K-Means Clustering and Classification of Kinetic Curves on Malignancy in Dynamic Breast MRI

S.H. Lee[1,3], J.H. Kim[1,2,3], K.G. Kim[2,3], S.J. Park[1,3] and W.K. Moon[2,3]

[1] Interdisciplinary Program in Radiation Applied Life Science major,
[2] Department of Radiology,
[3] Institute of Radiation Medicine, Seoul National University College of Medicine, Seoul, Korea

Abstract— **Contrast-Enhanced magnetic resonance of the breast (CE-MRI) is a useful tool for the diagnosis of breast malignancy. Because of its great sensitivity to malignancy, CE-MRI of the breast finds lesions that may not be found on mammography or second-look ultrasound. However, a typical CE-MRI examination produces numerous images which need to be analyzed by ones. The complete analysis of a single dataset requires a long time for the radiologist to diagnose and is an error-prone process due to the errors caused by the fatigue and the habituation of the radiologist. Computer aided diagnosis (CAD) related to CE-MRI is able to help the clinicians in the analysis of big datasets and in the differentiation of kinetic curve patterns. In this study, we present an examination of K-means classifier for the classification of signal intensity curves on malignancy extracted from CE-MRI images. We used 13 examinations with diagnostically confirmed results. From this data, we extracted 1734 malignant patterns and used them to train. The signal patterns were divided into 10 classes through K-means clustering algorithm. The kinetic curves on the tumor region were computationally classified with the centroid signals acquired from the clustering, and the distribution of each class was mapped on the tumor region. The result shows visually the classified distributions of the kinetic curves corresponding to the malignant tumor region. Moreover, various malignant cases could be analyzed by the rank of the kinetic distributions. This technique will help the radiologist particularly in exactly diagnosing a malignant tumor with heterogeneous patterns.**

Keywords— **Dynamic Breast MRI, Malignancy and K-means Clustering**

I. Introduction

Magnetic Resonance Imaging (MRI) in combination with T1 enhancement from gadolinium diethylene triamine pentaacetic acid (Gd-DTPA) [1] is a valuable imaging modality because of its high contrast sensitivity to a large number of characteristics of tissues and body fluids, including T1 relaxation times. Dynamic Contrast Enhanced MRI (DCE-MRI) is necessary to detect breast lesions, and to help differentiate malignant from benign tumors [2, 3]. Malignant breast lesions tend to enhance more rapidly than benign lesions. Observation of contrast enhancement is typically achieved using dynamic imaging techniques whereby the contrast agent is injected during acquisition of a dynamic image set.

The American College of Radiology notes that there are rare situations when traditional imaging modalities are unable to guide patient management. These include situations involving inconclusive or contradictory results on mammography or ultrasound and the search for an unknown primary tumor. Breast MRI may help to solve these problems before definitive treatment.

The basis of DCE-MRI is the fact that tumors need many blood vessels to grow, the concentration of the contrast agent at their location will be higher than in surrounding normal tissues and they will consequently appear as brighter areas in the images. Their dynamic and morphological patterns are related to the diffusion of the contrast agent and to the shape, edges, or the internal pattern of the enhancing region [4]. However, there is a considerable overlap between dynamic and morphological patterns of benign and malignant lesions. In this paper, we will study dynamic features only.

A DCE-MRI requires the acquisition of one series of images before the injection of the contrast agent, called pre-contrast series, and of several series of images, after injection, called post-contrast series. A DCE-MRI examinations produce one hundred images or so, which have to be analyzed by ones.

American college of Radiology (ACR) Breast Imaging and Reporting Data System (BI-RADS) [5] describe the slope from pre-contrast phase to the initial enhancement phase, and the curve change in the delayed phases for kinetic curve assessment. However, BI-RADS criteria do not consider in detail the intermediate enhancement patterns in the delayed phases. Thus we performed a cluster analysis technique as a part of the role of a computer-aided diagnosis (CAD), for providing the radiologist a second opinion to take into account in difficult cases. Our clustering algorithm was a trial to explore features which are not directly human readable, reaching higher level of accuracy.

In this study, we focused on the differentiation of malignant tumor patterns extracted from DCE-MRI examinations and presented the result of *k*-means clustering analysis applied to the problem of classification of dynamic patterns.

II. Materials and methods

A. Materials

Dynamic breast MR images from 13 patients with malignant lesions were acquired through 1.5 T Sonata (Siemens, Erlangen, Germany). First, pre-contrast T1 weighted three dimensional fast low angle shot (3D FLASH) sagittal image was obtained with fat suppression, and four consecutive post-contrast image using the same condition after the injection of 0.1 mmol/kg Gadolinium-DTPA (Magnevist, Schering, Berlin, Germany), respectively. Standard subtraction image, the image resulted after the first post-contrast image subtracts the pre-contrast image, was acquired as post-processing image.

B. Data Acquisition

Signal patterns were collected on a voxel by voxel from the enhanced malignant tumor area for each patient. The total number of the signal patterns was 1734 and these signals were used as the training data for the classification of the malignant signal patterns.

