

REVIEW

Editorial Process: Submission:03/01/2018 Acceptance:10/27/2018

An Effective Two Way Classification of Breast Cancer Images: A Detailed Review

Sinthia P*, Malathi M

Abstract

Cancer, a disease of cells, causes cell growth which differs from normal cell growth ratio, this cell growth spreads in the human body and kills the body cells. Breast cancer, it's a highly heterogeneous disease and western women commonly witness this. Mammography, a pre-screening X-ray based check is used to diagnose woman's breast cancer. This basic test mode helps in identifying breast cancer at early stage and this early stage detection would support in recovering more number of women from this serious disease. Medical centres deputed highly skilled radiologists and they were given the responsibility of analysing this mammography results but still human errors are inevitable. An error frequency ratio is high when radiologists exhausted in their analysis task and leads variations in either observations i.e., internal or external observation. Also, quality of the image plays vital role in Mammographic sensitivity and leads to variation. Several automation processes were tried in streamlining and standardising diagnosis analysis process and quality of breast cancer images were improved. This paper inducts a two way mode algorithm for grouping of breast cancer images to 1. benign (tumour growing, but not dangerous) and 2. malignant (cannot be controlled, it causes death) classes. Two-way mode data mining algorithms are used due to thinly dispersed distribution of abnormal mammograms. First type algorithm is k-means algorithm, which regroups the given data elements into clusters (i.e., prioritized by the users). Second type algorithm is Support Vector Machine (SVM), which is used to identify the most suitable function which differentiates the members based on the training data.

Keywords: Mammogram- breast cancer- k-means- SVM

Asian Pac J Cancer Prev, **19 (12)**, 3335-3339

Introduction

Unlimited multiplication of a specific group of cells in a particular area of the human body is referred as CANCER. A lumps or mass of an additional tissue will be formed on a group of divided cells which splits quickly. These lump or masses are identified as Tumours. Cancer cells are known as malignant tumours. Breast cells were the basis for formation of malignant tumour, known as Breast Cancer. Clusters of micro calcifications, architectural distortions and masses are notable and cautionable signs. The growth ratio of breast cancer seems to be reported very high in present years. At the same time, survival rate is also increased potentially over past, which is majorly due to improved efficiency in diagnosis and treatments.

Screening of Breast Cancer is primarily taken as anatomic approach through X-ray mammography, which requires the breast tumor to have developed to a stage where it is significantly more dense than healthy tissue. As a consequence, mammography misses 5%–15% of nonpalpable breast lesions that are not sufficiently denser than healthy tissue (Manoharan et al., 1998; Osteen et al., 1996). In addition, increased density is not always

tied to the presence of cancer: dense lesions of tissue that are further investigated via biopsy are often found to be benign (Manoharan et al., 1998). Instead of relying on density changes, cancer can also be detected by using early molecular signatures.

In 2011 United States, the American Cancer Society had come out with their pre-analysis report that nearly 230,480 fresh cases of invasive breast cancer and nearly 57,650 fresh cases of non-invasive breast cancer would be in treatment for breast cancer. 39,520 women would die out of the total affected cases. Mammography, a famous and well-known process in diagnosing Breast Cancer uses low-dose X-rays, high-contrast and high-resolution detectors and the X-ray system is designed exclusively to image the breasts. In Breast Cancer Screening and diagnosis, it is understood that Mammography serves the purpose of application. Mammography is of two types, 1. Screen Film Mammography (SFM) – film screen is acting as an end recording device and 2. Full-Field Digital Mammography (FFDM) – digital detectors acting as an end recording media. In Image Processing and further grading support, the FFDM produced digital images have more advantages rather traditional film screen.

Department of EIE, Saveetha Engineering College, Chennai, India. *For Correspondence: sinthiapanneerselvam@gmail.com

Digital Mammogram is one of the important methods to identify the Breast Cancer at an early stage at some extent. The advantages of digital mammography include the lack of ionizing radiation, its non-invasiveness, the relatively compact instrumentation, and its cost-effectiveness. As Siddiqui et al., (2005) mentioned, Mammography is very effective and the results were highly reliable in identifying breast cancer and it's proven, a minimum number of radiologists were tasked in interpreting and diagnosis of Mammograms which is more by and large from population screening. It was mentioned in the report by Wroblewska et al., (2003) that there is always a risk missing breast cancer cases, involved in mammographic image observance because unusual identifications were embedded and hidden by variance in structures of breast tissue.

Related Work

a. Studies on different techniques

Abou-Chadi et al., (2002), taken support of neural networks in identifying candidate circumscribed lesions in digitized mammograms. Back Propagation algorithm is used in training neural networks. The process of neural networks majorly differentiates the histogram of cancerous tissue and the normal tissue.

