Breast Cancer Diagnosis and Classification Improvement based on Deep Learning and image Processing

Mohsen Eftekharian[1], Ali Nodehi[2*]

[1] University of Applied Science and Technology, Center of Biarjomand Municipality

[2] Department of Computer Engineering, Gorgan Branch, Islamic Azad University, Gorgan, Iran

M.eftekharian@gorganiau.ac.ir

ali.nodehi@gorganiau.ac.ir

Abstract

Todays, medical intelligence detection systems thanks to artificial intelligence have been changed and also faced with some challenges. Breast cancer diagnosis and classification is one of medical intelligence system. There are a variety of screening techniques available to detect breast cancer such as mammography, magnetic resonance imaging and ultrasound. This research used MIAS mammography image dataset and try to diagnose and classify benign, malignant masses based on image processing and machine learning techniques. at first, apply pre-processing for noise reduction and image enhancement based on Quantum Inverse MFT, and then image segmentation with Social Spider Algorithm.The type of mass is then diagnosed by the deep neural network(CNN). Obtained results presented that proposed approach have better performance in comparison to others based on some evaluation criteria such as accuracy with 99.57%, sensitivity with 91%, specificity with 86.

Keywords: Breast Cancer, Diagnosis and Classification, Quantum Inverse MFT Algorithm, Social Spider Algorithm (SSA), CNN

Introduction

Breast cancer is a type of cancer that begins in the breast tissue of women with symptoms such as a mass in the breast, breast deformity, skin rash, discharge from the nipple, or partial scaling of the skin. In order to growing cancer, the gene must regulate growth and cell proliferation. These mutations will then become a mass through cell proliferation. Identifying the transporter gene of this cancer can be an important step in predicting breast cancer. High volume of genetic information is the one of the most important problems in representing the large structure and function of biological molecules. Also one of the most important challenges in bioinformatics is the need to design and produce methods, algorithms, and tools to convert this large volume of often heterogeneous (low-level) data to higher-level bio-knowledge [1]. The masses are divided into benign and malignant. Visually, benign masses have very smooth and uniform margins, whereas malignant masses have dark and prominent margins [2].

lecture review

The use of image processing principles and techniques, along with statistical and cognitive pattern recognition in the diagnosis and automated detection of mammograms has reduced human error and increased detection speed. The classification results for detection purposes in this paper are 99.11% for sensitivity, 98.25% for specificity and 98.54 % for accuracy criteria, respectively. Deep learning techniques in [3] used to diagnose and classify breast tumors. Three different deep learning architectures including GoogLeNet, VGGNet and ResNet have been considered and an analysis has been performed between these methods. The results of this method represented that the proposed approach had high accuracy in diagnosis and classification of tumor areas. [8] In this method, The MIAS and INbreast datasets were used, and the noise in the image dataset was reduced by using median and Gaussian filters. In this article, out of 7259 images, 6346 of which were used for training and 913 for testing. They used VGGNet, MobileNet, GoogLeNet, ResNet, DenseNet and proposed a deep ConvNet + SVM hybrid network. The accuracy of their proposed method is 97.8%. In [9], random forest, SVM, decision tree, K-nearest neighbor, and logistic regression methods were investigated with the Wisconsin breast cancer dataset to determine the best machine learning algorithm. From the point of view of the confusion matrix, accuracy and precision, it was found that the support vector machine performs better than other classifiers with an accuracy of 97.2%.. [10], in the pre-processing section, the noise of mammography images is reduced, then the input data is classified using machine learning techniques such as support vector machine, logistic regression, and K-nearest neighbor to data classification. 60% of the data is used for training and 40% for testing, and the accuracy is 97.7. [11] proposed a CAD system for automatic detection of tumor type by machine learning and using different algorithms. After examining different machine learning algorithms and different group models with experiments on two datasets, the results were compared.The results showed that the group method was better than other methods and achieved 98.80% accuracy. In [12] using graph-based semi-supervised learning method a new survival analysis model for analyzing the survival of breast cancer patients. In[13] Z-Curve mapping method has been utilized in order to conversion of DNA alphabetic strings to digital signals. This method has made use of Linear Predictive Coding model to analyze resultant data for feature extraction. In addition, this Method is beneficiary of a certain singular value decomposition computational approaches to select significant features for reduction of dimension.statistical parameters discriminate cancerous samples from non-cancerous ones.

