Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients

Lasya Priya Kotu a,*, Kjersti Engan a, Reza Borhani b, Aggelos K. Katsaggelos b, Stein Ørn c, Leik Woie c, Trygve Eftestøl a

a Department of Electrical Engineering and Computer Science, University of Stavanger, Kjell Arholms Gate 41, Stavanger 4036, Norway
b Department of Electrical Engineering and Computer Science, Northwestern University, 633 Clark St, Evanston, IL 60208, USA
c Department of Cardiology, Stavanger University Hospital, Gerd Ragna Bloch Thorsens Gate 8, Stavanger 4011, Norway

A R T I C L E   I N F O

Article history:
Received 30 September 2013
Received in revised form 8 June 2015
Accepted 25 June 2015

Keywords:
Local binary pattern
Sobel filter
k-Nearest neighbor classifier
Support vector machine classifier
Cardiac magnetic resonance image
High and low arrhythmic risk

A B S T R A C T

Introduction: Patients surviving myocardial infarction (MI) can be divided into high and low arrhythmic risk groups. Distinguishing between these two groups is of crucial importance since the high-risk group has been shown to benefit from implantable cardioverter defibrillator insertion; a costly surgical procedure with potential complications and no proven advantages for the low-risk group. Currently, markers such as left ventricular ejection fraction and myocardial scar size are used to evaluate arrhythmic risk.

Methods: In this paper, we propose quantitative discriminative features extracted from late gadolinium enhanced cardiac magnetic resonance images of post-MI patients, to distinguish between 20 high-risk and 34 low-risk patients. These features include size, location, and textural information concerning the scarred myocardium. To evaluate the discriminative power of the proposed features, we used several built-in classification schemes from matrix laboratory (MATLAB) and Waikato environment for knowledge analysis (WEKA) software, including k-nearest neighbor (k-NN), support vector machine (SVM), decision tree, and random forest.

Results: In Experiment 1, the leave-one-out cross-validation scheme is implemented in MATLAB to classify high- and low-risk groups with a classification accuracy of 94.44%, and an AUC of 0.965 for a feature combination that captures size, location and heterogeneity of the scar. In Experiment 2 with the help of WEKA, nested cross-validation is performed with k-NN, SVM, adjusting decision tree and random forest classifiers to differentiate high-risk and low-risk patients. SVM classifier provided average accuracy of 92.6%, and AUC of 0.921 for a feature combination capturing location and heterogeneity of the scar. Experiment 1 and Experiment 2 show that textural features from the scar are important for classification and that localization features provide an additional benefit.

Conclusion: These promising results suggest that the discriminative features introduced in this paper can be used by medical professionals, or in automatic decision support systems, along with the recognized risk markers, to improve arrhythmic risk stratification in post-MI patients.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Patients who have suffered a myocardial infarction (MI) are at risk of experiencing life-threatening arrhythmias due to impaired functioning of the scarred myocardium. The survival rate of patients susceptible to arrhythmia has been shown to be improved by the use of implantable cardioverter defibrillators (ICDs) [1]. However, patients with a low risk of arrhythmias do not benefit from ICD implantation because of the substantial cost and potential complications of this procedure. Moreover, magnetic resonance imaging is generally contraindicated post-ICD implantation. For these reasons, accurate arrhythmic risk stratification of post-MI patients is of significant importance. Currently, recognized risk markers such as reduced left ventricular ejection fraction (LVEF) and large scar size are used to evaluate patient eligibility for an ICD [1,2]. However, LVEF has its own limitations in assessing myocardial functional state [3]. In a population based study [4], only 30% of patients who suffered sudden cardiac death would have qualified as candidates for a prophylactic ICD based on the current guidelines. On the other hand, many patients with poor LVEF would not draw any advantages from ICD prophylaxis [5]. Therefore, there is a need for an alternative approach to risk assessment that is more efficacious and less expensive.
substantial need for improved strategies to distinguish between the high arrhythmic risk group (HAG) and the low arrhythmic risk group (LAG) of patients.