Brake et al., (1999) noted in his studies how digital mammograms were used in single and multi-layer detection of masses. In mammograms, scaling plays a vital role in automated process of detecting masses, it is mainly because of the possible range of masses can have. The work carried out, was experimented that if detection of masses can be done in single scale or might be suitable to use the result at various levels of scaling in multi-scale scheme.

Chan et al., (1988), has done research and introduced

a computerized mode of detecting micro-calcification in digital mammograms. This mode works on variation in image in which the signal suppressed image is subtracted from a signal enhanced image, which removes the structured background in the mammogram. For extracting micro-calcification signals, global and local threshold values based techniques are used.

Karssemeijer (1905) has done his studies and developed a data based calculate method for detecting of micro-calcifications in digital mammograms. Bayesian image analysis is base for the statistical models and general framework.

Nakayama et al., (2005) had taken support of filter bank in identifying linear and nodular patterns. The sub images were generated with the elements of a Hessian matrix at all resolution level with support from filter bank. The small and eigen values were calculated and a new filter bank resulted with three properties, follows, 1. Nodular patterns can be enhanced with various sizes, 2. Various sizes can be enhanced in both nodular and linear patterns and 3. By removing these patterns, an original image can be re-build. In mammograms, filter bank is used in enhancing micro-calcifications.

Yu et al., (2000) has given in proposal that two steps of CAD system for the automatic clustered micro-calcifications detection. In first step, wavelet and gray level statistical properties were used in potential micro-calcification pixels segment and establish them into objects of potential individual micro-calcification. In second step, 31 statistical properties were used to check these potential objects. Enough support was taken from Neural Networks too. The outcome results were promising but not guaranteed, it's due to training set usage in testing.

Kim (2016) has proposed a new algorithm weighted KM-SVM to circumvent the issues of the conventional

Table 1. Studies on Mammographic Image Classification.

Authors	Features	Classifier Algorithms	Accuracy (%)
Acharya et al., (2008)	Area, Homogeneity, Microcalcification	ANN (Artificial Neural Network) and GMM (Gaussian Mixture Model) (Multi-class classification)	ANN – 88.9, GMM – 94.4.
Andre and Rangayyan, (2003)	Shape factor measures, GLCM features	Multi-class classification (Perceptrons with several topologies)	Shape factors – 99, Texture feature – 63.
Chitre et al., (1993)	Texture measures	Neural Network Classifier (Multi-class classification)	87
Dehghan et al., (2008)	Wavelet features, gray level statistical features	SVM classifier with RBF kernel (Multi-class classification)	89.5
Ganesan et al., (2013)	Texture measures	Decision Tree and SVM	96
Kinoshita et al., (1998)	Shape and texture features	Three layer feed-forward neural network (Multi-class classification)	81
Priebe et al., (1994)	Fractal texture measures	Finite mixture model probability density estimation (Multi-class classification)	88
Rangayyan et al., (1997)	Region based edge-profile	Acutance measures (Multi-class classification)	92
Verma et al., (2005)	Statistical features	Fuzzy Neural Network (Multi-class classification)	83
Wei et al., (2001)	Statistical features	SVM and Kernel Fisher Discriminant (Multi-class classification)	85
Kim (2016)	Statistical features	Weighted k-means	85
Zheng et al., (2014)	Statistical features	Hybrid k-means	85
Sridevi and Murugan (2014)	Statistical features	k-means clustering	85
The current study	Two way classification	K-means algorithm and SVM algorithm.	Not yet implemented. This study is more accurate.

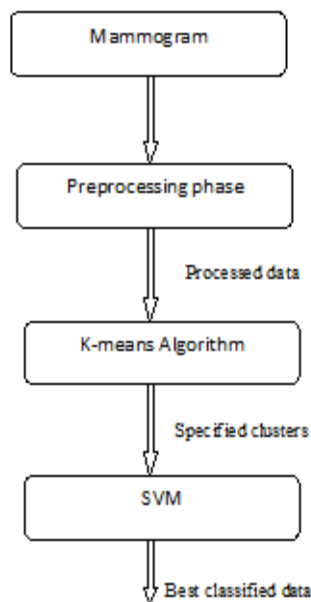


Figure 1. System Architecture

SVM algorithm. Initially the weighted SVM was performed and executed which made tractable process to classify the data set which can address the disease sub type identification.

Zheng et al., (2014) insisted that when the same data mining algorithm is applied to the same data set, the output may differ. Based on SVM classifier, three approaches, including GA, ant colony optimization (ACO) and particle swarm optimization (PSO), were utilized for selecting the most important features in the data set to be trained by the classification model.

