DonnishJournals

Donnish Journal of Mathematics and Computer Science Research
Vol. 4(1) pp. 006-010 November, 2018
http://www.donnishjournals.org/mcsr
ISSN: 2984-8628

Original Research Article

Multiclassification for Breast Cancer Image Using Voting Techniques

Hassan Abdalla Ahmed Ali[1]* and Mohamed Alhag Alobed[2]

[1]Faculty of Computer Science and Information Technology, Shendi University, Sudan.
[2]Shendi University, Sudan.
Accepted, 30th September, 2018.

Breast cancer is the most common malignancy disease that affects female population and the number of affected people is the second most common leading cause of cancer deaths among all cancer types in the developing countries. As mammography is an effective breast cancer detection tool at an early stage which is the most treatable stage, it is the primary imaging modality for diagnosis of breast cancer, the basic idea of this paper is to participate in the efforts of enhancing the accuracy in medical image classification. We presented a classification method based on multi-classifier voting method that can aid the physician in a mammogram image classification. The study emphasizes five phases starting with the collection of images, pre-processing (image cropping of ROI), features extracting, classification and Development of multi-classifier followed by testing and evaluation. The experimental results show that the voting achieves an accuracy of *90.04%*which is a good classification result compared to individual ones.

Keywords: Mammograms, Breast cancer, Multi-classifier voting, Early detection, Image classification.

INTRODUCTION

Breast cancer is the uncontrolled growth of cells in the breast region. Breast cancer is the second leading cause of cancer deaths in women today. Early detection of the cancer can reduce mortality rate. Mammography has reported cancer detection rate of 70-90% which means 10-30% of breast cancers are missed with mammography [1]. Early detection of breast cancer can be achieved using Digital Mammography, typically through detection of Characteristics of breast masses and/or microcalcifications.

A mammogram is an x-ray of the breast tissue which is designed to identify abnormalities. Studies have shown that radiologists can miss the detection of a significant proportion of abnormalities in addition to having high rates of false positives. Therefore it would be valuable to develop a computer-aided method for mass/tumor classification based on extracted features from the Region of Interest (ROI) in mammograms [3.4].

Breast cancer subtypes have some shared and unique causes, and contributing factors influencing prevention approaches. Mammography cannot stop or decrease breast cancer but are supportive only in detecting the breast cancer at early stages to increase the survival rate [5]. Regular screening can be a successful strategy to identify the early symptoms of breast cancer in mammographic images [6]. Medical images classification can play an important role in diagnostic and teaching purposes in medicine. And it is a form of data analysis that extracts models describing important data classes. Numerous methods have been created to classify masses into benign and malignant categories by using the different classification method [7]. In [8] the researcher presented a method for diagnosis using mammograms is aimed at classifying the detected cancerous regions as benign or malignant.

A review of several studies demonstrating how CAD tools help in tumor diagnosis. In [9], the researcher proposed a computer-aided diagnosis to detect cancer automatically in mammograms without any help of radiologist or medical specialist. After that, enhancement has been performed so that cancer can be clearly visible and identifiable. Results show that proposed method has achieved 96.74% accuracy as well as 98.34% sensitivity. In [10], researchers presented a computer-aided mass classification method in digitized mammograms using Artificial Neural Network (ANN) and performing benign-malignant classification on the region of interest (ROI) having mass.

A major mass classification mammographic characteristic is texture. ANN exploits this to classify mass as benign or malignant. Statistical textural features in characterizing masses are mean, standard deviation, entropy, sleekness, kurtosis, and uniformity. This method aims to increase classification process efficiency objectively to reduce many false positive of malignancies. Three layers artificial neural network (ANN) with

**Corresponding Author: Hassan Abdalla Ahmed Ali. Faculty of Computer Science & Information Technology, Shendi University, Sudan*
Email: hsn126@ush.sd

seven features was proposed to classify marked regions into benign or malignant achieving 90.91% sensitivity and 83.87% specificity which is promising compared to a radiologist's 75% sensitivity. Classification methods are becoming vast and constantly increasing [11]. The aim of this study is to evaluate the classification methods of medical images and the development of multiple mammograms based on the method of voting (fusion). Voting is an assembly method used to combine the decisions of multiple works.

In [12], researchers used a voting technique to choose which of the answers based on their functionality equivalent versions produce. More recent research presented in [13], concerned the identification of breast cancer patients for whom chemotherapy could prolong survival time and is treated here as a data mining problem.

In this paper, we use techniques of voting, Voting is an aggregation technique used to combine decisions of multiple classifiers, normal and abnormal (either benign or malignant) mammograms. In its simplest form that based on plurality or majority voting, each individual classifier contributes a single vote. The aggregation prediction is decided by the majority of the votes, i.e. the class with the most votes is finally classified. The remainder of this paper is organized as follows: Section 2 introduces the materials and methods, voting algorithm and technique. The experiment is given in Section 3. Results and discussions are provided in Section 4.Finally, Section 5 concludes the study.

