Journal of Integrative Oncology
ISSN: 2329-6771
Selim and Eldesoky, J Integr Oncol 2019, 8:2
Research Article
Open Access

Automated Classification of Breast Cancer/Ultrasound Image Based on Non-Negative Matrix Factorization Algorithm

Selim IM[1,2]* and Eldesoky A[3]

[1]*National Research Institute of Astronomy and Geophysics, Cairo, Egypt*
[2]*Computer Science Department, Integrated Thebes Institutes, Cairo, Egypt*
[3]*Higher Technological Institute (HTI), Department of Biomedical Engineering, Cairo, Egypt*

Abstract

The topic of morphological analysis has received much attention with the increasing demands in different applications spatial in bioinformatics and biomedical applications. This paper summarizes the recent advances automated machine learning supervised suitable method for morphological Breast Cancer/masses Image Classification based on Non-Negative Matrix Factorization Algorithm. This scheme is making distinctions between all types roughly corresponding to Breast Cancer/masses types. Among many factors that morphological Classification of Breast Cancer and statistics have made great contributions for a radiologist and detection. Morphological (texture features, echogenicity, and homogenous). Breast masses analysis assists classification whether the mass is benign or malignant and finds the Breast Cancer shape. The Experimental results show that Breast Cancer images from the dataset can be classified automatically. With performance (average 94% accuracy) on a large-scale dataset, this demonstrates the strength of our method in providing an efficient tool for breast cancer multi-classification in clinical settings.

Keywords: Classification scheme; Breast cancer; Non-negative matrix factorization

Introduction

Cancer is a massive public health problem around the world. The breast cancer type, (BC) is the most common for women, excluding skin cancer. BC is the most common malignancy of women and is very high when compared to other types of cancer, the second most common and leading cause of cancer deaths among them. At present, there are no effective ways to prevent breast cancer, because its cause is not yet fully known. Early detection is an effective way to diagnose and manage BC can give a better chance of full recovery. Therefore, the classification of BC can play an important role in understanding the formation and evaluation it, to reducing the associated morbidity and mortality rates. Despite significant progress reached by diagnostic imaging technologies, the final BC diagnosis, including grading and staging, continues being done by pathologists applying visual inspection of histological samples under the microscope. Most of these recent works related to BC classification are focused on Whole-Slide Imaging [1-5]. However, the broad adoption of Whole-Slide Imaging and other forms of digital pathology still facing obstacles such as the high cost of implementing and operating the technology.

Image processing has been a very important subject in recent years in all fields of interesting. For an example of the image processing and image, recognition is optical characters of images and image retrieval, etc. Each digital image is a rectangular array of pixels and each pixel is represented by its light intensity. The light intensity of each pixel (element) always is measured by a nonnegative value. Medical imaging is the technique, process, and art of creating visual representations of the interior of a body for clinical analysis and medical intervention. Medical imaging seeks to reveal internal structures hidden by the skin and bones, as well as to diagnose and treat disease [6]. Medical image analysis techniques have played an important role in several medical interests [7]. The automatic imaging processing for Medical images plays an important role in aiding diagnosis and treatment; they also can be useful in the education domain for health care students by explaining these images will help them in their studies. A cancer diagnosis has been explored as a topic of research for more than 40 years ago. Classification of images into distinct patterns, corresponding to different types is often the primordial goal in image analysis systems for cancer automatic Classification applications. The progress in digital imaging technologies created a large growth in the number of digital images taken in recent years [8]. The interpretation and analysis of medical images represent an important and exciting part of computer vision. Recent advances in image processing and machine learning techniques allow building Computer systems to classifying the images to different types. The challenge of such systems is dealing with the complexity of images [1]. Morphological image processing (or morphology) describes a range of image techniques deal with the shape of features in image morphological operations. There are two main approaches to the challenging problem of image classification: object-based and scene-based. Matrix factorization is an example of a prevalently used scene-based classification method. There are many well-established matrix factorization algorithms, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Non-negative Matrix Factorization (NMF) [6]. These methods all learn to represent data as a linear combination of basis images; however, each algorithm factorizes the input into these basis vectors subject to different constraints.

