Research Article

Segmentation of Scarred Myocardium in Cardiac Magnetic Resonance Images

Lasya Priya Kotu,[1] **Kjersti Engan,**[1] **Karl Skretting,**[1] **Stein Ørn,**[2]
Leik Woie,[2] **and Trygve Eftestøl**[1]

[1] *Department of Electrical Engineering and Computer Science, University of Stavanger, 4036 Stavanger, Norway*
[2] *Cardiology Department, Stavanger University Hospital, 4011 Stavanger, Norway*

Correspondence should be addressed to Lasya Priya Kotu; lasya.kotu@gmail.com and Kjersti Engan; kjersti.engan@uis.no

The segmentation of scarred and nonscarred myocardium in Cardiac Magnetic Resonance (CMR) is obtained using different features and feature combinations in a Bayes classifier. The used features are found as a local average of intensity values and the underlying texture information in scarred and nonscarred myocardium. The segmentation classifier was trained and tested with different experimental setups and parameter combinations and was cross validated due to limited data. The experimental results show that the intensity variations are indeed an important feature for good segmentation, and the average area under the Receiver Operating Characteristic (ROC) curve, that is, the AUC, is 91.58 ± 3.2%. The segmentation using texture features also gives good segmentation with average AUC values at 85.89 ± 5.8%, that is, lower than the direct current (DC) feature. However, the texture feature gives robust performance compared to a local mean (DC) feature in a test set simulated from the original CMR data. The segmentation of scarred myocardium is comparable to manual segmentation in all the cross validation cases.

1. Introduction

After a myocardial infarction (MI), the myocardium does not function properly due to scarring of the tissue. Late Gadolinium Enhanced Cardiac Magnetic Resonance (LGE-CMR) imaging is used for assessing morphology of myocardium after an MI.

Segmenting the scarred areas from the healthy myocardium is an important prerequisite for various diagnostic analyses. For example, the scar size is largely responsible in left ventricular remodeling [1]. The scar size and texture, as well as the left ventricle ejection fraction, are also important to identify patients with high risk of getting life threatening arrhythmias and decide who will benefit from implantation of implantable cardioverter-defibrillator (ICD) [2]. The scar size is also related to the heart rate of ventricular tachycardia [3].

Earlier work focuses mostly on manual or semiautomatic methods for segmenting the scarred area [2, 4]. At Stavanger University Hospital (SUS), as in many other hospitals today,

the cardiologists will provide their input in a semiautomatic system to segment the heart muscle from the surrounding areas as well as to segment scarred from healthy areas. This is a time consuming job and will be vulnerable to interobserver variability. In recent years, there has been some publications on fully automatic methods for segmenting the myocardial muscle [5, 6] and also methods to segment the scarred areas from the healthy parts of the heart muscle in LGE-CMR images [5, 7–9]. Algohary et al. [10] are proposing an automatic method for segmenting the scar using a combination of two imaging techniques, late enhancement and Strain Encoding (SENC) by using a combination of Otsu's method, morphological operations, and clustering.

Corsi et al. [8] are doing a simple thresholding as they define scarred area as all myocardial tissue with intensity value ≥80% of the maximum intensity value of the myocardium at each slice [8]. This method can only be used on slices where the presence of a scar is known. In a slice (or patient) without scar, this method will induce a scar area, and the paper fails to mention this problem. The method does

not take into account that the degree of damage can vary from patient to patient, and therefore the threshold of 80% is not suitable for all the patients. A simple threshold is almost guaranteed to produce some small false detection areas in some slices.

Dikici et al. [5] use a Support Vector Machine (SVM) classifier to distinguish between the scarred and healthy myocardium tissue. They argue to use a supervised classifier rather than a method based on thresholding because the degree of damage can vary from patient to patient and also because of the partial voluming effects. The three features they use in the classifier are (i) the intensity of a pixel relative to the average myocardial intensity, (ii) the standard deviation of the relative pixel intensities with respect to its next neighbors, and (iii) myocardial contrast defined as the ratio of the mean myocardial intensity over the mean intensity of the entire image. Features (i) and (ii) are calculated for every pixel, whereas feature (iii) is calculated for every slice. All the calculations and scaling are performed on a slice by slice level. In training and testing, they use only three slices for each patient, from the middle of the ventricle. The reported segmentation accuracy on these middle slices is 88.39% with a sensitivity of 81.34% and specificity of 92.28%. Feature (i) seems to be very dependent on the scar size at that particular slice and would possibly benefit from being calculated over the entire 3D volume instead. Since they only test the middle slices they will not experience the slices near the top or bottom of the heart where the scar can cover most of the myocardial muscle. Feature (ii) applies local characteristics and can be considered a textural measure.

The approach by Tao et al. [7] is a three step method: (i) initialization, (ii) false acceptance removal, and (iii) false rejection removal. The initialization is based on Otsu's method for thresholding; that is, the probability density function (PDF) of the intensity levels is estimated, and a threshold is found as a function of the estimated PDF. The technique expects a bimodal PDF for a two class problem and uses an optimization criterion to maximize the ratio of between-class variance and within-class variance. Starting with a segmented heart muscle this does not work very well because the scar is much smaller than the healthy area, and the bimodality of the PDF is not very evident. Tao et al. suggest to improve this by including the blood pool. Thus, they are actually finding a good threshold to separate the blood pool from the heart muscle. The scar will be closer to the blood pool than to the healthy myocardium in terms of intensity levels, and this way they separate the scarred area from the healthy myocardial area. The thresholding is performed jointly on the entire 3D volume (all slices). It is followed by false acceptance removal by connectivity filtering using a two pass algorithm [11]. The false acceptance removal is performed on a slice to slice level and will remove small and thin areas labeled as scar after the thresholding. Region growing is done as a last step. Voxels that are connected to the scarred region and having intensity values larger than $2 \times SD$ of the healthy region are included in the scarred region. The region growing is done on a slice by slice level, so that the SD is calculated for each slice. Tao et al. report experiments of 20 postinfarction patients compared to manually traced scars

from two individual observers using the Dice index metrics. The method is dependent on the blood pool intensity level, which might be questionable. Local textures are not taken into account.

Recent studies showed that the scar tissue is heterogeneous in nature and that the mortality of patients with reduced LVEF depends on the heterogeneity of scar tissue [12, 13]. Texture features from gray-level co-occurrence matrices and Local Binary Patterns (LBPs) were used in our group's previous work to classify patients with high and low risk of getting arrhythmia [14, 15]. Other results from our recent work [16] strengthened our claim that there are textural differences between the healthy myocardium and the scar tissue; thus we want to explore this further. We are concerned about the fact that not all slices (or patients) have scar. We also know that the degree of damage will vary from patient to patient and thus the intensity levels of the scar *relative* to the healthy myocardium might differ from patient to patient, making relative features problematic. On the other hand, the intensity values of the healthy myocardium might vary from patient to patient, making absolute features problematic. Thus, we do not want to base the method on thresholding but rather train a classifier with both textural and intensity based features, conceptually related to Dikici's method. However, we believe that we need to consider the whole 3D volume of the myocardium as the basis in all our calculations. This will include slices without scar occasionally and/or slices with scars covering the entire myocardium. To avoid the problem of the varying degree of damage (or even *no* damage) we classify the LGE-CMR images *without* scaling relative to assumed healthy areas, or maximum values (assumed to be scar). The problem of the intensity values varying from patient to patient, also in the healthy areas, is difficult to address since the degree of damage can vary at the same time. Thus, we carefully exclude a couple of outliers from our training set; however they are included in the test set to examine the robustness of the method. The exclusion is done automatically in the training process.