METHODOLOGY

This study emphasizes on five phases starting with images collection, pre-processing, features extracting, individual classification and Development of multi-classifier followed by testing and evaluation. Figure 1 shows the five steps of the research method.

Mammogram images collection

Dataset used in this study is downloaded from the MIAS (Mammographic Image Analysis) database website [14]. This dataset was recently used by many researchers. MIA's dataset is used for experimentation purpose in this study which is a standard and publicly available dataset. The size of each mammogram is 1024 × 1024 pixels and 200 micron resolution. MIAS contains a total of 322 mammograms of both breasts (left and right) of 161 patients.

Pre-processing images

After collecting mammogram images, the next step is to determine the region of interest ROI for mammogram images. ROI extracted by entering coordinates X, Y and radius in pixels, according to data provided by the MIAS database for each abnormal mammogram image. A random 60x60 pixels region was extracted for the normal mammogram images, After that, we applied crop technique to the images; a cropping operation was employed in order to cut the interest parts of the image. Cropping removed the unwanted parts of the image usually peripheral to the regions of interest as shown in Figures 2 and 3.

Feature extraction

The accurate classification and diagnostic rate mainly depend upon robust features, particularly while dealing with mammograms, after cropping the Region of Interest (ROI) from [x] position to [y] position and [radius] depends on the MIAS dataset. This stage applies the six functions (Mean, Standard Deviation, Skewness, Kurtosis, Contrast, and Smoothness) to extract the feature values from each mammogram image. The following paragraphs give more details about the six functions used to extract features values.

Individual Classification

The result of the previous three phases converts the data to numeric values. In this stage, we apply five individual classifiers, namely Decision Tree, K-nearest Neighbors, and Artificial Neural Network. The process of classifying features into their respective classes, such as normal and abnormal or benign and malignant, is known as classification. In this paper, we used the voting method on three classifiers (Decision Tree, ANN, and KNN) to apply on medical image that is extracted from MIA's data set. In the next paragraphs, we review and present a brief overview of the five classifiers that are used in the classification stage of the mammogram images.

Decision tree

Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node [15].

K-nearest Neighbors classifier

Pattern classification the k-Nearest Neighbor (K-NN) is a non-parametric algorithm. The K-nearest-neighbor method was first described in the early 1950s. The method is labor intensive when given large training sets, and did not gain popularity until the 1960s when increased computing power became available. It has since been widely used in the area of pattern recognition, Nearest-neighbor classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it [16].

Artificial neural network classification

Artificial Neural Network (ANN) has emerged as an important tool for classification. Neural networks were introduced by McCollum and Pitts in 1943. The artificial neuron is a computer simulated model stimulated from the natural neurons. The neuron is starting to work and send a signal through the axon once the signal extends to a certain threshold. This signal then transfers through to other neurons and may get to the control unit (the brain) for a proper action [17].

Development of multi-classifier based on voting method

In this phase, we proposed a multi-classifier based on the individual results obtained by each single classifier discussed above. The concept of our proposed approach depends on the voting method. Majority of the voting techniques are used to perform the final output of the given data. The voting technique presented by selecting the majority output from the experimental results of the five algorithms. The included Mammogram Image and transport data classification have five classes of output. The voting technique becomes difficult when the results of the five algorithms output equally during majority vote. Figure 4 describes the voting algorithm.

EXPERIMENT

The study contains two main processes the first one is built for each classifier using the 40,50,60 percentage (120 mammogram 48 images, 60 images, 72 images) to training dataset from the dataset and after building the classifier, the 60,50,40 percentage (72 images , 60% images, 48 images) of data is used in the test stage. The results are presented in the upcoming section. To test the performance of the proposed method, different quantitative measures have been used. Accuracy has been used. These can be calculated by using mathematical equation 1:

$$\mathrm{CR}=\frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (1)$$

Where TP is True positive, FP is false positive FN is false negative and TN is true negative.

RESULTS AND DISCUSSION

In this study, MIAS data set was used for three individual classifiers and applied multi-classifier voting based on continuous data set. The highest precision was given with a good accuracy for 60% of data splitting, which was *90.04%*, while in 50% the accuracy was *87.28 %* and in 40 % the accuracy was *85.66 %*. Generally, the accuracy was increased after applying voting in the three precisions as shown in *Table 1*.

After applying three different sizes of training and testing we calculated the overall accuracy, the final results are shown in Table 1 and Figure 5. As a result, our method, namely multi-classifier, outperformed single classifiers. Even the voting produced higher accuracy than these methods. This result shows the accuracy of our method consisting of some classifiers. We compared three classifiers methods in this experiment: multi- classifiers (Decision Tree, ANN, and KNN) and the proposed method based on voting. Figure 6 shows the experimental results of the multi-classifier and voting method.

