev. Sci., Technol. Sustainable Dev. 14 (1) (2018) 52–71.
[3] P. Parthasarathy, S. Vivekanandan, A typical IoT architecture-based regular monitoring of arthritis disease using time wrapping algorithm, Int. J. Comput. Appl. (2018) 1–11.
[4] J. Blagojce, K. Ivan, T. Katarina, D. Ivika, L. Suzana, Mammographic Image Classification Using Texture Features, 9th Conference for Informatics and Information Technology, 2012.
[5] P. Parthasarathy, S. Vivekanandan, Investigation on uric acid biosensor model for enzyme layer thickness for the application of arthritis disease diagnosis, Health Inf. Sci. Syst. 6 (2018) 1–6.
[6] R. Varadharajan, M.K. Priyan, P. Panchatcharam, S. Vivekanandan, M. Gunasekaran, A new approach for prediction of lung carcinoma using back propagation neural network with decision tree classifiers, J. Ambient Intell. Hum. Comput. (2018) 1–12.
[7] N. Karssemeijer, Automated classification of parenchymal patterns in mammograms, Phys. Med. Biol. 43 (2) (1998) 365–389.
[8] S. Petroudi, T. Kadir, M. Brady, Automatic Classification of Mammographic Parenchymal Patterns: A Statistical Approach, Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancum, Mexico, 2003, pp. 798–801.
[9] A. Oliver, J. Freixenet, R. Zwiggelaar, Automatic Classification of Breast Density, in: In Proceedings of the IEEE International Conference on Image Processing, ICIP, 2005, pp. 1258–1261.
[10] K. Bovis, S. Singh, Classification of Mammographic Breast Density using a Combined Classifier Paradigm, In medical image understanding and analysis (MIUA) conference, Portsmouth(C), 2002, pp. 1–4.
[11] M.G. Mustra, K. Delac, Breast density classification using multiple features selection, AUTOMATIK, J. Control Meas., Electron. Comput. Commun. 53 (4) (2012) 362–372.
[12] T.V. Padmavathy, M.N. Vimalkumar, S. Nagarajan, G.C. Babu, P. Parthasarathy, Performance analysis of pre-cancerous mammographic image enhancement feature using non-subsampled shearlet transform, Multimedia Tools Appl. (2018) 1–16.
[13] Z. Chen, E. Denton, R. Zwiggelaar, Local feature based mammographic tissue pattern modelling and breast density classification, InProceedings of 4th International Conference on Biomedical engineering and Informatics, Shanghai, 2011, pp. 15351–17355.
[14] Y. Qu, C. Shang, Q. Shen, Evolutionary Fuzzy extreme learning machine for mammographic risk analysis, Int. J. Fuzzy Syst. 13 (4) (2011) 282–291.
[15] P. Parthasarathy, S. Vivekanandan, Urate crystal deposition, prevention and various diagnosis techniques of GOUT arthritis disease: a comprehensive review, Health Inf. Sci. Syst. 6 (1) (2018) 19.

[16] S. Lokesh, P.M. Kumar, M.R. Devi, P. Parthasarathy, C. Gokulnath, An automatic Tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map, Neural Comput. Appl. (2018) 1–11.
[17] B. Kanisha, S. Lokesh, P.M. Kumar, P. Parthasarathy, G. Chandra Babu, Speech recognition with improved support vector machine using dual classifiers and cross fitness validation, Pers. Ubiquitous Comput. (2018) 1–9.
[18] P. Parthasarathy, S. Vivekanandan, A numerical modelling of an amperometric-enzymatic based uric acid biosensor for GOUT arthritis diseases, Inf. Med. Unlocked (2018).
[19] P.M. Kumar, S. Lokesh, R. Varatharajan, G.C. Babu, P. Parthasarathy, Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier, Future Gener. Comput. Syst. 86 (2018) 527–534.
[20] K. Mathan, P.M. Kumar, P. Panchatcharam, G. Manogaran, R. Varadharajan, A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease, Des. Autom. Embedded Syst. (2018) 1–18.
[21] A.A. Basha, S. Vivekanandan, P. Parthasarathy, Evolution of blood pressure control identification in lieu of post-surgery diabetic patients: a review, Health Inf. Sci. Syst. 6 (1) (2018) 17.