Proposed Method

In the proposed approach, there are three main parts, including pre-processing to improve and reduce noise of mammographic images using a method called Quantum Inverse MFT Filtering.Because These images are noisy due to the way they are created. They have salt and pepper noises that may be confused with microcalcifications found in mammography images. For this reason, they must be Detected and destroyed .Then the image segmentation phase is implemented using the Social Spider Algorithm (SSA). Finally The type of mass is then diagnosed by the CNN[1].

Pre-processing Phase

Each single image in the combination of local threshold and Active Contour is represented by a two-dimensional array of pixels whose values are integers in the range [0,255]. Local thresholding does the initialization of the images in two steps. Initially, the input noise image is considered as the initial image, which will be used to eliminate the image noise. Thus, at the end of the first step, there will be a decomposed image. In the second step, the thresholding is performed on the detail coefficients and one of these decomposed parts is randomly selected and sent to a reconstruction operation by Gauss Fading.

At the first step, we define a threshold value for noise reduction method and then this Quantum Inverse MFT Filtering will be apply by three parts in equation (1) to (3) and in this noise reduction steps which determine some kinds of noises. Active Contour will be apply to determine some variation about these noises to help Quantum Inverse MFT Filtering for more noise reduction. Quantum Inverse MFT Filtering-based local thresholding and Active Contour can produce a much smoother display. A local threshold Function and Active Contour with Quantum Inverse MFT Filtering have two main features, first of which the function is oscillatory or has a wave appearance such as equation (1).

$$\int_{-\infty}^{0} \Psi(t)|^2 dt < \infty \qquad (1)$$

[1] Convolutional Neural Network

Local thresholding are values which have [0 1] or [0 255] colors. In this case most of the energy at $\Psi(t)$ is limited to a finite period of time whose relation is in form equation (2).

$$\int_{-\infty}^{0} \Psi(t)dt = 0 \qquad (2)$$

The proposed method is generally calculated to reduce the noise in equation (3).

$$Method(I) = (\sum_{\Omega} \sqrt{1+\beta^2|\nabla I|^2}) + \frac{\lambda}{2}(I - I_0)^2 \qquad (3)$$

Function (3) is aware of the edges of the image and tries to preserve the important features of the image. $(I - I_0)^2$ Section guarantees a certain degree of validity between the evaluated image and the original image, in which the I evaluated image and the I_0 image are noisy. Parameter ∇I is the period for adjusting the sum of the variations, β and λ are the balancing parameters and Ω is the sum of the points in the image. It should be noted that ∇I as a parameter to adjust the sum of the variations means that one mammography image may be have some noises such as Gaussian, Salt and Pepper or also blur effects. So this variation used to determine the kinds of these noise variation and calculate its sum. By minimizing equation (3), the goal is to reduce the overall image variability by maintaining accuracy[13].

Image Segmentation Phase

Social Spider Algorithm is used for image segmentation. The Social Spider Algorithm formulates the search space of the optimization problem as a very high spider web. Each position on the web presents a practical solution to the optimization problem and all practical answers to the problem that has the answer on the web. The thread is used as the vibration transfer medium produced by the spiders. Each spider on the web demonstrates the ability to find the position and fit the solution based on the objective function as well as the ability to find food sources in the position. Spiders can move freely on the web, but cannot leave the web until they have found a practical answer to the optimization problem. When a spider moves to a new position, it produces a vibration that propagates throughout the web. Each vibration holds the information of one spider, and the other spiders receive information until it receives the vibration.

In fact, the image is referred to as the search space or the web in the Social Spider Algorithm. It is necessary to create an initial population of spiders on the search space or image that is web. The spiders move on the web, aiming to accurately identify boundary areas, which have different brightness and light intensity than other parts, based on color means black and white histograms for gray-scaled images or in RGB modes. In fact, spiders walk on the edges, moving to the right for precise alignment by shaking the lines or to recognize edges and light intensity to segment image. The fit function is the segmentation of the whole image area which is the termination condition.Edges are in mammogram images and in segmentation after defining initial parameters of SSA, spiders should move to find edges due to their brightness value which determined in pre-processing by local thresolding in [0 1] or [0 255] colors. Figure(1) shows segmentation with SSA algorithm.

Fig (1) - segmentation with SSA algorithm.