The main measurement of comparison is accuracy. In a previous study [18] researchers proposed a method to classify movie documenting to positive or negative opinions, consisted of three classifiers based on Decision Tree, ME and Score calculation. Using two voting method (Naïve and weighted and integration with SVMs, Classification accuracy is achieved by Naïve voting is 85.8%, Weighted voting is 86.4%, SVM is 87.1%. The output results are comparable to the work in the literature which achieves *90.04%* accuracy. Future work can explore optimizing the classifiers for improving the accuracy.

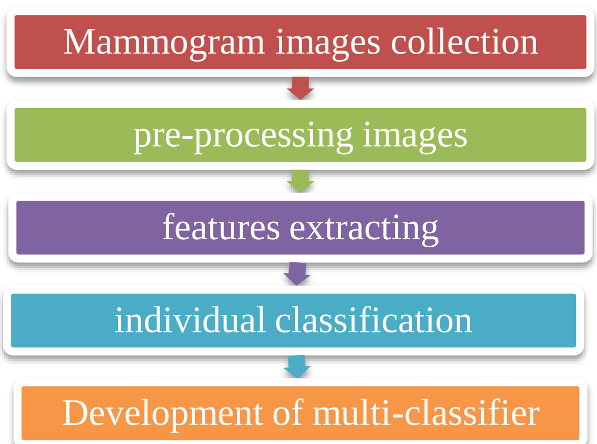

Figure 1: Research phases

Figure 3. Mammogram with Image Cropping

Figure 2. Full Mammogram with detected (ROI) detected

Begin

Read class labels predict by DT classifier
Read class labels predict by ANN classifier
Read class labels predict by KNN classifier
s=size of test instances
For counter=1 to s1
Sum=summation of 3 predicted instances
If sum >=2 Then
Set voting_output =1
Else
Set voting_output =0
End if
End for
End

Figure 4. Voting algorithm

Table 1: Results

	DT	ANN	KNN	voting
40-60	78.34 %	67.75 %	71.05 %	85.66%
50-50	82.20 %	69.50 %	73.09 %	87.28%
60- 40	86.28 %	74.04 %	79.34 %	90.04%