C. K-means Clustering

The *k*-means clustering is a popular method used to divide *n* patterns $\{x_1, \ldots, x_n\}$ in *d* dimensional space into k clusters. The result is a set of k centers, each of which is located at the centroid of the partitioned dataset. This algorithm can be summarized in the following steps:

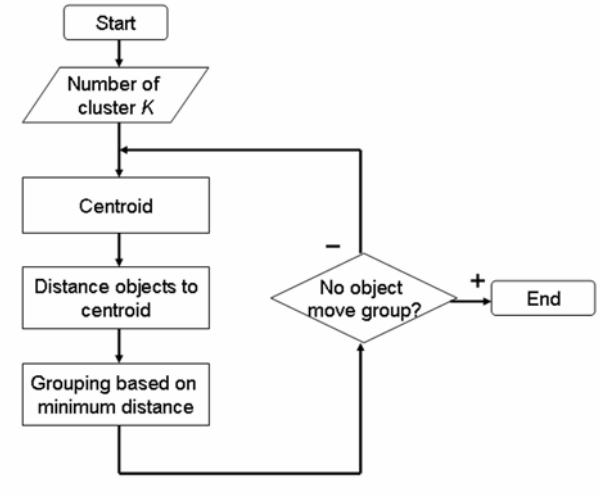


Fig. 1 Steps of *k*-means algorithm

1. Choose the number of clusters *k* and input a dataset of n patterns $X=\{x_1, \ldots, x_n\}$. Randomly select the initial candidates for k cluster centers matrix $V^{(0)}$ from the dataset.
2. Assign each pattern to the nearest cluster using a distance measure. For each pattern x_i, computer its membership function $m(C_j \mid x_i)$ in each cluster C_j. The membership function $m(C_j \mid x_i)$ defines the proportion of pattern x_i that belongs to the *j*th cluster *Cj*. The k-means algorithm uses a hard membership function, that is the membership $m(C_j \mid x_i) \in \{0, 1\}$. If the pattern x_i is closest to cluster C_j (i.e., the distance between x_i and cluster center v_j is minimal), then $m(C_j \mid x_i)=1$; otherwise $m(C_j \mid x_i)=0$.
3. Recompute the centroids (centers) of these k clusters to find new cluster centers v_j, and compute the sum of square error *E*.

$$v_j = \frac{\sum_{i=1}^{n} m(C_j \mid x_i) x_i}{\sum_{i=1}^{n} m(C_j \mid x_i) x_i} \quad \text{for } j = 1, \ldots, k. \tag{1}$$

$$E = \sum_{j=1}^{k} \sum_{x_i \in C_j} \| x_i - v_j \|^2 \quad \text{for } i=1, \ldots, n; j=1, \ldots, k. \tag{2}$$

4. Repeat step 2 and 3 until convergence. Typical convergence criteria are: no more reassignment of patterns to new clusters, the change in error function E fails below a threshold, or a predetermined number of iterations have been reached.

In *k*-means clustering, choosing a good set of initial cluster centers is very important to reduce the distance calculations. However, it is difficult to select a good set of initial clusters randomly. Our method used as initial clusters the centroids which are empirically acquired by training of dataset. Table 1 shows the notation used in describing the algorithm.

Table 1 Notation used in our algorithm

k	The number of clusters
n	The number of patterns in a dataset
x_j	The *i*th data element (pattern)
d	The number of dimensions
v_j	The centroid of the *j*th cluster
C_j	The *j*th cluster

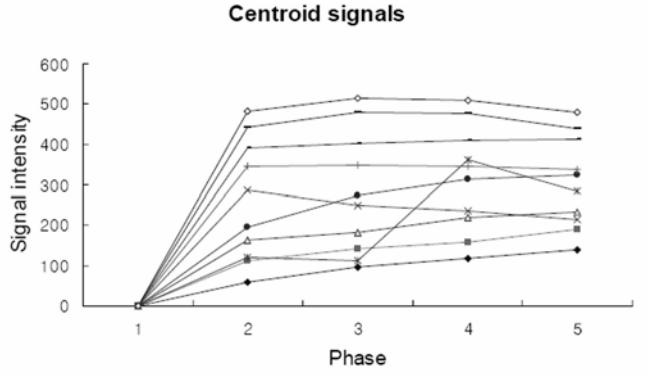


Fig. 2 Ten centroid signals computed from 1734 kinetic curves in malignant tumors

D. Classified Image Mapping

The centroids of the signal patterns computed through *k*-means clustering analysis were used for classifying a random dynamic breast MR image with a malignant lesion. The criterion of classification is to assign each pattern extracted from the MR image to the nearest reference centroid using distance measure. In this study, the number of clusters, *k* was 10. The voxels corresponding to the position of each signal patterns were labeled as its class name and mapped to an image.