Several studies suggest that scarred myocardial tissue in cardiac magnetic resonance (CMR) images contains useful information for identifying patients who are susceptible to fatal arrhythmias. Bello et al. [6] showed that the risk of arrhythmia is correlated with the infarct size. The location of the scar in the myocardium is another factor in identifying patients with a high risk of arrhythmia, which needs to be further explored. Recent studies indicate that the heterogeneous texture of scarred tissue is likely to be responsible for irregular heartbeat (arrhythmia) [2,7,8]. Motivated by these findings, several studies have been conducted by our group to examine the role of features extracted from the scarred myocardium in arrhythmic risk stratification. Engan et al. [9] were able to classify patients implanted with an ICD into two groups (high and low risk of fatal arrhythmia) using features that describe size, statistics and texture (Haralick features from co-occurrence matrices) of the scarred myocardium. In [10], the same textural features were used along with LVEF and infarct size to classify HAG and LAG patients. Although these studies were exploratory in the sense that a large number (several hundreds) of features were explored on a small dataset, they substantiated the potential use of textural information in predicting serious ventricular arrhythmias. Local binary patterns (LBPs), introduced by Ojala et al. [11], are powerful texture descriptors which have been used in many biomedical applications [12,13]. In [14], we successfully used LBPs along with a simple non-parametric classifier to distinguish between HAG and LAG patients with a sensitivity of 75.0% and specificity of 83.3%. However, in order to be validated and used as a clinical prognostic tool, both of these discriminative measures need to be improved. In the present work, we aim to do so by introducing new quantitative CMR-based image features and using them in combination with widely used classification algorithms, namely, k-nearest neighbor (k-NN) and support vector machine (SVM), adjusting decision (AD) tree and random forest.

The main aim of the manuscript is to differentiate the HAG and LAG patients using CMR image-based features with the help of image processing and pattern recognition techniques. To our knowledge this study is the first attempt in this regard. In previous works, [1,2], the authors used recognized risk markers, LVEF and scar size computed from the CMR images to differentiate HAG and LAG. Therefore, we used the classification results obtained by LVEF and scar size (computed by the cardiologists) to compare our classification results from different CMR image-based features. We have also compared the classification results in this manuscript to our group’s previous works [10,14].

2. Data acquisition

The Department of Cardiology at Stavanger University Hospital provided the late gadolinium (LG)-enhanced CMR images of 54 post-myocardial infarction patients. The CMR images of the HAG patients were captured between February 2005 and March 2007, and the CMR images of the LAG patients were captured between September 2005 and September 2007. The use of LG enhancement techniques during CMR examinations allows precise depiction of scar characteristics and cardiac morphology. Experienced cardiologists divided the patients into high and low arrhythmic risk groups. Out of the 54 patients, 20 were categorized as high risk (HAG) and the remaining 34 were categorized as low risk (LAG). The decision to categorize a patient as high risk is based on a combination of LVEF (less than 30%), cardiac history, i.e., having demonstrated ventricular tachycardia in the past, and possibly other biomarkers (note that LVEF is an important factor in the decision making process). Each one of the HAG patients was later implanted with an ICD, and during the first year following implantation, the ICDs of all HAG patients recorded incidents of VT. The LAG patients had acute myocardial infarction treated with percutaneous coronary intervention (PCI) one year before the CMR images were taken, and during this year no arrhythmia was detected. All the CMR images were obtained from a 1.5 Tesla MRI machine using the same protocol. Resting left ventricular (LV) function was determined from cine images using a steady-state free precession (SSFP) technique in short axis and long axis views, in the true heart axis. The SSFP technique provides very good contrast between the blood and myocardium in CMR imaging [15]. The short axis view constitutes the slices perpendicular to the line between the center of the mitral valve and the apex (long axis) [16]. In-plane resolution was 1.3mm × 1.3 mm with a slice thickness of 8 mm and an interslice gap of 2 mm. The whole LV was covered with 10–14 contiguous slices. The left ventricular mass, LVEF, and left ventricular volumes were determined using short axis volumetry. After acquiring the functional images, a dosage of 0.25 mmol/kg of gadolinium-based contrast agent was infused via an antecubital vein and delayed hyper-enhancement (DHE) images were acquired 10–15 min after the infusion, using an inversion recovery prepared T1-weighted gradient echo (T1-GRE) sequence, with inversion time (TI) = 200–290 ms, repetition time (TR) = 4.1 ms and echo time (TE) = 1.3 ms. Inversion recovery (IR) pulses are used in delayed enhanced CMR imaging to nullify the signal from the healthy myocardium. IR makes healthy myocardium dark in contrast to the enhanced scar. IR pulses are specified in terms of three time parameters: TI, TR and TE. TI is the time it takes for the healthy myocardium turns dark, and was individually adapted to null the healthy myocardium (typically 200–300 ms). Images were acquired with a pixel size of 0.82 mm × 0.82 mm, covering the whole ventricle with short-axis slices of 10 mm thickness, without interslice gaps. Mean heart rate was 57.5 bpm (range: 52–68 bpm). Myocardial scars are bright voxels on LG-enhanced CMR images. Bright voxels have high signal intensity and are more fibrotic than voxels that are less bright. These CMR images were stored according to the Digital Imaging and Communications in Medicine (DICOM) format with a 512 × 512 pixel resolution. In all of our experiments, the short axis image slices showing visible scars were used. Fig. 1 shows segmentation of the healthy and scarred myocardium, superimposed on the original CMR images of one HAG and one LAG patient. The fact that the difference in the scar texture between the two patients is not visually recognizable necessitates the use of image processing techniques to extract appropriate discriminative features.

