Cancer Computing : The Early Detection Of Breast Cancer

Student: Chaima DEROUICHE ; Supervisor: Akram BOUKHAMLA

UNIVERSITY KASDI MERBAH OUARGLA: DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

I.INTRODUCTION

Breast cancer is the most common cause for women's death[4] ,for reason of its cause is not fully known yet. As its prevention has proven to be a difficult task; the diagnosis of breast cancer in its early stage of development is recommended indeed.

Mammography is one of early detection tools of breast cancer. It is used to see inside the breast using a low dose of X-ray systems to obtain a medical image; this latter would be used by a radiologist to detect abnormalities inside the breast. However, radiologist can miss the detection of tiny abnormalities and false positive rates. In order to avoid this latter problem; Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADi) techniques are suggested to be as a second reading for mammogram; where CADe technique is used to identify potential abnormalities, while CADi is used to classify the detected abnormal entity (benign or malignant).

Before applying CAD algorithms; preprocessing algorithms are used to improve the quality of mammogram, therefor segmentation techniques are used to remove the pectoral muscle that can bias the result of CAD algorithms.

This paper concerns an approach based on the similarity between intensities to delineate the pectoral muscle boundary using features of measure of semantic similarity between words in Natural Language Process (NLP) and Information Retrieval (IR) fields.

II.METHOD

In Natural Language Process (NLP) and Information Retrieval (IR) fields, the polysemy and synonymy characteristics in words of natural language create a challenge to humans which is the difficulty in choosing the adequate word that refers to the intended meaning .In order to find out relation between words , ontologies are proposed to provide a formal specification of a shared conceptualization between words (i.e.taxonomy).

There are several approach have been proposed for measuring semantic similarity between words and concepts using ontologies such as; based on Information Content(IC) which refers to the probability of the presence of a concept in ontology. The IC of concept can be quantified as follows : $\mathbf{IC(c)=\log^{-1}P(C),}$

where c refers to the concept while P(c) refers to the probability of the presence of a concept in ontology.

Jiang et al.[1] proposed an approach based on Information Content which has been evaluated on Wordnet(an ontology that organizes more than 100,000 general English concepts, which are semantically structured),the similarity between words (c1,c2) is defined as : $\mathbf{Sim(c1;c2) = IC(c1) + IC(c2) - 2IC(LSuper(c1; c2))}$,

where Lsueprs refers to the (lowest super-ordinate) refers to the shortest path from c1 to c2,

Our approach uses the feature of semantic similarity between words to select the pixels that are similar intensity. The similarity between intensities can be quantified as following : $\mathbf{Sim(p1; p2) = IC(P1)+IC(P2)-2IC(LSuper(P1; P2)),}$

where IC(P) refers to the probability of the presence of same intensity pixel in a selected region (i.e slid window) while LSuper refers to the the nearest intensity value to both pixels(P1 and P2) in the selected region.

III.RESULTS

Pectoral muscle removal is an important step before applying the CAD techniques; because this latter will be applied only inside the breast region. Accurate removing of pectoral muscle based on accurate detection of pectoral muscle boundary.

Our method has been evaluated using MATLAB, it consists on two steps :

1- Radiopaque artifacts suppression,

Radiopaque artifacts suppression consist in separating breast profile from background of mammogram. There are several approach have been proposed for this suppression such as:

Nagi et al.[2] proposed an algorithm based on morphological operations to remove radiopaque artifact from mammogram.

First; the image is converted to binary image ,next; this latter is labeled using threshold T=18 , then; the region which has big number of pixels has been extracted, after; morphological operations were applied to remove noise produced due previous

steps, finally; the resulting image is multiplied by original image to obtain a mammogram without radiopaque artifact.

Fig,1: Artifacts Suppression using morphological operation

2-Pectoral muscle boundary detection:

The pectoral muscle boundary is delineated using the proposed approach . the similarity is calculated between selected pixel(Ps) and the rest of pixels(Pi) in image using a window sliding(region) to select the nearest intensity between (Ps) and (Pi), window sliding is reshaped to array which takes ascent order then; the nearest intensity is the previous position of the smallest intensity between (Ps , Pi) in array.

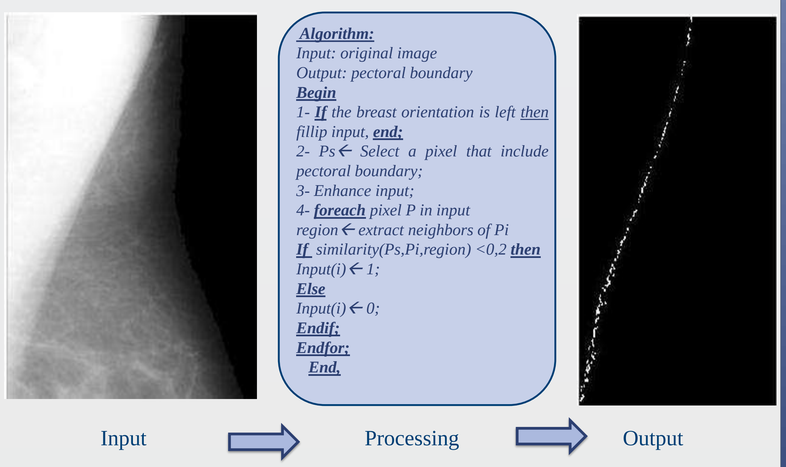


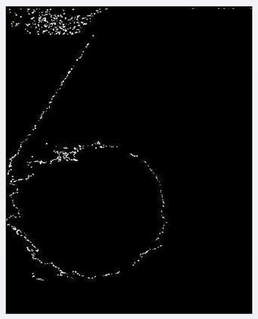
Fig2, Experimental result

IV.DISCUSSION

The proposed method have been applied on Mini-MIAS[3] database(Mammographic Image Analysis Society) and it delineates the accurate pectoral muscle boundary, However ; this approach faces some problems which are:

1-Sometimes, the breast profile boundary is delineate; that because the intensity of both boundaries are similar,

2-The pectoral muscle boundary is not always detected cause of the selected pixel is not include the pectoral muscle boundary,



Fig,3 Breast profile boundary

Fig,4 Invalid pectoral boundary

V.CONCLUSION

Finding an approach capable of delineating the pectoral muscle boundary has proven to be a difficult task; for many reasons such as:

- There are no method that can highlight just the pectoral muscle in mammogram
- Therefore; the presence of some regions in breast profile which have the similar intensity of pectoral muscle boundary .

We have approached the problem from feature of semantic similarity between words which is used to select the pixels that are similar in intensity, as a result the accurate pectoral muscle boundary is delineated.

Our future perspective is to improve the quality of resulting image which can be used in further process in CAD algorithms.

REFERENCES

[1] Jiang, Jay J., and David W. Conrath. "Semantic similarity based on corpus statistics and lexical taxonomy." arXiv preprint cmp-lg/9709008 (1997).

[2] Nagi, Jawad, et al. "Automated breast profile segmentation for ROI detection using digital mammograms." Biomedical Engineering and Sciences (IECBES), 20,

[3] Suckling, John, et al. "The mammographic image analysis society digital mammogram database." Exerpta Medica. International Congress Series. Vol. 1069. 1994.

[4] Kowal, Marek, and Pawe Filipczuk. "Nuclei segmentation for computer-aided diagnosis of breast cancer." International Journal of Applied Mathematics and Computer Science 24.1 (2014): 19-31.

CONTACT

Chaima DEROUICHE email: derchayoume@gmail.com
Akram BOUKHAMLA email: akr_lmd@hotmail.com