E. Removal of False Positives

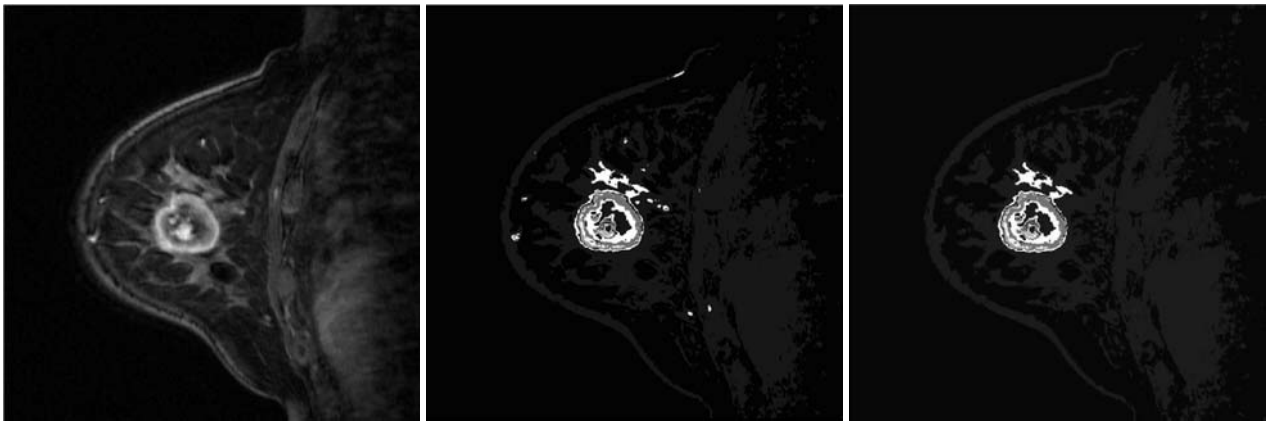
The signals as the cause of the false positives resulted from our clustering algorithm can be the artery signals similar to the kinetic patterns of typical malignant tumors with rapid initial enhancement and washout delayed phases, or the signals of normal tissues resembling the benign-like kinetic patterns, that is, the persisted enhancement patterns of untypical malignant tumors. Thus, we analyzed the size and the circularity of the voxels corresponding to the false positives using ImageJ free software (National Institute of Health, USA) and adequately removed them by thresholding.

F. Development Environments

In this study, the C++ programming language and Insight Segmentation and Registration Toolkit (ITK) open-source software system sponsored by the National Library of Medicine were used for image analysis, and ImageJ software for the removal of false positives.

III. Results

Our classification results were displayed on an image as the texture with different colors for each class of malignant signal patterns. The various classes within a malignant tumor could be visually identified. The use of the more subdivided classification than the existing BI-RADS criteria allowed radiologist the possible to be able to provide the information of more detailed cancer diagnosis.



(a) (b) (c)

Fig. 3 An example of multi-dimensional detection on dynamic breast MR images of a patient with a single breast cancer; (a) the DCE-MR image in the initial phase, (b) the classification results from k-means clustering, (c) the result after removing false positives

IV. Discussion and Conclusion

In this study, we presented the experimentation of *k*-means classifier based on automatic detection of tumor specific patterns, considering the intermediate kinetic patterns in the delayed phases. Such experimentation strategy can discriminate the heterogeneous patterns within malignant tumor and characterize the multi-dimensional patterns in detail, but still fail, when malignant lesions present atypical kinetic patterns similar to the signal patterns of benign tumors, as the case may be. Such observation confirms that, by itself, kinetic patterns are not able to distinguish correctly between benign and malignant lesions, and that morphological analysis of various specific tumors is to be combined.

There are several ways we plan our future work:

1. The short-term goal is to enrich the set of kinetic features used by the classifiers and try to improve the classification of contrast-enhancement patterns. In addition, other advanced classifiers need to be tested.
2. The long-term goal is to introduce morphological analysis. The combination with morphological analysis makes possible the comparison with the fully ranked classification through BI-RADS criteria.

References

1. Watson AD, Rocklage SM, Carvlin MJ. (1992) Magnetic Resonance Imaging. St. Louis, USA.
2. Heywang SH, Hahn D, Schmidt H et al. (1986) MR imaging of the breast using Gadolinium-DTPA. J. Comp. Ass. Tomography, 1986, vol. 10, pp 199-204.
3. Kaiser WA, Zeitler E. (1989) MR Imaging of the breast: fast imaging sequences with and without Gd-DTPA. Radiology, 1989, vol. 170, pp 681-686.
4. Kuhl CK, Schild HH. (2000) Dynamic interpretation of MRI of the breast, J Magn Reson Imaging, 2000, vol. 12, pp 965–974
5. ACR BI-RADS-MRI first edition American College of Radiology, 2003.

Address of the corresponding author:

Author: Jong Hyo Kim
Institute: Interdisciplinary Program in Radiation Applied Life Science major, Department of Radiology and Institute of Radiation Medicine
Street: 28, Yongon-Dong, Chongno-Gu
City: Seoul
Country: Korea
Email: kimjh@radcom.snu.ac.kr