Section 2 discusses material and methods. Section 3 shows the experimental framework. The results are presented and discussed in Section 4. Finally, the paper is concluded and expected future work is suggested.

2. Material and Methods

This section explains the material and methods used in this work. Section 2.1 discusses the CMR material used in the experiments. The proposed segmentation scheme using Bayes classifier and the feature extraction are illustrated in Section 2.2 and Section 2.3, respectively.

2.1. Material. The Department of Cardiology in Stavanger University Hospital provided the LGE-CMR images for our experiments, and the experiments were conducted in MATLAB. The LGE-CMR images is a group of 24 patients, all with high risk of getting arrhythmia, and they were stored according to the Digital imaging and communications in medicine (DICOM) format with 512×512 pixel resolution. The number of image slices with visible scar in each patient varies approximately from 5 to 12 depending on the size

of scar and heart. Only short-axis CMR images were used in our experiments. The LGE-CMR images were acquired from 1.5 Tesla Philips Intera machine. Images were obtained with a pixel size of $0.82 \times 0.82\,\mathrm{mm}^2$, covering the whole ventricle with short-axis slices of 10 m thickness, without interslice gap. As shown in Figure 1 manual segmentation of the myocardium and infarction tissues were used to form the labeled training sets for the both tissues. The CMR images were not preprocessed before experimenting.

2.2. Segmentation Using Bayes Decision Theory.

The segmentation of the scarred and nonscarred myocardium is obtained by using Bayes decision theory. The class specific probability density function (PDF) modeling specific feature vector values discriminating the scar or the healthy myocardial areas is estimated as $p(\mathbf{v} \mid \mathrm{scar})$ and $p(\mathbf{v} \mid \mathrm{myo})$, respectively. In addition, prior probabilities expressing the proportions of the number of observed pixels being either scar or healthy myocardium are estimated as $P(\mathrm{scar})$ and $P(\mathrm{myo})$. According to Bayes rule, the posterior probability of a pixel being scarred and healthy myocardium can be calculated as

$$P(\mathrm{scar} \mid \mathbf{v}) = \frac{(P(\mathrm{scar})\, p(\mathbf{v} \mid \mathrm{scar}))}{p(\mathbf{v})}, \tag{1}$$

$$P(\mathrm{myo} \mid \mathbf{v}) = \frac{(P(\mathrm{myo})\, p(\mathbf{v} \mid \mathrm{myo}))}{p(\mathbf{v})}, \tag{2}$$

where $p(\mathbf{v}) = P(\mathrm{scar})p(\mathbf{v} \mid \mathrm{scar}) + P(\mathrm{myo})p(\mathbf{v} \mid \mathrm{myo})$ and $\mathbf{v}$ is the feature vector. The class specific PDFs can be estimated using parametric or nonparametric methods. In this work, the parameters in the class specific PDFs are calculated using Maximum Likelihood (ML) estimation. The ML [17] technique is a popular method for parametric estimation of an unknown PDF. The ML estimates of the mean $\mathbf{m}_{\mathrm{ML}}$ and the covariance matrix $\mathbf{S}_{\mathrm{ML}}$ of normally distributed data $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t, \ldots, \mathbf{v}_L\}$ are given below:

$$\mathbf{m}_{\mathrm{ML}} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{v}_i,$$

$$\mathbf{S}_{\mathrm{ML}} = \frac{1}{L} \sum_{i=1}^{L} (\mathbf{v}_i - \mathbf{m}_{\mathrm{ML}})(\mathbf{v}_i - \mathbf{m}_{\mathrm{ML}})^T, \tag{3}$$

where L is the number of training feature vectors.

For a specific image, $\mathbf{v}$ can be computed for every pixel in the segmented myocardium and the value of $P(\mathrm{scar} \mid \mathbf{v})$ and $P(\mathrm{myo} \mid \mathbf{v})$ can be computed for each of these pixels according to (1). There could be different ways of finding the features of the myocardium. The feature value $\mathbf{v}$ can be of any dimension, and in this work they are based on mean intensity and texture of each pixel in the myocardium. The extraction of feature vectors for the myocardium is illustrated in the following Section 2.3. Feature vector $\mathbf{v}$ of myocardium in general can be represented as

$$\mathbf{v} = [\mathrm{feature}_1, \mathrm{feature}_2, \ldots, \mathrm{feature}_l], \tag{4}$$

where l is the dimension of feature vector $\mathbf{v}$. Finally, the segmentation of scarred and nonscarred myocardium is obtained by assigning the pixel to the class that has greater posterior probability.
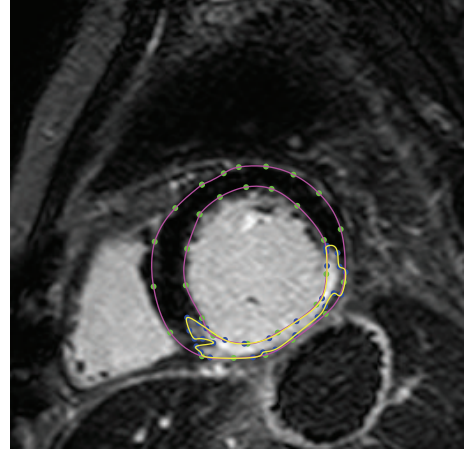

FIGURE 1: Cropped short-axis CMR image showing manual segmentation of myocardium and scar tissues. The green and blue dots in the image are manually marked (by cardiologist) coordinates to segment the myocardium and scar. The magenta and yellow contours generated by cubic spline interpolations of the above coordinates show the myocardium and scar tissues, respectively.

2.3. Extraction of Features.

As discussed in Section 1, features based on intensity and texture information are used for the segmentation of scarred and nonscarred myocardium. The mean intensity value of sliding window image patches is used to obtain the DC feature for each pixel, $\mathrm{dc}(i, j)$. The texture features are calculated using sparse representation by capturing the underlying texture information using dictionary learning techniques. The texture features are named $R_s(i, j)$, $R_m(i, j)$, and $R_p(i, j)$ and will be discussed in detail. $R_m(i, j)$ is correlated with textural features of the healthy myocardial region and $R_s(i, j)$ is correlated with textural features of the scarred region. $R_p(i, j)$ is a scaled value of the above two texture features. The four combination of features used in this work are represented as follows:

$$\mathbf{v} = [\mathrm{dc}(i, j)]; \qquad \mathbf{v} = [\mathrm{dc}(i, j)\, R_s(i, j)\, R_m(i, j)]';$$
$$\mathbf{v} = [\mathrm{dc}(i, j)\, R_p(i, j)]', \qquad \mathbf{v} = [R_p(i, j)]. \tag{5}$$

The detailed process of extracting the DC and texture features is discussed in the following Sections 2.3.1 and 2.3.2.