In this work, we evaluate the deep learning approach for the problem of mass/cancer image classification. We investigate this method to deal with high-resolution images without changing the architecture used for low-resolution images. The best performance, though, is obtained by combining different.

***Corresponding author:** Selim IM, National Research Institute of Astronomy and Geophysics, Cairo, Egypt, Tel: 002-0104498964; E-mail: i_selim@yahoo.com

Received April 16, 2019; **Accepted** April 22, 2019; **Published** May 06, 2019

Citation: Selim IM, Eldesoky A (2019) Automated Classification of Breast Cancer/Ultrasound Image Based on Non-Negative Matrix Factorization Algorithm. J Integr Oncol 8: 229.

Related Works

Ultrasound is a technique that has advantages over other imaging techniques. There is no ionizing radiation or magnetic fields that may affect tissues [9-11]. So, that ultrasound images enable detecting breast tumors safely. But due to low-resolution, attenuation and noise, doctors preferred to use mammogram. To use Ultrasound for detecting dense tissue, it was needed to increase transducer elements [12]. Ultrasound is an excellent tool to organize masses and evaluate the progress of therapy, as doctors tend not to use mammography for young ladies especially, in case of pregnant females [13-15]. Ultrasound is preferred diagnostic imaging rather than mammogram especially for women with dense breast tissue. There are many attempts for automatic diagnosis of cancer from images through three-step; pre-processing, extraction of features, and diagnosis [16]. Techniques as Artificial Neural Networks, (ANN) [12], support vector machine [17,18] and Bayesian neural are used for these processes [19]. Tumour detection and classification using the large dataset for artificial neural network technique depends on features, a number of layers and backpropagation to correct error [20].

Support vector machine is also used for detection and classification, pattern recognition and field of bioinformatics [18]. It is a method for automatic classification and determination normal or abnormal based on coefficient and features from training data. The accuracy is calculated and guides the researcher to best feature used [21]. Using Fuzzy logic technique also used to eliminate noise and extract information from ultrasound image [22]. By defuzzification, ultrasound images enhancement can be processed [22].

The neuro-fuzzy depends on database and membership functions and parameter detected through learning algorithms of neural networks [22,23]. Using cross-correlation, it becomes easy to identify data so, reduce matrix size. Neuro-fuzzy obtains two advantages, first of self-learning and adaptation from the neural network and second, membership and logic functions. Image classification depends on features like texture, color, and shape for the extracted region [24]. This can be used for many applications where probability are accepted and where data available is limited [23,25]. Other techniques as Genetic Algorithms (GA), (GA) combined with Neural Networks, Fuzzy Support Vector Machines are also used for classification [26]. In this work, we improved a classification used Non-negative Matrix Factorization (NMF) method depending on the shape, texture features, echogenicity, and homogenous from the database.

Radiology Background Knowledge

Normal breast displays as fibro-glandular tissue with different ratio of fat layers [27]. Fibro glandular is echogenic surrounded by hypoechoic fat. Hypoechoic means an area looks darker on ultrasound than the surrounding tissue while Hyper means a lot of echoes look bright. The contour of mass, echogenicity and the shape of the mass are the main features used in classification. Fluid accumulates causes seen as round or oval, anechoic structure with a well-defined, thin wall. Blood or milk of calcium crystals may show internal echoes [28]. Breast benign lesions show Smooth and well-circumscribed texture. Rather than, Hyperechoic, or mildly hypoechoic and thin echogenic capsule. Oval shape with few lobules and wider than deep. Malignant lesions are commonly hypoechoic lesions with irregular borders. The malignant lesion is taller than broad. Small microsimulations and heterogeneous echo texture. Some blood vessels are seen. Finally, elastography shows benign lesions compress while malignant lesions don't change in height.

The Proposed Method

Morphological analysis is often studied as the shape appearances of objects and the surfaces of the images, with intensity, is seen as height and texture appearing as a relief. The formalisation of morphological features is of benefit to computerized calculation and more efficient than manual morphological quantification, which is still laborious and subjective. The morphology characteristics can be described by shape, geometrical, intensity, and texture analysis.

