A New Approach to the Detection of Lesions in Mammography using Fuzzy Clustering

Y Wang1, H Shi2 and S Ma1

1Department of Computational Mathematics, Jilin University, Changchun, China; 2Department of Radiology, The First Affiliated Hospital of Qiqihaer Medical College, Qiqihaer, China

Breast cancer is a leading cause of female mortality and its early detection is an important means of reducing this. The present study investigated an approach, based on fuzzy clustering, to detect small lesions, such as microcalcifications and other masses, that are hard to recognize in breast cancer screening. A total of 180 mammograms were analysed and classified by radiologists into three groups (n = 60 per group): those with microcalcifications; those with tumours; and those with no lesions. Twenty mammograms were taken as training data sets from each of the groups. The algorithm was then applied to the data not taken for training. Analysis by fuzzy clustering achieved a mean accuracy of 99.7% compared with the radiologists’ findings. It was concluded that the fuzzy clustering algorithm allowed for more efficient and accurate detection of breast lesions and may improve the early detection of breast tumours.

KEY WORDS: BREAST CANCER; MICROCALCIFICATIONS; MAMMOGRAMS; COMPUTER-AIDED DETECTION; FUZZY CLUSTERING

Introduction

Breast cancer is the second leading cause of death for women worldwide and it affects > 8% of women during their lifetime. The World Health Organization’s International Agency for Research on Cancer estimates that more than a million cases of breast cancer occur worldwide annually and more than 400 000 women die each year as a result of it.1

Early detection is the key to improving breast cancer prognosis since the earlier the cancer is detected, the earlier proper treatment can be provided. Mammography is one of the most reliable methods for early detection of breast cancer. For women whose tumours are discovered early by mammograms, the 5-year survival rate is about 82%, as opposed to 60% in those patients where breast cancer is not diagnosed early.2 It is, however, difficult for radiologists to provide both accurate and uniform evaluation for the enormous number of mammograms generated by widespread screening, and reading mammograms requires well-trained and experienced radiologists. Additionally, even well-trained experts may have a high interobserver variation rate.3 With advances in digital image processing and artificial
intelligence, radiologists have an opportunity to improve the diagnosis of breast cancer with the aid of computer systems.4 – 17

Microcalcifications are tiny deposits of calcium in the breast tissue that appear on a mammogram as small clusters of a few pixels. They are the primary signs of breast cancer, where early detection is important to prevent and treat the disease. Detecting all microcalcifications is not an easy task, as there is often poor contrast on mammograms between microcalcifications and the surrounding tissue. In previous studies, several methodologies have been proposed to classify clusters of microcalcifications on digitized mammograms based on artificial neural networks,18 – 22 wavelet analysis,23 – 25 mathematical morphology,26 – 28 Bayesian analysis models,29 and fuzzy logic systems.30 – 32

Fuzzy clustering is a useful tool to deal with uncertainty. In fuzzy clustering, data elements can belong to more than one cluster and associated with each element is a set of membership levels that indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels and then using them to assign data elements to one or more clusters. It is well known that mammograms have some degree of fuzziness, such as indistinct borders, ill-defined shapes and different densities. Due to the nature of mammography and breast structure, fuzzy clustering may be a better choice to handle the fuzziness of mammograms than traditional methods.33 The present study describes the development of an algorithm for mammogram analysis by fuzzy clustering that allows the accurate detection of microcalcifications and breast tumour masses and may aid the early diagnosis of breast cancer.

Patients and methods

STUDY POPULATION

Mammograms (one image per patient) taken from a database of images of consecutive patients screened at the Department of Radiology, The First Affiliated Hospital of Qiqhaer Medical College, Qiqhaer, China between 1 August 2008 and 1 January 2011 were examined. As this was a non-interventional retrospective study of images from a database, patient consent and ethical approval were not required, although patient approval was obtained for the use of the mammograms in medical research.

INTENSITY RANGE IDENTIFICATION

The mammograms were classified into three groups by 11 radiologists: (i) those with microcalcifications; (ii) those with tumours; (iii) and those with no lesions. Each of the 11 radiologists was experienced in mammography and had read a minimum of 4000 mammograms annually for the past 5 years. A simple cyst was considered benign. Lesions highly suspected as malignant, with infiltration into the surrounding fatty tissue or other associated features of malignancy, were classified as malignant. Lesion size was measured (in mm) and the location was recorded. The inter-rater reliabilities of the classification were then evaluated.