2.3.1. DC Feature.

Historically, in electronics field, the *mean* is commonly referred to as DC (direct current) value [18]. We define the feature $\mathrm{dc}(i, j) = \mathrm{mean}(I_{\sqrt{N} \times \sqrt{N}}(i, j))$ (where $I_{\sqrt{N} \times \sqrt{N}}$ is the neighborhood around the pixel $I(i, j)$), such that a new image I_{dc}, with $\mathrm{dc}(i, j)$ as values at pixel position (i, j), is made as a sliding window averaging over the myocardium. DC-value $\mathrm{dc}(i, j)$ has been used to segment scarred myocardium because of the intensity differences present in the healthy and scarred myocardium tissues enhanced with the contrast agents. The intensity variations of the scarred and healthy myocardium in LG enhanced CMR image are visible in Figure 1. This is probably the feature that would match the cardiologists segmentation best, since

it reflects the type of information (intensity variations) they use to manually segment the scarred tissue from the healthy myocardium in the CMR images.

2.3.2. Dictionary-Based Textural Features.

Sparse representations and learned dictionaries have been shown to work well for texture classification by Skretting and Husoy in [19] and by Mairal et al. [20]. Sparse representation of learned dictionary atoms reflects the typical texture present in the training set. For each pixel in the myocardium, two textural features are calculated using dictionary learning and sparse representation. One is correlated with textural features of healthy myocardial region $R_m(i, j)$ and the other one is correlated with textural features of scarred region $R_s(i, j)$. In this paper, Recursive Least Squares Dictionary Learning Algorithm (RLS-DLA) presented in [21] is used for dictionary learning and ORMP vector selection algorithm presented in [22] is used for sparse representation.

2.3.3. Recursive Least Squares Dictionary Learning Algorithm.

A dictionary D is an ensemble of finite number of atoms and can be used to represent signals. A linear combination of some of the atoms in the dictionary gives exact or approximate representation of the original signal. Let a column vector x of finite length N represent the original signal. The dictionary atoms are arranged as columns in a $N \times K$ matrix D. The representation, or the approximation of the signal, $\tilde{x}$, and the representation error, r, can be expressed as

$$\tilde{x} = \sum_{k=1}^{K} w(k) d^{(k)} = Dw, \qquad r = x - \tilde{x} = x - Dw, \quad (6)$$

where w is sparse coefficient vector.

Dictionary learning is the task of learning or training a dictionary on a available training set such that it adapts well to represent that specific class of signals. The training vectors and the sparse coefficients are arranged as columns in the matrices X and W, respectively. The objective in dictionary learning is to give a sparse representation of the training set X in order to minimize the sum of the squared error. The cost function is formulated as

$$\underset{D,W}{\operatorname{argmin}} F(D, W) = \underset{D,W}{\operatorname{argmin}} \sum_i \|r_i\|_2^2 \quad \text{s.t } \|w\|_0 \le s, \quad (7)$$

where s is sparsity. The pseudonorm $\| \cdot \|_0$ is the number of nonzero elements. This is a very hard optimization problem. The problem can be solved in two steps algorithm. Step (1) find W using vector selection algorithms, keeping D fixed. Each column in W is given as

$$\hat{w}_i = \underset{w}{\operatorname{argmin}} \|x_i - Dw_i\|_2 \quad \text{s.t } \|w\|_0 \le s. \quad (8)$$

Step (2) keeping W fixed, find D. The dictionary update step depends on the dictionary learning method we choose to use.

The RLS-DLA algorithm presented in [21] is an online dictionary learning algorithm that address the above problems. It updates the dictionary with the arrival of each new training vector. In deriving updating rules for RLS-DLA, we define $X_t = [x_1, x_2, \ldots, x_t]$ of size $N \times t$, $W_t = [w_1, w_2, \ldots, w_t]$ of size $K \times t$ and $C_t = (W_t W_t^T)^{-1}$ for the "time step" t. At each times step the dictionary D_{t-1} and the C matrix are updated so that they obey the least squares solution $D_t = (X_t W_t^T)(W_t W_t^T)^{-1}$. The matrix inversion lemma (Woodbury matrix identity) is applied on C_t to get the following update rules:

$$\begin{aligned} C_t &= C_{t-1} - \alpha u u^T, \\ D_t &= D_{t-1} + \alpha r_t u^T, \end{aligned} \qquad (9)$$

where $u = C_{t-1} w_t$ and $\alpha = 1/(1 + w_t^T u)$ and $r_t = x_t - D_{t-1} w_t$ is the approximation error.

With the inclusion of an adaptive forgetting factor λ_t in RLS-DLA, the update equation (9) is changed to

$$C_t = \left(\lambda_t^{-1} C_{t-1} \right) - \alpha u u^T, \qquad (10)$$

where $u = (\lambda_t^{-1} C_{t-1}) w_t$. The remaining equations remain the same. Due to introduction of the forgetting factor, RLS-DLA has good converging properties and is less dependent on the initial dictionary. RLS-DLA requires less training time due to updating the dictionaries on-line.

2.3.4. Texture Feature Extraction Using Sparse Representation.

In Frame Texture Classification Method (FTCM) presented by Skretting and Husoy [19], texture in a small image patch is modeled as sparse linear combination of dictionary atoms. FTCM is developed by modeling a texture as a tiled floor where all the tiles are identical. The color or the gray level, at a given position in the floor, is given by an underlying continuous periodic two-dimensional function. It is shown based on this model that a vector from spatial neighborhood is indeed a sparse combination of finite dictionary atoms.

Texture feature extraction requires testing and training phases and the whole CMR data set is divided into training and testing images. In training phase, texture feature, R_p, is computed by sparse representation of training images with the help of dictionaries learned on the training images. The algorithm for sparse representation in our work proceeds as follows. Consider the myocardium in a CMR image I that contains two texture classes: healthy and scarred myocardium. The training vector y_l for each pixel in the training image is made from that specific pixel and its neighborhood $\sqrt{N} \times \sqrt{N}$. In the training set, each pixel is categorized into a specified texture class. Then, the dictionaries D_s and D_m are trained for the predefined texture classes (scar and myocardium) using RLS-DLA. Using the two trained dictionaries, each training vector y_l is then represented sparsely using ORMP vector selection algorithm [22]. For training set, the residual images R_s and R_m which are of same size as the original image are calculated for the two texture classes. For each pixel in the myocardium of a training image, the residuals (or representation errors) for the dictionaries D_s and D_m are calculated as

$$R_s(i, j) = \|y_l - D_s w_l^s\|, \qquad R_m(i, j) = \|y_l - D_m w_l^m\|, \quad (11)$$

where w_l^s and w_l^m are sparse coefficient vectors.