System architecture

In order to classify images from cancer dataset, each image could be represented as a non-negative matrix. The architecture of the system goes as follows:

1) Image pre-processing, where images are getting resized, scaled, rotated and centered. So, it becomes for the second phase

2) Feature extraction. Where morphological features are extracted from the images. This phase is generally used to minimize the dimensionality of Breast Cancer Image data

3) Machine learning

4) Classification

Performed method

As we know to make an object the status of parts is represented by nonnegative numbers, where 0 represents an absence and a positive number represents a presence with some degree. Furthermore, the objects are also represented by a set of nonnegative numbers, e.g., light intensities. Because of that, non-negativity is needed to maintain during the analysis of images. So, we used better tools (non-negative matrix factorization) to study the data stored in the matrix. Non-negative matrix factorization is a linear dimensionality reduction technique which is useful in handling nonnegative data [29]. It allows only non-subtractive combinations of nonnegative basis vectors, leading to (possibly) a parts-based representation and that`s what makes it distinguished from other methods like Principle Component Analysis by its non-negativity constraints. These constraints lead to part-based representation because they allow only additive not subtractive combinations of the original data [30]. Given an initial database of matrix V with dimensions where m is the number of examples in the dataset. Matrix V is factorized into to matrices W with dimensions and 'H' with dimensions, usually 'r' is chosen to be smaller than 'n' or 'm' so that the two matrices W and H becomes less than the original matrix 'V'.

A common way to find 'W' and 'H' is by minimizing the Euclidean distance between 'V' and 'WH' [31].

$\sum \sum$

Each non-negative inequality is a "bound constraint," as it relates to only a single variable. We also note that

$\sum \sum$

The first is the training step in which each training sample is normalized using–norm to make all images have the same scale. Then the NMF algorithm is used to decompose these sets into basis matrix W and coefficient matrix H. The second main step is the test or prediction step, where the test set S is normalized. The coefficient matrix of the S is computed based on basis matrix W, where the basis matrix W contains all information about different breast cancer types and these basis make the process of separating the images that belong to the same type is easy (like blind source separation).

To predict the class of an unknown sample, we used a MAX rule that selects the maximum coefficient in (the coefficient vector of), and then assign the class label of the corresponding training sample to this new sample. The pseudocode of our algorithm is given in algorithm 1.

Algorithm: NMF algorithm

(A) First step

Input:: a training set

: cluster numbers

: unknown samples without labels

Output:: predicted class labels of the p unknown samples

Training step:

1. Normalize training set to have-norm
2. Solve the NMF optimization problem:

[ ]=NMF( )

Test step:

1. Normalize test set to have-norm
2. Solve the NMF optimization problem:
3. Predict the class label:

Return

(B) Second step

Input: input matrix

Output: basis matrix, H coefficients matrix

Initial: ,

for =1, 2, ..., maximum iterations

1.

2. If

satisfies (7)

do

Until

does not satisfy (7) or

Else do //decreaseUntil

satisfies (7).

End IF

3. ( )

4. Can be easily obtained by repeating steps similar to (1)-(3). But in (1) consider and as input matrices.

Experimental Results

Dataset

Large database of breast mass cases is used. It contains 600 objects from different types (malignant, benign and normal). These images got from the diagnostic center. Cases include all ages. To examine different lesions, there are many objects per category that span large in-class variations while still clearly belonging to the category. Each object is represented by 200 images. This allows for analyzing the performance of different types.

These samples were prepared in digital images form and available in 3-channel RGB (Red-Green-Blue), True Colour (24-bit color depth, 8 bits per color channel) color space, and the dimension of 756 × 460 pixels. Figure 1 presents samples from this set show malignant and benign masses.