3. Feature extraction

As discussed in Section 1, several studies suggest that information concerning the scarred myocardium, including the size, location and textural properties of the scar might be useful in identifying patients who are susceptible to arrhythmia. In this section, we introduce quantitative CMR image features that capture this information and which will later be used to classify HAG and LAG patients.

3.1. Relative scar volume

Scar size is currently used as a marker for prophylactic ICD implantation since HAG patients generally have larger scars than LAG patients. Different methods have been proposed in the literature to quantify the scar size. In this work, we use the segmented CMR slices, as shown in Fig. 1, to calculate the relative scar volume (RSV) as the ratio of the number of pixels present in the scar to the
number of pixels present in the myocardium, across all the CMR slices.

Let $S_t$ and $M_t$ denote the scar region and the myocardium region, respectively, in the $t$th CMR slice. RSV is then defined as:

$$RSV = \frac{\sum_{t=1}^{T} |S_t|}{\sum_{t=1}^{T} |M_t|},\quad (1)$$

where $|S_t|$ and $|M_t|$ represent the number of pixels in the associated region and $T$ is the total number of slices in the CMR image sequence.

### 3.2. Localization features

In order to take into account the morphological information concerning the scarred tissue, we need a reference point on any CMR slice $t$, with respect to which the localization features are calculated. This reference point, denoted by $C_t$, is obtained by taking the average position of the centroids of all pixels in the epicardial, myocardial, and endocardial regions. The angular position, $\phi(p)$, of any pixel $p$ is defined as the angle (ranging from 0 to $2\pi$) between the line connecting that pixel to $C_t$ and the left ventricle (LV) axis manually marked by cardiologists. The LV axis is drawn in reference to the left and right ventricles. In some short axis image slices, the right ventricle is not visible. In these cases, the LV axis is marked in reference to the previous slice in which both ventricles are visible.

Although the mapping $\phi(p)$ captures angular information about the pixel, $p$, we still need another quantity in order to uniquely localize each pixel within the myocardium. The radial position, $\lambda(p)$, of any pixel $p$ within the myocardium is a number between 0 and 1, representing the relative distance from that pixel to the inner wall of the myocardium such that the pixels close to this wall have relatively small values of $\lambda$, whereas those close to the outer wall have larger values. The calculation process for the angular and radial positions of an arbitrary pixel within the myocardium is presented in Fig. 2. More details on the localization features can be found in [17].

For all pixels, $p_x$, within the scar volume, angular position values are calculated and stacked into the vector $\Phi$ whose statistics, i.e., $\mu_\phi$ (mean), $\sigma_\phi$ (standard deviation), and $\eta_\phi$ (median) can be easily calculated. A similar procedure is applied to the radial position of the pixels within the scar volume and the corresponding statistics are calculated for the vector $\Lambda$. Finally, the 6-tuple $(\mu_{\phi}, \sigma_{\phi}, \eta_{\phi}, \mu_{\lambda}, \sigma_{\lambda}, \eta_{\lambda})$ represents the localization features used throughout this paper.