A pixel should give less error or residual to the dictionary it belongs. Finally, the residuals R_s and R_m are combined to form the texture feature, R_p. The texture feature is calculated as the ratio of the residual using D_m to the sum of the residuals using D_s and D_m. It is shown as follows:

$$R_p(i, j) = \frac{R_m(i, j)}{(R_s(i, j) + R_m(i, j))}. \qquad (12)$$

Pixel by pixel R_p value can be interpreted as the scaling of residuals R_s and R_m. Smaller values of R_p means that the pixel is not likely to be scar (i.e., healthy myocardium) and larger values means that it is likely to be scar. The texture features calculated from the training set are used to estimate the prior probabilities and PDFs. The test feature vectors are collected from the test image set, in the same way as training feature vectors. The residual images R_s and R_m for test images are calculated for D_s and D_m as training images. Texture is defined as spatial distribution of gray levels. Texture is not pixel-by-pixel local except in the edge between two textures. Therefore, some sort of smoothing is used on the test residual images before the final segmentation of the scarred myocardium [19]. Before calculating R_p, the test residual images R_s and R_m are smoothed using a $a \times a$ pixel separable Gaussian low pass filter with variance σ^2.

3. Experimental Setup

The main aim of our experiments was to compute different features and compare their ability in segmenting the scar from the healthy myocardium using Bayes classifier. The objectives that were explored in our experiments are the discriminative power of the features, the robustness of the features, and examples of the segmented images for illustration. The subsequent subsections discuss about the experimental setup.

3.1. Experimental Details. Our experiments consist of three cases or setups. Each experimental case used different parameters for finding the DC and texture features. Table 1 shows details of the used parameters. The three cases were chosen in order to investigate the following: (1) No. of training patients required for training the classifier, (2) neighborhood size needed to find the features of scarred myocardium, and (3) to find sparsity required to represent the original image patch in scarred myocardium. The three cases were separately cross validated four times to test the robustness of our segmentation technique. The 24 patients were divided into four groups (G1, G2, G3, and G4) each containing 6 patients. In case (i) and case (ii), we used one group for training and the remaining three groups for testing, with fourfold cross validation. In case (iii), two groups were used for training and the remaining two groups were used for testing purposes. While selecting two groups for training, there were six possible ways of combining the four groups of patients. All the six possible combinations were used to train the dictionaries and the classifier in case (iii). In G3, the intensity values of one specific patient are four times higher than the average intensity values of the entire CMR data. It

TABLE 1: The three experimental cases used for segmentation of scarred and nonscarred myocardium.

Experimental parameters	Cases		
	Case (i)	Case (ii)	Case (iii)
Window size	3×3	3×3	5×5
Sparsity	2	4	2
No. of training groups (no. of patients)	3 (18)	3 (18)	2 (12)
No. of testing groups (no. of patients)	1 (6)	1 (6)	2 (12)

was considered as an outlier while training the Bayes classifier and hence the patient was removed while training the Bayes classifier for all feature combinations in three cases. However, the patient was allowed to be used as a test patient.

3.2. Training and Testing Phase. In all CMR Images, we take into account only myocardium segmented by cardiologists. The training phase involves extraction of the DC and texture features in all the three cases. The process was illustrated in Sections 3.2.1 and 3.2.2.

3.2.1. Extraction of DC Feature $dc(i, j)$. Two sets of training vectors were generated from scar and healthy myocardium. The neighborhood size 3×3 and 5×5 were used to form training vectors as explained in Section 2.3.1. The same neighborhood size must be used while training and finding the DC images I_{dc}. The DC images obtained from the training images were used to form the training feature set. The parameters of the class specific PDFs: the mean and the standard deviation were found using ML estimation according to (1) using the training feature set. DC values were scaled to have zero mean and unit variance before finding the ML estimates. The scaling coefficients from the training were stored to scale the test vectors.

3.2.2. Extraction of Texture Features $R_s(i, j)$, $R_m(i, j)$, and $R_p(i, j)$. The texture features were calculated using the same training and test set employed in finding the DC features. Two sets of training vectors were generated from the scarred and nonscarred myocardium segmented by cardiologists. The training vector and test vectors were generated in the same way as in the DC feature experiment using the same neighborhood sizes. Consider the pixels on the border zones, their neighborhood extends into other regions that are not under consideration. If we use training vectors from border regions, then the dictionaries might learn the texture properties of other regions along with the texture properties they were intended to learn. So, the training vectors for the pixels whose neighborhood span other regions were not considered in our experiments. This is depicted in Figure 2. The dictionaries were learned by RLS-DLA as explained in Section 2.3.3 after generating the training vectors from both areas. The dictionary sizes of 9×90 and 25×150 were used in the experiments with window sizes 3×3 and 5×5, respectively. The initial dictionaries were formed by
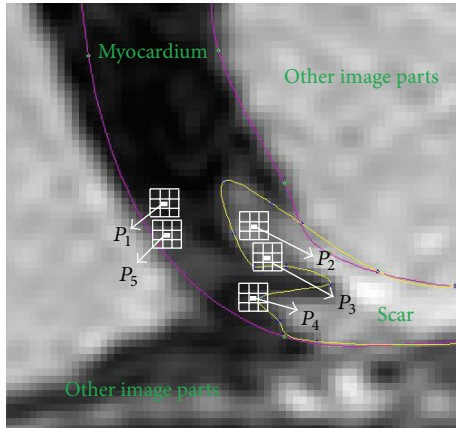
FIGURE 2: The training vector of a pixel is extracted as long as its neighborhood is within one texture area. The neighborhood of pixels P_3, P_4, and P_5 includes more than one texture, and the corresponding feature vectors are excluded from the training set. P_1 and P_2 have the entire neighborhood within one texture region, and hence, the corresponding feature vector is included in that texture's training set.

randomly selecting 90(150) vectors of length 9(25) from the training sets. The forgetting factor was initialized to 0.995 and slowly increased towards 1 according to the exponential method described in [21]. The sparsities s used in our work are 2 and 4. For each pixel in the myocardium of training images, the scaled residual $R_p(i, j)$ was found from residual images as explained in Section 2.3.4. The training feature set was generated from $R_s(i, j)$, $R_m(i, j)$, and $R_p(i, j)$. The training feature set generated from texture features was used to train the Bayes classifier as described in Section 2.2.

In the testing phase, the DC and texture features were generated as in the training phase. As described in Section 2.3.4, the only difference in the testing phase is that the test residual images $R_s(i, j)$, $R_m(i, j)$, and $R_p(i, j)$ were smoothed using low pass Gaussian filter with $\sigma = 5$ and 9×9 window size. Finally, the segmentation results of the scarred myocardium were generated employing Bayes classifier as explained in Section 2.2.

3.3. ROC Analysis. The DC and texture features R_s, R_m, R_p could be combined in different ways. In all the three experimental cases, four combinations of feature vectors were tested. They are (1) dc, (2) dc, R_s, R_m,   (3) dc, R_p, and (4)R_p. The segmentation results obtained in all the three cases with the four combinations of features were compared to the manual segmentation to obtain Receiver Operating Characteristic (ROC) curves [23]. Area under curve (AUC) was used to quantify the performance of the classifier. In all the three cases, each feature combination was subjected to fourfold cross validation. In order to find the discriminative power of the four feature combinations, the true positives and true negatives of all the test patients in the fourfold cross validation were averaged to find the AUC values in the three cases.