A total of 20 mammograms from each of the groups were taken as training sets to determine the optimum range of image intensities needed to identify microcalcifications and tumours in the breast tissue. This procedure was necessary due to the low contrast between breast tissue and bone. To increase the performance of the system, an intensity range of 173 – 227 was identified by the 11 radiologists as effective for the detection of breast tissue lesions (microcalcifications, malignant and benign tumours).
LESION DETECTION USING THE EFCM ALGORITHM

Fuzzy data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. In traditional clustering, data are divided into distinct clusters, where each data element belongs to exactly one cluster. In extended fuzzy C-means (EFCM) clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership functions. The membership function of a fuzzy set is a generalization of the indicator function, which indicates the strength of the association between that data element and a particular cluster. EFCM clustering is a process of assigning these membership functions and then using them to assign data elements to one or more clusters. This algorithm is based on optimization of the cost function and is represented by the following equation:

\[ J(U, V; X) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m \| X_k - V_i \|^2 \]  

(1)

where the membership matrix, \( U = u_{ik} \) is a fuzzy partition, \( X_k \) represents the feature data (intensity value) of the \( k \)th pixel and \( V = v_1, v_2, \ldots, v_c \) is the vector of prototypes of the clusters. The optimal partition \( U^* \) is reached through the couple \( (U^*, V^*) \), which minimizes (locally) the objective function \( J \) according to an alternating optimization iterative process. The EFCM clustering algorithm is composed of the following steps: (i) initialize the membership matrix \( U \) with random values between 0 and 1; (ii) compute the cost function according to equation 1 and stop if it is below a certain tolerance value; and (iii) determine the membership matrix \( U \) according to the result of step ii.

DETERMINATION OF THE MEMBERSHIP FUNCTIONS

Three membership matrixes were determined by the proposed algorithm in order to distinguish problematic areas: (A) the membership function of an area with a lower intensity than that of a lesion; (B) the membership function that accurately reflects an area with intensity similar to that of a lesion; and (C) the membership function of an area with a higher intensity than that of a lesion. These membership functions determined the probability that a pixel with a particular intensity on the mammogram would be linked to one of these functions, i.e. identified the membership levels. The ultimate fuzzy partition membership functions A, B and C are shown in Fig. 1 and demonstrate overlap between them. If the possibility that a region contains lesions is strictly > 50% (i.e. membership function B value > 0.5), it would be decided that the region belonged to the lesion cluster. Thus, according to the decision rule, a particular intensity value that fell within the shaded region of Fig. 1 would be classed as indicative of a lesion cluster. After classification, the results were tested with the data sets that had not been used as training data and compared with the findings of the radiologists.

SUMMARY OF THE EFCM CLUSTERING ALGORITHM

The EFCM clustering algorithm can be summarized as follows: (i) identify the intensity range for different tissues; (ii) use the EFCM algorithm to calculate the membership level of the pixels in the mammograms; and (iii) determine which pixels are associated with which tissue, according to the membership function.

EFCM CLUSTERING ALGORITHM VERSUS EXISTING METHODS

The EFCM clustering algorithm was
Breast cancer detection using fuzzy clustering compared with three other existing methods for the detection of microcalcifications and breast tumours: (i) the support vector machine (SVM) method;\(^{35}\) (ii) the multilayer propagation neural network (MLP) method;\(^{36}\) and (iii) the fuzzy C-means (FCM) method.\(^{37}\)

**Results**

A total of 180 mammograms from patients aged 18 – 64 years were examined and classified into the three groups: (i) those with microcalcifications \((n = 60)\); (ii) those with tumours \((n = 60)\); (iii) and those with no lesions \((n = 60)\). The mean intra-class correlation coefficient for inter-rater reliabilities was 0.962.

Applying the EFCM clustering algorithm to digital mammograms detected microcalcifications and tumour masses of various densities. Comparisons of two sets of results obtained by the EFCM clustering algorithm and the fuzzy C-means clustering methods are shown in Fig. 2. In these two samples, four and three microcalcifications, respectively, were recognized following careful inspection of the mammograms (Figs 2A, 2D).