Classification

Now the output of the noise reduction and mammography imaging phase enters as input to the convolutional neural network. Then, the dot Product between the input and the parameters of each neuron is performed and finally the convolution operation is performed in each layer. After calculating the network output, these results are used to adjust the network training parameters and calculate the error rate. In the next step, based on the calculated error, the back propagation phase begins. Now the gradient of each parameter is calculated according to the Rule chain and all the related parameters of the neural network change according to their effect. After updating the parameters, the feed forward step begins. The trainning ends after a certain number of repetitions. The structure of our proposed convolutional neural network is shown in Figure (2) and table(1). As can be seen, 16 layers are used.

Simulation and Results

We used MATLAB platform to simulate our algorithm. The hardware specifications of our device include a 7-core processor with a 3.4 GHz Intel processor, 6 MB of cache memory, and 6 GB of RAM. In this research, the MIAS dataset is illustrated using statistical properties. In this dataset, there are images with both breast cancer and non-breast cancer features.This dataset can be downloaded at

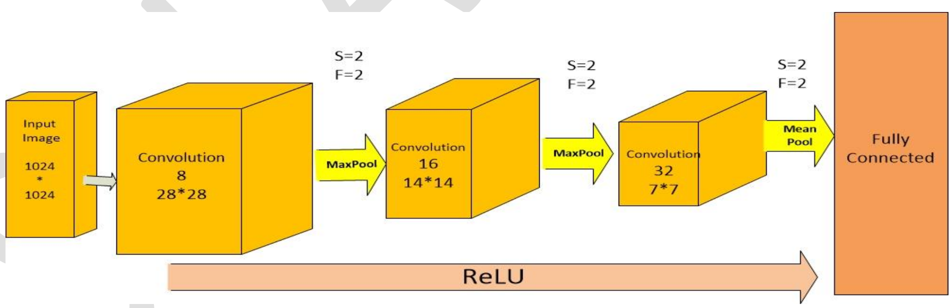

Fig (2) - Proposed CNN deep neural network structure

http://peipa.essex.ac.uk/info/mias.html link.After simulation and implementation of the proposed algorithm, good results were obtained.The mini Mias image dataset used in this article and some

Table(1) - The proposed CNN neural network structure

Layer Number	Layer Name	Description
1	Image Input	1204*1204*3
2	Convolution	8 28*28 Convolutions with Stride [2 2] and Padding 'same'

3	Batch Normalization	Batch normalization
4	ReLU	ReLU
5	Max Pooling	2*2 Max pooling with stride[2 2] and padding [0 0 0 0]
6	Convolution	16 14*14 Convolution with stride [2 2] and padding 'same'
7	Batch Normalization	Batch normalization
8	ReLU	ReLU
9	Max Pooling	2*2 Max pooling with stride[2 2] and padding [0 0 0 0]
10	Convolution	32 7*7 Convolution with stride [1 1] and padding 'same'and padding 'same'
11	Batch Normalization	Batch normalization
12	ReLU	ReLU
13	Max Pooling	2*2 Mean pooling with stride[2 2] and padding [0 0 0 0]
14	Fully Connected	100 Fully Connected layers
15	Soft Max	Soft Max
16	Classification Out Put	Cross Entropyex

evaluation criteria as ROC and AUC, Mean Square Error (MSE), Accuracy, Sensitivity and also Specificity used for classification part. The evaluation criteria is as Table (2).

Table (2) proposed method evaluation criteria

AUC	Sensitivity	Specificity	Accuracy (%)	Mean Square Error (MSE)
0.87	91.00 %	86.00 %	99.57 %	0.01

As it can be seen, our method have good results in classification part. 99.57 % accuracy obtained in whole fusion method. Finally, a comparison should be made in terms of evaluation criteria. This comparison requires the use of a series of prior articles that have used the MIAS dataset that can be compared under the same conditions. The main comparison criterion for the diagnosis and classification of breast cancer is the accuracy in percent. Table (3) show the results of the comparison of the proposed approach with other previous methods.

Table (3) comparison of proposed approach with recent methods in terms of accuracy

References	Accuracy (%)
M.Toğaçar., 2020 [4]	98.58%
Mahmood, Li et al.,2021 [5]	97.8%
AmineNaji, Aarika et al., 2021[6]	97.2%
Roger ,Lincoln S, et al., 2021[7]	97.18%
Mahmood, Pei , et al., 2021[8]	97.8%
AmineNaji M, Aarika K, , et al., 2021[9]	97.2%
Usman N, Junaid R, et al., 2022[10]	97.7%
Sadia s, Rizwan M, et al.,2022[11]	98.8%
Proposed Approach	99.57 %

Conclusion

this approach proposed a three level processing. Quantum Inverse MFT applied due to enhancement and noise reduction in mammography images.The image segmentation operation is then defined by the operators of the Social Spider and CNN for classification. The obtained results represented that proposed method have 99.57 % accuracy which gained better performance in comparison to recent methods.

future work

In this section, suggestions for future work are given, which include:

-Use other standard data sets. For example, DDSM

-Use other deep learning structures.