The standard deviation of AUC values was computed to examine the robustness of the DC and texture features. The standard deviation of AUC value was computed from the

TABLE 2: Comparison of the average AUC values from the ROC analysis of the four different feature combinations in the three experimental cases.

| | Testing | | |
| | Average of AUC values | | |
Cases (window size, sparsity)	dc	R_s, R_m, dc	R_p, dc	R_p
Case (i) (3×3, 2)	0.92	0.92	0.92	0.85
Case (ii) (3×3, 4)	0.92	0.92	0.92	0.86
Case (iii) (5×5, 2)	0.91	0.90	0.91	0.86

mean true positives and true negative for all the 24 patients. The sensitivity (true positives) and specificity (true negatives) of each patient in all the cross validations of case (i), case (ii), and case (iii) had to be averaged before finding the standard deviation due to the multiple testing of each patient in all the cases. The confidence interval in Figure 4 was also found in the same way. The results are discussed in the following section.

4. Results and Discussion

Referring to the ROC analysis, the discriminative power of the four feature combinations and the robustness of the DC and texture feature are discussed in the subsequent sections.

4.1. Discriminative Power of Features. The average ROC curves of the three cases are shown in Figure 3 and the corresponding AUC values are tabulated in Table 2. The DC feature performed well as seen in Figure 3 and Table 2. From Table 2, it can be found that the feature combinations (1) dc, (2) dc, R_s, R_m, and (3) dc, R_p give almost similar AUC values in all the three cases. The performance of DC feature alone as well as other combinations using the DC feature gives almost same performance with different number of training patients and window size. The texture features did not add any discriminative power to the DC based classifier. Complex classifiers and improved ways of combining DC and texture features might be required to take advantage of the DC and texture feature combinations.

The average AUC values of texture feature, R_p, are less than the average AUC values of the other feature combinations. This could be partly due to the smoothing of residuals images during final segmentation. The smoothing helped in increasing the classification power between the two texture regions, but it might had introduced false negatives at border areas. The sparsity, s, was chosen as two, and four as the discriminative power of residuals of both texture classes will be less when we use more dictionary atoms to represent the test vector. The texture feature, R_p performed better in case (iii), and it shows that the texture feature might need more training patients or a window size of 5×5 for better segmentation. There has been good agreement between the average AUC values of training and test patients in all cases and feature combinations.

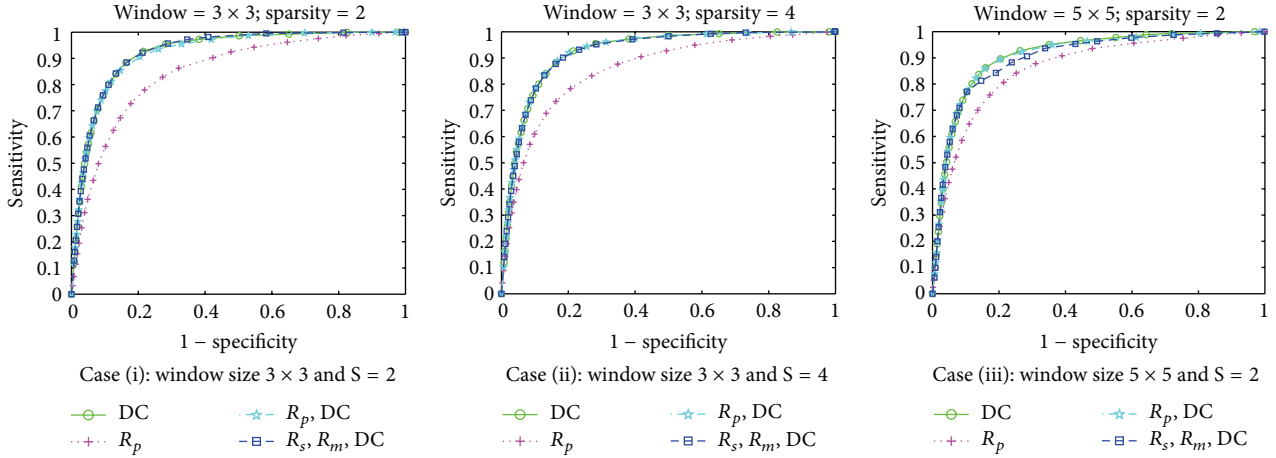Figure 7 shows the segmentation of scarred myocardium using the DC and texture feature, R_p, at 80% sensitivity

FIGURE 3: ROC curves for the three experimental cases and feature combinations: (i) dc, (ii) dc, R_m, R_s, (iii) dc, R_p, and (iv) R_p. In all the three cases, the first three feature combinations have similar performance, and the fourth feature performance is low when compared to other feature combinations.
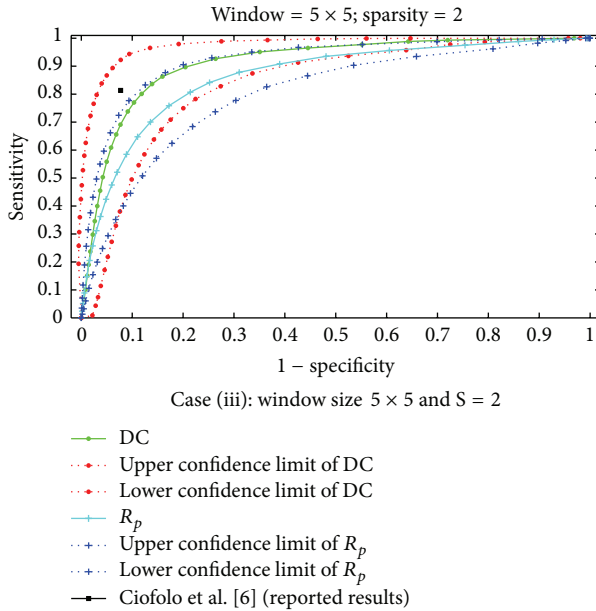

FIGURE 4: ROC curves of the DC, dc, and texture feature, R_p, with confidence interval in case (iii).

TABLE 3: Comparison of the standard deviation AUC values of 24 patients from the fourfold cross validation of four different feature combinations in the three experimental cases.

Standard deviation of AUC values of 24 patients				
Cases (window size, sparsity)	dc	R_s, R_m, dc	R_p, dc	R_p
Case (i) (3×3, 2)	0.032	0.031	0.035	0.072
Case (ii) (3×3, 4)	0.032	0.032	0.034	0.074
Case (iii) (5×5, 2)	0.035	0.039	0.041	0.070

area where the healthy and the scarred myocardium are interwoven together.

4.2. Robustness of Features. Patient specific scaling of intensity values might be a problem as for the robustness of the analyzed features. Hence, we focused to analyze the robustness of DC and texture features individually. Table 3 shows the standard deviation of the AUC values. The standard deviation of AUC values of DC feature and its combinations are less compared to the texture feature, R_p. The standard deviation of AUC values of R_p in case (iii) is slightly less compared to other cases of R_p. Though the number of test patients are different in the cases, the reduction in standard deviation of R_p of case (iii) shows that the texture feature might require more training patients and a window size of 5×5. Figure 4 shows ROC curves of DC and texture feature, R_p, with confidence limits of case (iii). In case (iii), where the texture feature performs well, the upper confidence limit of the texture feature, R_p, almost coincidences with average ROC curve of the DC feature. This indicates that the performance of DC and texture is not significantly different.