Application of the EFCM clustering algorithm detected all four microcalcifications (Fig. 2C), whereas the FCM method only detected two of them (Fig. 2B). The FCM method detected many false microcalcifications but did not detect the true microcalcifications (Fig. 2E), while the EFCM clustering algorithm revealed the true microcalcifications by eliminating false microcalcifications and background noise (Fig. 2F).

Original mammograms showing a malignant breast tumour are presented in Figs 3A and 3C, and the results obtained using the EFCM clustering method are shown in Figs 3B and 3D. The identification and classification of a malignant breast tumour from a mammogram by analysis of the grey level threshold, as is usually done by a medical expert, is difficult because fatty tissue and fatty glandular tissue have similar intensities and it is not easy to distinguish these two different regions. The EFCM clustering algorithm provided improved grey level contrast between the tumour and surrounding tissue.

Comparison of the EFCM clustering algorithm with the observations of the 11
Radiologists showed 98.5%, 99.7% and 100% agreement for microcalcifications, tumours and no lesions, respectively. The EFCM clustering method compared with the SVM, \(^{35}\)
MLP\textsuperscript{36} and the traditional FCM methods\textsuperscript{37} showed improved sensitivity, accuracy and specificity of approximately 5\% (Table 1).

**Discussion**

The low resolution of the breast structure on some mammograms means that a diagnosis of cancer can be delayed or undetected. The EFCM clustering algorithm could serve as a useful tool for the early diagnosis of breast cancer and be incorporated into the lesion detection software of computer-aided mammography screening systems. Considering the rate of undiagnosed breast cancer cases worldwide, even a 1\% improvement in early diagnosis as a result of using this algorithm would greatly increase the number of patients who survive.

The advantages of the EFCM clustering algorithm are that lesions can be accurately detected even in mammograms of very dense breasts and irrelevant breast structures can be easily identified and removed from the analysis. The experimental results of the EFCM clustering algorithm compared with the SVM,\textsuperscript{35} MLP\textsuperscript{36} and traditional FCM methods\textsuperscript{37} showed improved sensitivity, accuracy and specificity and, therefore, are encouraging in terms of lesion detection and providing a good platform for further processes such as categorizing lesions as benign or malignant.

In conclusion, the present study demonstrated that the EFCM clustering algorithm can detect lesions more accurately and efficiently than the existing methods and may be useful for the accurate, early diagnosis of breast tumours.

**Conflicts of interest**

The authors had no conflicts of interest to declare in relation to this article.

---

**TABLE 1:**

Comparison of the performance of the extended fuzzy C-means (EFCM) algorithm with that of other diagnostic methods

<table>
<thead>
<tr>
<th>Index</th>
<th>EFCM</th>
<th>SVM</th>
<th>MLP</th>
<th>FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy, %</td>
<td>98.50</td>
<td>92.60</td>
<td>90.80</td>
<td>88.10</td>
</tr>
<tr>
<td>Sensitivity, %</td>
<td>99.72</td>
<td>94.45</td>
<td>95.55</td>
<td>94.72</td>
</tr>
<tr>
<td>Specificity, %</td>
<td>97.42</td>
<td>92.86</td>
<td>90.14</td>
<td>89.83</td>
</tr>
<tr>
<td>Positive predictive value, %</td>
<td>97.21</td>
<td>90.21</td>
<td>91.47</td>
<td>89.39</td>
</tr>
<tr>
<td>Negative predictive value, %</td>
<td>97.68</td>
<td>92.61</td>
<td>93.74</td>
<td>91.00</td>
</tr>
</tbody>
</table>

EFCM, extended fuzzy C-means; SVM, support vector machine; MLP, multilayer propagation neural network; FCM, fuzzy C-means.

---

References

5. Chan HP, Doi K, Vyborny CJ, et al: *Computer-
Breast cancer detection using fuzzy clustering

Y Wang, H Shi, S Ma

32 Bhattacharya M, Das A: Fuzzy logic based segmentation of microcalcification in breast using digital mammograms considering...


Author’s address for correspondence

Dr Yang Wang
Room 631, Mathematics Building, Jilin University, 2699 Qianjin Street, Changchun 130012, China.
E-mail: wyangjlu@gmail.com