-In the segmentation part, to increase the accuracy, use the combination of the social spider algorithm with another classifier such as SVM.

References

[1] N.Kavya,N.Sriraam,N.Usha,B. Hiremath,A. Suresh,D. Sharath , et al." Breast Cancer Lesion Detection From Cranial-Caudal View of Mammogram Images Using Statistical and Texture Features Extraction", International Journal of Biomedical and Clinical Engineering , Vol.9,No.1,2020, PP.16-32.

[2] N.Kavya,N.Sriraam,N.Usha,B. Hiremath,A. Suresh,D. Sharath , et al." Breast Cancer Lesion Detection From Cranial-Caudal View of Mammogram Images Using Statistical and Texture Features Extraction", International Journal of Biomedical and Clinical Engineering , Vol.9,No.1,2020, PP.16-32.

[3]M.Kikuchi,T. Hayashida,R. Watanuki,A. Nakashoji,Y. Kawai,A. Nagayama . "Diagnostic system of breast ultrasound images using Convolutional Neural Network", Cancer Research, Vol.80,No.4, February 2020.

[4]M.Toğaçar, B.Ergen, Z.Cömert. " Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders", Medical Hypotheses, Vol.135, February2020.

[5] Mahmood, T.; Li, J.; Pei, Y.; Akhtar, F. An Automated In-Depth Feature Learning Algorithm for Breast Abnormality Prognosis and Robust Characterization from Mammography Images Using Deep Transfer Learning. Biology 2021, 10, 859

[6] AmineNaji M, Aarika K, HabibBenlahmar EL, AitAbdelouhahid R, Debauche O. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, Volume 191, 2021, Pages 487-492

[7] Roger Resmini, Lincoln Silva, Adriel S. Araujo, Petrucio Medeiros , Débora Muchaluat-Saade Aura Conc, Combining Genetic Algorithms and SVM for Breast Cancer Diagnosis Using Infrared Thermography,sensors,2021

[8] Mahmood, T.; Li, J.; Pei, Y.; Akhtar, F. An Automated In-Depth Feature Learning Algorithm for Breast Abnormality Prognosis and Robust Characterization from Mammography Images Using Deep Transfer Learning. Biology 2021, 10, 859

[9] AmineNaji M, Aarika K, HabibBenlahmar EL, AitAbdelouhahid R, Debauche O. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, Volume 191, 2021, Pages 487-492

[10]. Usman N, Junaid R, Liaqat A, Jungeun K, Qazi Emad Ul Haq, An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers, Ieee Access, 12 May 2022, Vol.10, PP: 78242 - 78252

[11]. Sadia s, Rizwan M, Reddy Gadekallu T , Javed A, Rahmani M , Jawad Kh,, Bhatia S, Bio-Imaging-Based Machine Learning Algorithm for Breast Cancer Detection, diagnostics, 3 May 2022

[۱۲] تیموری یانسری رمضان،میرزا رضایی میترا،صادقی مهدی،نجار اعرابی بابک، بهبود کاهش نویز، پیش‌بینی زمان بقا عاری از بیماری تهاجمی در بیماران مبتلا به سرطان پستان با به‌کارگیری روش‌های یادگیری ماشین نیمه نظارتی مبتنی بر گراف،،محاسبات نرم،دوره۱۰،صفحات ۶۹-۴۸، شهریور ۱۴۰۰

[۱۳] خدائی امین، مظفری تازه کند، روشی جدید برای تفکیک و طبقه‌بندی توالی‌های سرطانی و غیرسرطانی DNA با استفاده از الگوریتم‌های مبتنی بر LPC و SVD ،،محاسبات نرم،دوره۳، صفحات ۵۳-۴۲، شهریور ۱۳۹۳

[۱۴] افتخاریان محسن ، نودهی علی، عنایتی فر رسول، بهبود کاهش نویز، تقطیع و طبقه‌بندی توده‌های سرطان توسط فیلتر تطبیقی معکوس کوانتومی، عنکبوت اجتماعی و ELM بهبود یافته،محاسبات نرم،دوره۱۰،صفحات ۸۹-۷۰، شهریور ۱۴۰۰