Sridevi and Murugan (2014) introduced that Rough set theory is often applied to feature reduction using the data alone, requiring no additional information and widely used for classification tool in data mining. k-means clustering algorithm is applied to partition the given information system and further rough set theory was implemented on the data set to generate feature subset. The classification process by means of SVM is performed by using the remaining features. Wisconsin Breast Cancer datasets derived from UCI machine learning database are used for the purpose of testing the proposed hybrid model and the success rate of hybrid model is determined as 99%.

b. Studies on Mammographic Image Classification Algorithms used

a. K-means Algorithm

A set $D = \{x_i \mid i = 1, \dots, N\}$, where x_i denotes the i th data point

\Rightarrow Set of d -dimensional vectors. The process was initiated with k points chosen from the initial k cluster data or “centroids”. The initial value was taken by using sampling at random on dataset, fixing it as the clustering solution, a small data subset or unsettled global mean of k times data.

Repeat this algorithm process till convergence,

Step 1: Assigning Data from set D

Every data point from set D is assigned to its closest centroid, with ties arbitrarily broken. Data partitioning is resulted.

Step 2: “means” Relocation

Every cluster representative data is replaced to the center (mena) of all the data points assigned to it. The replacement is to the expectations (weighted mean) of the data partitions taking place if the data points reached the probability measure (weights).

Euclidean distance is the default measure of closeness, during this scenario, non-negative cost function applies always,

$$\sum_{i=1}^N \left(\underset{j}{\operatorname{argmin}} \left\| X_i - C_j \right\|^2 \right)$$

b. Support Vector Machines

Support Vector Machines (SVM) (Vapnik, 1995) is very popular application mode because it results in robust, reliable and accurate method, while comparing with other process and algorithms. Numbers of dimensions are insensitive and needs only 12 examples for training, sound theoretical foundation. Moreover, improved methods are developed at rapidly in training SVM.

In order to determine a maximum margin hyperplane, Support Vector Machine takes data from two classes. The hyperplane is resulted in, distance from the hyperplane taken and to the next nearest data points on either side, which is called support vectors, bring the results maximum. In support from Kernel function application, non-linearly separable data to make them linearly separable – to perform this Support Vector Machine (SVM) is extendable (Muller et al., 2001). We have taken support from the linear kernel in this paperwork, polynomial kernel of orders 1, 2 and 3 and the radial basis function kernel. Similar kernel techniques were quoted and used in wavelet SVM (Shen et al., 2010).

System Architecture

Mammogram result is taken as an input and given to preprocessing phase for filtering the data. In low-level image processing, pre-processing becomes an inevitable problem. By using various filtering techniques, noise presented in the image can be filtered out. The gray level of an image reduces while high pass filter passes the changes to a low pass filter. It means, the value smoothens and sharp edges were removed frequently, while applying low pass filter. The Median Filter is the best of low pass filters. The filter considers an image of area 3×3 , 5×5 , 7×7 , etc., an element array is resulted by taking all the pixel values. The median value of an array is calculated and resulted by ordering element array. A famous sorting technique, Bubble sort is used in this element array sorting in an Ascending order, which returns a median value from the middle elements of the sorted array. The set, the median values of the array elements calculated for all the pixels, were resulted to an output image array (Gonzalez and Woods, 2007). The complete image array is arrived by repeating the Median Filter process.

End of pre-processing phase, all the processed data is fed into first classification algorithm (i.e. k-means algorithm). With the help of k-means algorithm, processed data can be converted into specified clustered data. Then clustered data is given as input in SVM algorithm and produces best classified data.

In conclusion, Extensive Review has been made from various papers and the modified versions of the conventional k-means and SVM have been reviewed. The review showed that these techniques are found to have proved to be a novel framework for two-way classification methodology in mammographic image analysis. From a study of the available literature, we find that the application of two way classification to the problem of mammographic image analysis is rare. We strongly believe that the proposed system's performance can be scaled up and further enhanced by framing new functional features that are more adaptable to mammograms.

This article presents a very general overview of two way classification architecture. It demonstrates how an abstract structure allows us to discover effective classification of breast cancer images.

This algorithm will be implemented in future because it is simple and the results were encouraging, this will lead to a real-time breast cancer diagnosis system.