In which the input to this algorithm is training set and test set (where our algorithm does not need to dimension reduction as pre-processing

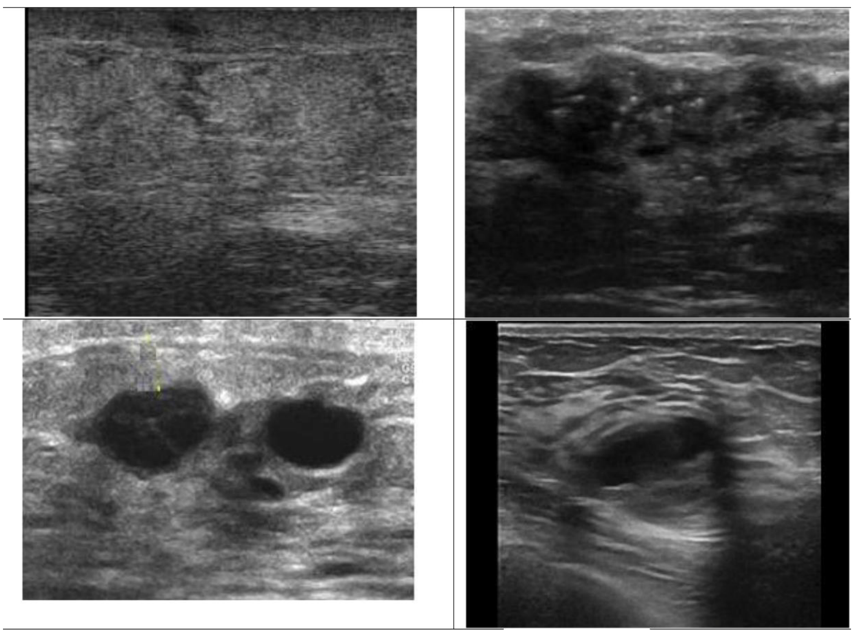

Figure 1: Image samples from database (Benign lesions and malignant).

step since (r<<min (n, m)). Also, the number of breast cancer type (groups) is input (i.e. if it is determined from prior knowledge about catalog). Otherwise, (if it is unknown) we determine the best number of breast cancer type by using the consensus matrix, where it reflects the probability that images and belong to the same type. It is defined as the average of connectivity matrix overruns, where the consensus matrix is defined as selected from all images as in (Table 1). Different percentage of training data set and test data set are used.

In the experiment, the grey-scale image is used to test images and normalized all images with respect to the intensity with the same standard deviation. The normalization ensured that the order will be determined by the shape, with no impact of color or brightness (Figure 1) [32].

On the training set of Breast mass image, using a set of images for testing and obtaining results as showed in Table 1, we obtained about 93% of correct classification within Breast Cancer Image. Preliminary experimental results using 10-fold cross-validation shows that the homogeneous ensemble of locally weighted regression produces the best results, with over 93% accuracy when considering images types. It is also noticeable that Breast Cancer Image types are different. This result indicates that the percent value different across different Breast Cancer Image morphological types shown in Figure 2a-2f. The obtained accuracy is satisfactory and is comparable to that of a specialist. Image-level accuracy simply corresponds to the score from the total number of correctly-classified images. We let *t_im* be the total number of images in the dataset, and *c_im* the total of correctly-classified images, image-level accuracy is defined as Image-level accuracy = *c_im /t_im*

Test data set	Training data set	Accuracy of result
90%	10%	95%
80%	20%	94%
70%	30%	91%

Table 1: Test, training data description and accuracy of result.

Conclusion

Automated classification of Breast Cancer Image methods are very important to understanding the properties of the past, present, and future of Breast Cancer, it is also identified and analyze Breast Cancer that cannot be associated with a defined morphological stage. In this study, we have proposed a computer-based approach to automatically classification the breast cancer morphology using the supervised machine learning system based on Non-Negative matrix factorization algorithm, that can derive the automatically classify images of Breast Cancer and masses. The growth of scope development and accurate database capturing is considered a freshly used technique in this area, satisfactory results were obtained comparing to results used by other authors using other methods taking in consideration the infancy of the area of Breast Cancer.

The detection mainly determined by the shape, with no impact of color or brightness with TP of 93 and accuracy of 94%. The accuracy is high compared to that used CNN and CAD system used where ROI input image is selected by the physician. The accuracy of that system was 88.80%.