In order to show that the texture feature performance is robust on the outliers, we simulated a test set by scaling the original intensities of CMR slices. The CMR images used in our work are from the same MRI device. Even though the MRI machine automatically tries to produce MRI images to be in the same scale of intensities, the CMR images of all the

level in the second trial of case (i) (the color description is explained under the figure). The segmentation results from DC and texture feature in Figure 7 shows that, at the same sensitivity level, the texture feature gives more false positives in the first three CMR slices.

Figure 8 shows the segmentation of scarred myocardium at various sensitivity levels (95%, 90%, 85%, 80%, and 75%) on the ROC curves of DC and texture feature, R_p. The specificity decreases faster for texture feature, R_p, compared to DC feature at the same sensitivity levels. The false positives segmented with the texture feature in Figure 8 at varying sensitivity levels are connected together instead of being random. One can speculate that this can be gray-zone area of the scarred myocardium. Gray-zone area [12, 13] is the
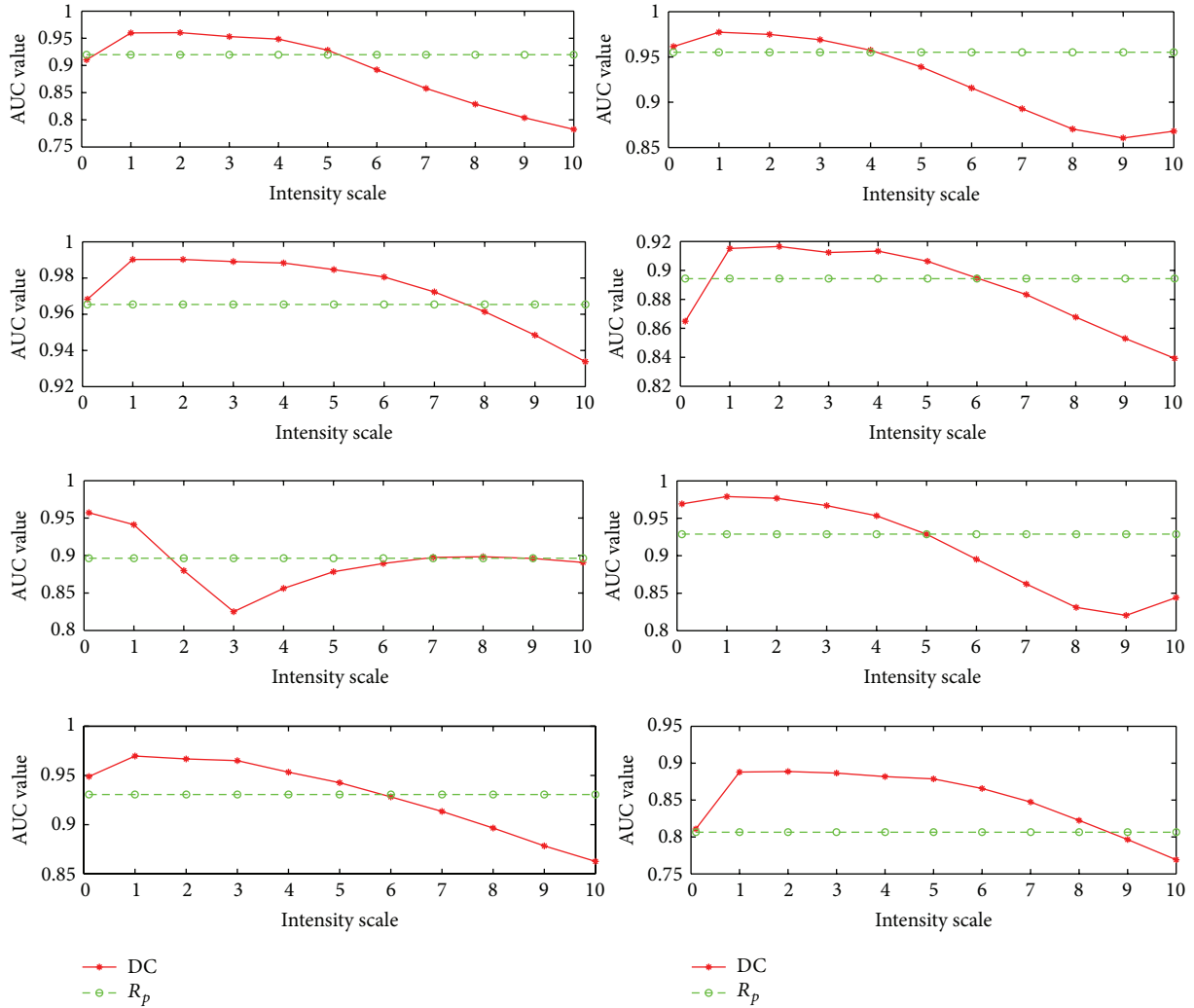
FIGURE 5: AUC values of 8 patients whose CMR slices are scaled at different factors.
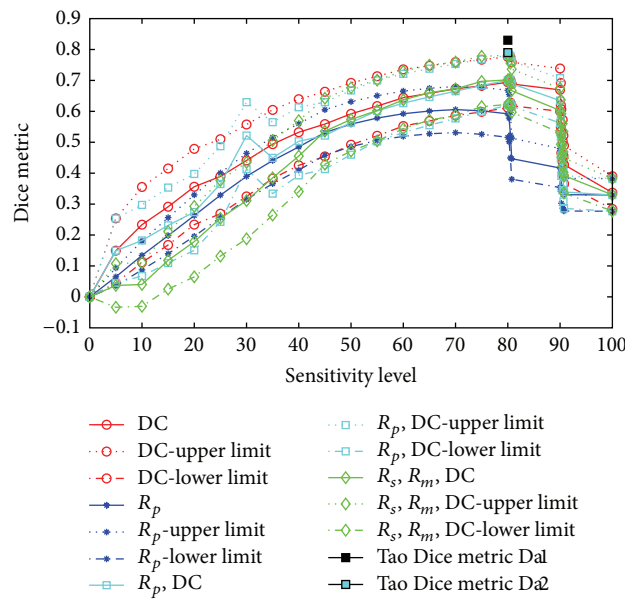