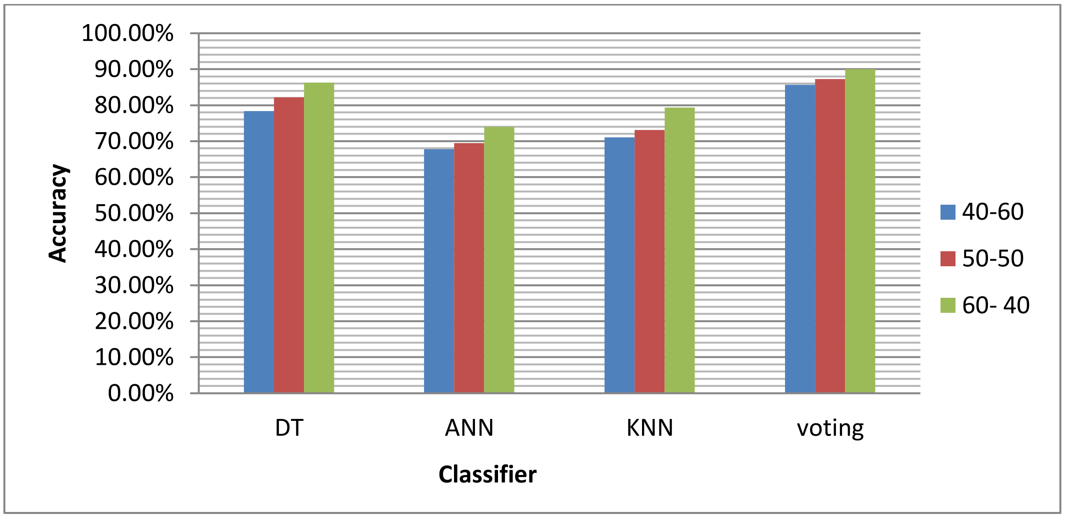

Figure 5: Result of classification and voting accuracy

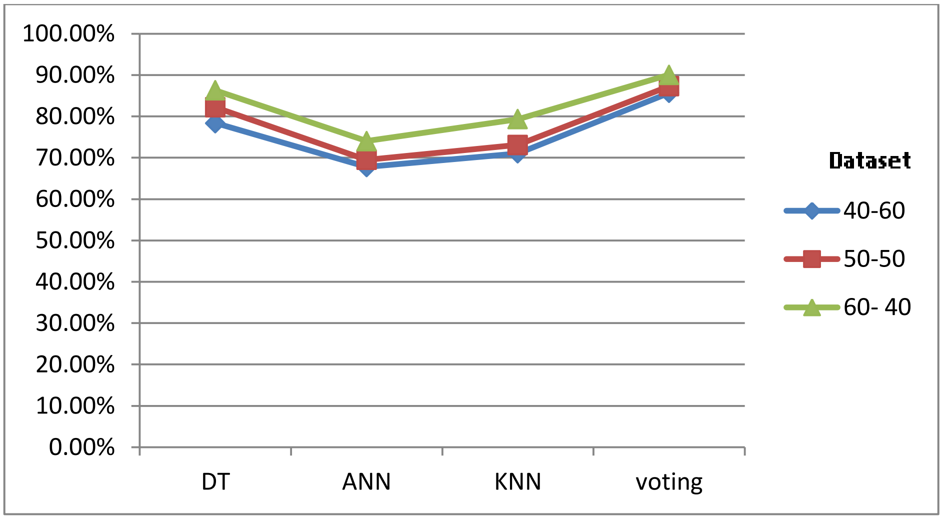

Figure 6. The compared results multi-classifier and voting method

CONCLUSION

This study aimed to build and implement the voting method on three classifiers (Decision Tree, ANN, KNN). The classifiers are applied on the medical image that is extracted from MIAS data set. The study contains two main processes the first one is built for each classifier using the 40,50,60 percentage to training set from the data set and after building the classifier, the 60,50,40 percentage of data is used in the test stage. The accuracy of the voting is 90.04%.

REFERENCES

1. Majid, E. de Paredes, R. Doherty, N. Sharma, and X. Salvador, “Missed breast carcinoma: Pitfalls and pearls1,” Radiographic, vol. 23, no. 4, p.881, 2003.
2. O. Mangasarian, W. Street, and W. Wolberg, “Breast cancer diagnosis and prognosis via linear programming,” Operations Research, vol. 43, no. 4, pp. 570–577, 1995.
3. M. Sampat, M. Markey, A. Bovik et al., “Computer-aided detection and diagnosis in mammography,” Handbook of image and video processing, vol. 10, no. 4, pp. 1195–1217, 2005.
4. Perou, C.M., et al., Molecular portraits of human breast tumours. Nature, 2000. 406(6797): p.747-752.
5. Mencattini, A., et al., Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. IEEE transactions on instrumentation and measurement, 2008. 57(7): p. 1422- 1430.
6. Zhang, G.,et al.Acomputeraided diagnosis system in mammography using artificial neural networks.in BioMedical Engineering and Informatics, 2008.BMEI 2008.International Conference on. 2008: IEEE.
7. Smith, R.A., V. Cokkinides, and H.J. Eyre, American Cancer Society guidelines for the early detection of cancer, 2006. CA: a cancer journal for clinicians, 56(1): p.11-25.
8. Saeed Khodary M Hamouda1,(2017) Enhancement Accuracy of Breast Tumor Diagnosis in Digital Mammograms. Journal of Biomedical Sciences ISSN 2254-609X SN 2254-609Xurna.
9. Jaffar, M.A., Hybrid Texture based Classification of Brea Mammograms using Ad boost Classifier. International Journal of Advanced Computer Science andApplications,2017.8(5).
10. Islam, M. J.,Ahmadi, M., & Sid-Ahmed, M. A. (2010). An efficient automatic mass classification method in digitized mammograms using artificial neural network. arXiv preprint arXiv: 1007.5129.
11. Anunciaçao, O., et al. A Data Mining Approach for the Detection of High Risk Breast Cancer Groups. InIWPACBB. 2010: Springer.
12. Vouk, M.A., et al., An empirical evaluation of consensus voting and consensus recovery block reliability in the presence of failure correlation. Journal of Computer and Software Engineering, 1993. 1(4): p.367-388.
13. Y.J. Lee, O.L.M.W.H.W. Survival -Time Classification of Breast Cancer Patients. 2008 [cited 2017; Available from:http://www.cs.wisc.edu/dmi/annrev/rev0601/uj.ppt.
14. Clark, A.F. The mini-MIAS database of mammograms.2012 [cited 2017; Available from: http://peipa.essex.ac.uk/info/mias.html.
15. Usha, S. and S. Arumugam, Calcification Classification in Mammograms Using Decision Trees. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2016. 9(9): p.2127-2131.
16. Min Dong, Z.W., Chenghui Dong, Xiaomin Mu, Yide Ma, Classification of Region of Interest in MammogramsUsingDualContourletTransformandImprovedKNN. JournalofSensors.2017:p.15.
17. Gershenson, C., Artificial neural networks for beginners. arXivpreprintcs/0308031,2003.
18. Nosayba Mustafa, Multi-classification for Medical Images Using Voting Method Apr 2017.