References

- Acharya UR, Ng EYK, Hong Y, Jie Y, Kaw GJA (2008). Automatic identification of breast cancer using mammogram. *J Med Syst*, **32**, 499–507.
- Andre TCSS, Rangayyan RM (2003). Classification of tumors and masses in mammograms using neural networks with shape and texture features. in Proc. 25th Ann Int Conf IEEE EMBS, **3**, pp 2261–4.
- Chan HP, Do K, Vyborny CJ, Lam KL, Schmidt RA (1988). Computer-aided detection of microcalcifications in mammograms methodology and preliminary clinical study. *Invest Radiol*, **23**, 664–71.
- Chitre Y, Dhawan AP, Moskowitz M (1993). Artificial neural network based classification of mammographic microcalcifications using image structure features. *Int J Pattern Recognit Artif Intell*, **7**, 1377–1402.
- Dehghan Abrishami-Moghaddam H, Giti M (2008). Automatic detection of clustered microcalcifications in digital mammograms: Study on applying adaboost with SVM-based component classifiers,” in Proc. 30th Annu Int Conf IEEE EMBS, pp 4789–92.
- Ganesan K, Acharya R, Chua KC, Min LC, Abraham KT (2013). Decision support system for breast cancer detection using mammograms. *Proc Inst Mech Eng Part H J Eng Med*, **227**, 21–732.
- Gonzalez RC, Woods RE (2007). Digital Image processing. Prentice Hall.
- Guido M, Brake T, Karssemeijer N (1999). Single and multiscale detection of masses in digital mammograms. *IEEE transactions on medical imaging*, **18**, 7.
- Karssemeijer N (1993). Recognition of clustered microcalcifications using a random field mode, biomedical image processing and biomedical visualization. *Proc SPIE*, **1905**, 776–86.
- Kinoshita SK, Marques PMA, Slaters AFA, et al (1998). Detection and characterization of mammographic masses by artificial neural network,” in Proc. 4th Int. Workshop Digit Mammography, pp 489–90.
- Manoharan R, Shafer K, Perelman L, et al (1998). Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging. *Photochem Photobiol*, **67**, 15–22.
- Mencattini A, Salmeri M, Lojaco R, Frigerio M, Caselli F (2008). Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. *IEEE Trans Instrum Meas*, **57**, 1422–30.
- Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B (2001). An introduction to kernel based learning algorithms. *IEEE Trans Neural Netw*, **12**, 181–201.
- Nakayama R, Uchiyama Y (2005). Development of new filter bank for detection of nodular patterns and linear patterns in medical images. *Sys Comput Japan*, **36**, 13.
- Osteen RT, Connolly JL, Costanza ME, Harris JR, Hayes DF (1996). Cancer of the breast, in Cancer Manual, 9th ed. New York: Am. Cancer Soc., pp 320–39.
- Pisano ED, Gatsonis C, Hendrick RE, et al (2005). Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*, **353**, 1773–83.
- Pisano ED, Hendrick RE, Yaffe MJ, et al (2008). Diagnostic accuracy of digital versus film mammography: Exploratory analysis of selected population subgroups in DMIST. *Radiology*, **246**, 376–83.
- Polat K, Genes S (2007). Breast cancer diagnosis using least square support vector machine. *Digit Signal Process*, **17**, 694–701.
- Priebe CE, Lorey RA, Marchette DJ, Solka JL, Rogers DW (1994). Nonparametric spatio-temporal change point analysis for early detection in mammography. in Proc. 2nd Int. Workshop Digit. Mammography, **10**, 111–20.
- Rangayyan RM, El-Faramawy NM, Desautels JEL, Alim OA (1997). Measures of acutance and shape for classification of breast tumours. *IEEE Trans Med Imag*, **16**, 799–810.
- Schulz-Wendtland R, Fuchsjgerb M, Wackerc T, Hermannnd KP (2009). Digital mammography: An update. *Eur J Radiol*, **72**, 258–65.
- Shen M, Lin L, Chen J, Chang CQ (2010). A prediction approach for multichannel EEG signals modeling using local wavelet SVM. *IEEE Trans Instrum Meas*, **59**, 1485–92.
- Siddiqui M, Anand M, Mehrotra P, Sarangi R, Mathur N (2005). Biomonitoring of organochlorines in women with benign and malignant breast disease. *Environ Res*, **98**, 250-7.
- Sridevi T, Murugan A (2014). An intelligent classifier for breast cancer diagnosis based on k-means clustering and rough set. *Int J Comput Appl*, **85**, 11.
- SungHwan K (2016). Weighted K-means support vector machine for cancer prediction. *Springerplus*, **5**, 1162.
- Tax D, Duin R (2001). Uniform object generation for optimizing one-class classifiers. *J Mach Learn Res*, **2**, 155–73.
- Vapnik V (1995) The nature of statistical learning theory. Springer, New York.
- Verma B, Zakos J (2001). A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE Trans Inf Technol Biomed*, **5**, 46–54.
- Wei L, Yang Y, Nishikawa RM, Jiang Y (2005). A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. *IEEE Trans Med Imag*, **24**, 371–80.
- Wroblewska A, Boninski P, Przelaskowski A, Kazubek M (2003). Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms, *Opto-Electron Rev*, **11**, 227–35.
- Youssry N, Fatma EZ, Abou-Chadi, Alaa M El-Sayad (2002). A neural network approach for mass detection in digitized

mammograms. ACBME, 2002.

Yu S, Guan L (2000). A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram 3films. *IEEE Trans Med Imag*, **19**, 115-26.

Zheng B, Yoon SW, Sarah SL (2014). Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithm. *Expert Syst Appl*, **41**, 1476 – 82.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.