FIGURE 6: Similarity measure—Dice index [7] calculated on our CMR data at different sensitivity levels of the ROC analysis. The Dice index at various sensitivity levels is compared to the reported results by Tao et al. in [7].
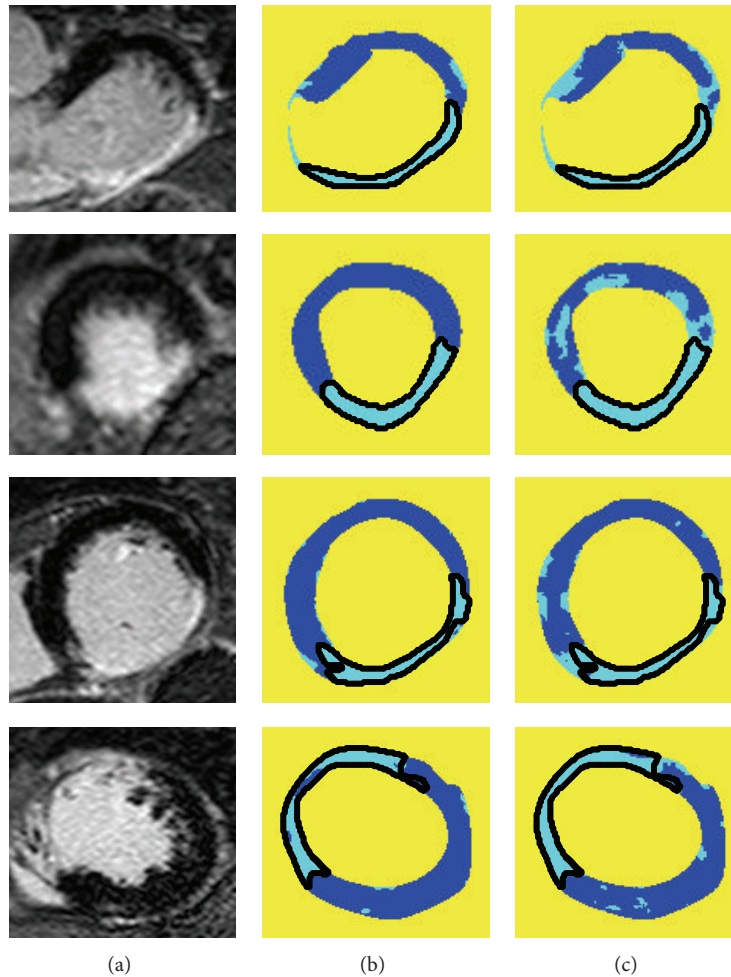
FIGURE 7: The segmentation plots of the scarred myocardium in the CMR images of the patients implanted with ICD using the DC value and texture feature, R_p, in the second trial of case (i) at 80% sensitivity level on the ROC curve. Column (a) the cropped CMR images containing left ventricle. (b) The manually segmented myocardium along with the segmentation results using the DC feature. (c) The segmentation results of the scarred myocardium using the texture feature, R_p. Color Specification: blue: myocardium, cyan: scar, and yellow: other image parts and black contour is the cardiologists segmentation of the scarred myocardium.

patients used in our work vary from one patient to another. With our CMR data, the DC feature performed better than the texture feature. The AUC values for the simulated test were calculated using the already trained classifier in case (iii). The AUC values of the simulated testing set of a patient are plotted against the scaling factor in Figure 5. Figure 5 shows the plots of several AUC values versus intensity scaling factor of eight patients. The initial impression of Figure 5 is that the texture features give the same performance for different scaling factors whereas the DC feature performance varies, either by increasing or decreasing the scaling factor. Figure 5 infers that the DC feature gives good performance on the test patient whose intensities fall in the range of the training patient set whereas texture feature, R_p, performance is not altered with intensity variations. The robust performance of the texture feature on different scaling factors might imply that the dictionaries captured the physiological difference between the scarred and nonscarred myocardium. Therefore, we belief that the textural features might give details that are

not perceived by the cardiologist and not credited in the ROC analysis.

4.3. Comparison to Previous Works.
The segmentation of the scarred and nonscarred myocardium using the DC and texture feature was compared to Dikici et al. [5]. The main results reported by Dikici et al. [5] (sensitivity: 81.34% and specificity: 92.28%) were plotted as one isolated point in Figure 4, compared to the ROC plots of DC and texture features in case (iii). This point lies above both these ROC curves but not above the 95% confidence interval of the DC feature. The result is however not calculated on our study material but is included for comparison. Dikici et al. [5] does not include all the CMRI slices of a patient for training and testing where as in our work we include all the CMRI slices, that is, the scar volume.

Figure 6 shows the similarity measure, Dice index [7] calculated on volume of the scar at different sensitivity levels
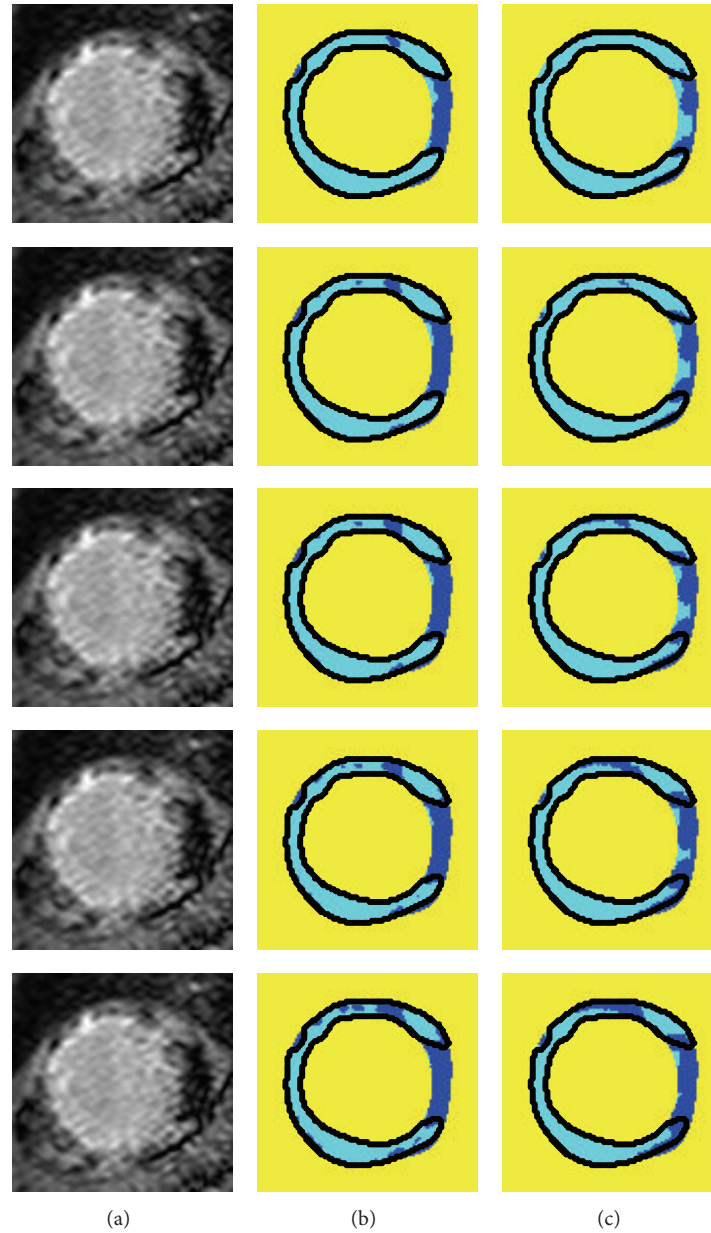
(a)   (b)   (c)

FIGURE 8: The segmentation results of the scarred myocardium in the CMR image of a patient implanted with ICD using the DC and texture feature, R_p, in the second trial of case (i) at 95%, 90%, 85%, 80%, and 75% sensitivity levels on the ROC curve. Column (a) the cropped CMR images containing left ventricle. (b) The manually segmented myocardium along with the segmentation results using the DC feature. (c) The segmentation results of the scarred myocardium using the texture feature, R_p. Color Specification: blue: myocardium, cyan: scar, and yellow: other image parts and black contour is the cardiologists segmentation of the scarred myocardium.