Typical contributions are addressed for initialization, shape analysis for morphological Breast Cancer Image, extended morphological segmentation, are introduced. The results obtained show that automated grading is feasible and that discrimination between different levels of tubule formation can be performed with moderate to high accuracy. The analysis is performed such that the algorithm determines the morphology types of Breast Cancer from different morphological classes automatically, without human guidance. While the extreme morphological structures according to tubule formation in the breast cancer tissue are discriminated with high accuracy, the recognition of the intermediate class should still be improved. We assume that it will be reasonable to be used for other problems in morphological analysis.

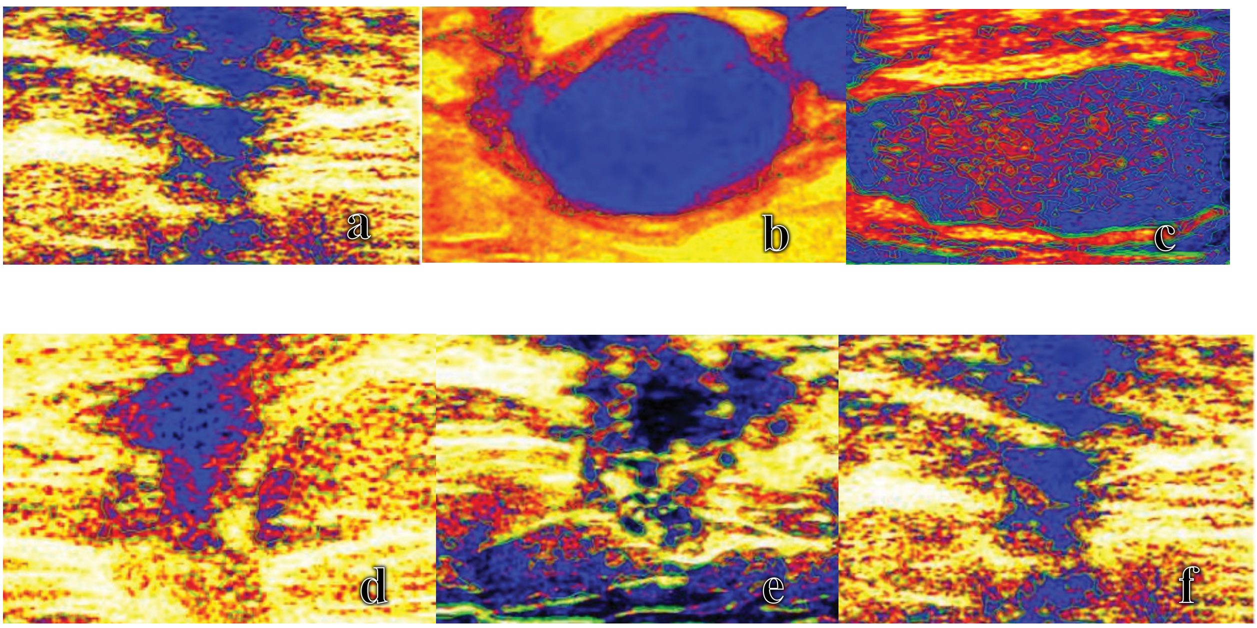

Figure 2: The morphological types of breast mass image (a): Benign lesions-inflammation; (b): Benign lesions-infected cysts; (c): Benign non-mass lesion; (d) Malignant lesions-carcinoma tumor types; (e): Malignant-breast carcinoma diagnostic criteria; (f): Malignant lesions-metastases.