on our CMR data. Dice index gives the percentage of overlap between manual and automatic segmentation [7]. The Dice index increases till 80% sensitivity level and falls later due to the increase in false positives. The Dice index measures reported by Tao et al. in [7] were compared to Dice index measures calculated at 80% sensitivity level and they were plotted as two isolated points (compared to two manual segmentations) in Figure 6. Figure 6 shows that one point is above all the Dice index curves whereas the other point lies on the upper limit of the DC feature and its combinations. The Dice index of DC and other DC combinations performed better than the texture feature. While Tao et al. calculated Dice index at volumetric level in [7], CMR image slices with small scar size were removed totally as the slices with small scar sizes give zero Dice index, and it effects the average Dice index measures. In our Dice index calculations, we considered all image slices regardless of the size of the scar. This could be a reason for the low Dice index measures on our data compared to Tao et al. [7] reported results. Another reason for the low Dice index measures of our method could be due to the false acceptance and false rejection removal stages after thresholding in [7].

5. Conclusion

The DC feature segments the scarred myocardium comparatively better than the texture feature on our CMR data, and the combination of the DC and texture features does not seem to improve the overall performance compared to the DC feature alone. The DC feature also needs fewer training patients than the texture feature method seems to need. However, the texture feature from the learned dictionaries and sparse representation is shown to be much more robust when it comes to scaling variations, that is, variations in the intensity value range. The DC feature seems to reflect the way cardiologist perceive the scarred myocardium in CMR images whereas our results indicate that the texture features can be explored to investigate the heterogeneous nature of the scarred myocardium. Our belief that the texture feature can be used to explore the properties of the scarred myocardium got strengthened with the cardiac segments experiment based on the probability mapping of the scarred myocardium in our recent paper [24]. Our method uses the entire scar volume for training and testing the classifier and works well in comparison to the previous works which did not use the the entire scar volume either during training or testing [5, 7]. Other ways of combining DC and texture features and other texture features will be explored in future work.

Conflict of Interests

The authors declare that they have no conflict of Interests.

References

[1] S. Ørn, C. Manhenke, I. S. Anand et al., "Effect of left ventricular scar size, location, and transmurality on left ventricular remodeling with healed myocardial infarction," *American Journal of Cardiology*, vol. 99, no. 8, pp. 1109–1114, 2007.

[2] A. T. Yan, A. J. Shayne, K. A. Brown et al., "Characterization of the peri-infarct zone by contrast-enhanced cardiac magnetic resonance imaging is a powerful predictor of post-myocardial infarction mortality," *Circulation*, vol. 114, no. 1, pp. 32–39, 2006.

[3] L. Woie, T. Eftestøl, K. Engan, J. T. Kvaløy, D. W. T. Nilsen, and S. Ørn, "The heart rate of ventricular tachycardia following an old myocardial infarction is inversely related to the size of scarring," *Europace*, vol. 13, no. 6, pp. 864–868, 2011.

[4] E. Heiberg, H. Engblom, J. Engvall, E. Hedström, M. Ugander, and H. Arheden, "Semi-automatic quantification of myocardial infarction from delayed contrast enhanced magnetic resonance imaging," *Scandinavian Cardiovascular Journal*, vol. 39, no. 5, pp. 267–275, 2005.

[5] E. Dikici, T. O'Donnell, R. Setser, and R. D. White, "Quantification of delayed enhancement MR images," in *Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '04)*, vol. 3216, pp. 250–257, September 2004.

[6] C. Ciofolo, M. Fradkin, B. Mory, G. Hautvast, and M. Breeuwer, "Automatic myocardium segmentation in late-enhancement MRI," in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: from Nano to Macro (ISBI '08)*, pp. 225–228, May 2008.

[7] Q. Tao, J. Milles, K. Zeppenfeld et al., "Automated segmentation of myocardial scar in late enhancement MRI using combined intensity and spatial information," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 586–594, 2010.

[8] C. Corsi, G. Tarroni, A. Tarroni et al., "Automatic quantification of cardiac scar extent from late gadolinium enhancement mri," in *Proceedings of the Computing in Cardiology*, Hangzhou, China, 2011, http://www.cinc.org/2011/preprints/205.pdf.

[9] A. Kolipaka et al., "Segmentation of non-viable myocardium in delayed enhancement magnetic resonance images," *Scandinavian Cardiovascular Journal*, vol. 39, no. 5, pp. 267–275, 2005.

[10] A. O. Algohary, A. M. El-Bialy et al., "Detection of cardiac infarction in MRI C-SENC images," *Universal Journal of Computer Science Engineering Technology*, pp. 36–40, 2010.

[11] L. Shapiro and G. Stockman, *Computer Vision*, Prentice Hall, Upper Saddle River, NJ, USA, 2001.

[12] S. D. Roes, C. J. W. Borleffs, R. J. Van Der Geest et al., "Infarct tissue heterogeneity assessed with contrast-enhanced mri predicts spontaneous ventricular arrhythmia in patients with ischemic cardiomyopathy and implantable cardioverter-defibrillator," *Circulation*, vol. 2, no. 3, pp. 183–190, 2009.

[13] A. Schmidt, C. F. Azevedo, A. Cheng et al., "Infarct tissue heterogeneity by magnetic resonance imaging identifies enhanced cardiac arrhythmia susceptibility in patients with left ventricular dysfunction," *Circulation*, vol. 115, no. 15, pp. 2006–2014, 2007.

[14] K. Engan, T. Eftestøl, S. Ørn, J. T. Kvaløy, and L. Woie, "Exploratory data analysis of image texture and statistical features on myocardium and infarction areas in cardiac magnetic resonance images," in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '10)*, pp. 5728–5731, September 2010.

[15] L. P. Kotu, K. Engan, T. Eftestøl, S. Ørn, and L. Woie, "Texture classification of scarred and non-scarred myocardium in cardiac MRI using learned dictionaries," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 65–68, September 2011.

[16] L. P. Kotu, K. Engan, T. Eftestøl et al., "Local binary patterns used on cardiac mri to classify high and low risk patient groups," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO '12)*, pp. 2586–2590, 2012.

[17] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 4th edition, 2008.

[18] K. Sayood, *Data Compression*, Academic Press, London, UK, 2nd edition, 2000.

[19] K. Skretting and J. H. Husoy, "Texture classification using sparse frame-based representations," *Eurasip Journal on Applied Signal Processing*, vol. 2006, pp. 1–11, 2006.

[20] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.

[21] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2121–2130, 2010.

[22] M. Gharavi-Alkhansari and T. S. Huang, "Fast orthogonal matching pursuit algorithm," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '98)*, pp. 1389–1392, May 1998.

[23] T. Eftestøl, "Controlling true positive rate in ROC analysis," in *Proceedings of the 36th Annual Conference of Computers in Cardiology, CinC 2009*, pp. 353–356, September 2009.

[24] L. P. Kotu et al., "Probability mapping of scarred myocardium using texture and intensity features in CMR images," *BioMedical Engineering Online*, vol. 12, article 91, 2013.