References

1. Yajie S (2004) Normal mammogram analysis, PhD thesis, Department of Computer Science, University of Purdue, USA.
2. George YM, Zayed HH, Roushdy MI, Elbagoury BM (2014) Remote computer-aided breast cancer detection and diagnosis system based on cytological images. IEEE Sys J 8: 949-964.
3. Zhang Y, Zhang B, Coenen F, Lu W (2013) Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles. Mach Vision Appl 24: 1405-1420.
4. Zhang Y, Zhang B, Coenen F, Xiao J, Lu W (2014) One-class kernel subspace ensemble for medical image classification. EURASIP J Adv Signal Process 2014: 17.
5. Doyle S, Agner S, Madabhushi A, Feldman M, Tomazewski J (2008) Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. Proc IEEE Int Symp Biomed Imaging 61: 496-499.
6. Sunil R, Veena R (2014) Loss less data hiding technique for medical images using LSB substitution. ISRJ.
7. Pham TA (2010) Optimization of texture feature extraction algorithm. MSc Thesis, Delft University of Technology, Netherlands.
8. Kowal M, Filipczuk P, Obuchowicz A, Korbicz J, Monczak R (2013) Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Comput Biol Med 43: 1563-1572.
9. Hoskins PR, Martin K, Thrush A (2010) Diagnostic ultrasound: Physics and equipment. Cambridge University Press.
10. Narouze SN (2018) Atlas of ultrasound-guided procedures in interventional pain management. Springer.
11. Zagzebski JA (1996) Essentials of ultrasound physics. Mosby-Year Book. Chapter 7, pp: 123-145.
12. Cheng HD, Shan J, Ju W, Guo Y, Zhang L (2010) Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recognit 43: 299-317.
13. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, et al. (2015) Ultrasound as the primary screening test for breast cancer: Analysis from ACRIN 6666. J Natl Cancer Inst 108: 367.
14. Barr RG, Nakashima K, Amy D, Cosgrove D, Farrokh A, et al. (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 2: breast. Ultrasound Med Biol 41: 1148-1160.
15. Weinstein SP, Conant EF, Orel SG, Zuckerman JA, Bellah R (2000) Spectrum of US findings in pediatric and adolescent patients with palpable breast masses. Radiographics 20: 1613-1621.
16. Demir C, Yener B (2005) Automated cancer diagnosis based on histopathological images: A systematic survey. Rensselaer Polytechnic Institute, Tech Rep.
17. Piliouras N, Kalatzis I, Dimitropoulos N, Cavouras D (2004) Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound. Comput Med Imaging Graph 28: 247-255.
18. Huang YL, Chen DR (2005) Support vector machines in sonography: Application to decision making in the diagnosis of breast cancer. Clin Imaging 29: 179-184.
19. Chen DR, Chang RF, Chen CJ, Ho MF, Kuo SJ, et al. (2005) Classification of breast ultrasound images using fractal feature. Clinic Imaging 29: 235-245.
20. Chen DR, Chang RF, Huang YL (2000) Breast cancer diagnosis using self-organizing map for sonography. Ultrasound Med Biol 26: 405-411.
21. Abdullah N, Ngah UK, Aziz SA (2011) Image classification of brain MRI using support vector machine. In 2011 IEEE international conference on imaging systems and techniques. IEEE 242-247.
22. Guo Y, Cheng HD, Huang J, Tian J, Zhao W, et al. (2006) Breast ultrasound image enhancement using fuzzy logic. Ultrasound in Medicine and Biology 32: 237-247.
23. Dixit R, Singh H (2012) Comparison of detection and classification algorithms using boolean and fuzzy techniques. Adv Fuzzy Syst 2012: 5.
24. Panigrahi S, Verma T (2013) Texture image classification using neurofuzzy approach. Int J Eng Comput Sci 2: 2309-2313.
25. Khan SIA, Barnes CF (2011) Using residual vector quantization for image content classification. In 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE 1041-1044.
26. Seetha M, Muralikrishna IV, Deekshatulu BL, Malleswari BL, Hegde P (2008) Artificial neural networks and other methods of image classification. J Theor Appl Inf Technol 4.
27. Masciadri N, Ferranti C (2011) Benign breast lesions: Ultrasound. J Ultrasound 14: 55-65.
28. Klein S (2005) Evaluation of palpable breast masses. Breast J 100: 28-33.
29. Kowal M, Filipczuk P, Obuchowicz A, Korbicz J, Monczak R (2013) Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Comput Biol Med 43: 1563-1572.
30. Filipczuk P, Fevens T, Krzyzak A, Monczak R (2013) Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. IEEE Trans Med Imaging 32: 2169-2178.
31. Niu XX, Suen CY (2012) A novel hybrid CNN-SVM classifier for recognizing handwritten digits. Pattern Recognit 45: 1318-1325.
32. Shamir L, Orlov N, Macura T, Eckley DM, Johnston J, et al. (2008) Wndchrm-an open Source utility for biological image analysis. BMC Source Code Biol Med 3: 13.