### 3.3. Scar gradient

Infarct tissue heterogeneity has been shown to be a strong predictor for spontaneous arrhythmias [7,8]. Fig. 3 depicts two sample 3 x 3 image patches with different degrees of heterogeneity. Of these two patches, the one on the right is more likely to belong to a high-risk patient due to the drastic changes in pixel intensity. To capture and quantify this type of heterogeneity in a gray-scale CMR image, we make use of the derivative images which measure the rate of change in the pixel intensity in the $x$- and $y$-directions. For any pixel $(i, j)$ within the scar region of the $t$th slice, the magnitude of the gradient $D^{(w)}_t(i, j)$ is calculated as:

$$D^{(w)}_t(i, j) = \sqrt{|G^{(w)}_{t,\chi}(i, j)|^2 + |G^{(w)}_{t,\gamma}(i, j)|^2}, \quad (2)$$

where the directional derivative images $G^{(w)}_{t,\chi}$ and $G^{(w)}_{t,\gamma}$ result from convolving the scar image $S_t$ with the $x$- and $y$-direction Sobel kernels of size $w$, respectively. Different $x$-direction kernels ($w = 3, 5, 7$) used in this paper are shown below.

$$K^{(3)}_x = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -2 & 0 \\ -1 & 0 & 1 \end{bmatrix},$$

$$K^{(5)}_x = \begin{bmatrix} 1 & 0 & -1 & -2 & -3 \\ 2 & 1 & 0 & -1 & -2 \\ 3 & 2 & 0 & -2 & -3 \\ 4 & 3 & 0 & -3 & -4 \\ 5 & 4 & 3 & 0 & -5 \end{bmatrix},$$

$$K^{(7)}_x = \begin{bmatrix} 1 & 0 & -1 & -2 & -3 & -4 & -5 \\ 2 & 1 & 0 & -1 & -2 & -3 & -4 \\ 3 & 2 & 0 & -2 & -3 & -4 & -5 \\ 4 & 3 & 2 & 0 & -3 & -4 & -5 \\ 5 & 4 & 3 & 0 & -4 & -5 & -6 \\ 6 & 5 & 4 & 3 & 0 & -6 & -7 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}. \quad (3)$$

Please cite this article in press as: Kotu LP, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. Artif Intell Med (2015), http://dx.doi.org/10.1016/j.artmed.2015.06.001
The $y$-direction kernels are obtained by simply transposing the $x$-direction kernels, i.e., $K_{y}^{(w)} = (K_{x}^{(w)})^{T}$. Finally, the feature $\text{grad}_{\text{scar}}^{(w)}$ is formed by taking the sum of the gradient magnitude of all pixels within the scar volume:

$$\text{grad}_{\text{scar}}^{(w)} = \sum_{t=1}^{T} \sum_{i,j} \text{grad}_{i,j}^{(w)}$$

(4)

### 3.4. Local binary patterns

Local binary patterns (LBPs) are powerful texture descriptors that are used in a variety of computer vision applications due to their computational simplicity. The LBP image is formed by replacing each pixel in the original gray-scale image with a binary string.

This string is obtained by thresholding a circular neighborhood around each pixel and assigning 1 to the neighbors with larger values than that of the center pixel and 0 to the others. The two parameters that play a role in finding the LBP strings are the radius of the circular neighborhood, $R$, and the number of neighbors, $P$. The decimal representation of the LBP binary string is calculated as:

$$LBP_{P,R} = \sum_{i=0}^{P-1} u(g_{i} - g_{c})2^{i},$$

(5)

where $u(\cdot)$ is the unit step function, $g_{c}$ is the gray-scale value of the center pixel and $g_{i}$'s $(0 \leq i \leq P-1)$ are the gray-scale values of its $P$ neighbors located on a circular neighborhood of radius $R$. Fig. 4 demonstrates the calculation of the LBP string for the center pixel of a sample neighborhood.

The number of bitwise transitions ($0 \rightarrow 1$ or $1 \rightarrow 0$) in a circular LBP string of a pixel provides a measure of heterogeneity for the neighborhood around it. There are $2^{P}$ possible binary strings, out of which many are just the rotated version of another string. The set of all circular binary strings is divided into subsets of rotationally equivalent strings in the sense that any string can be obtained from other strings in the same subset via rotation. All strings in a subset can then be mapped to one member of that subset which is called the representative pattern. There are exactly $P+1$ representative patterns with a maximum transition number of 2, which are referred to as uniform patterns [11]. Patterns with a transition number of more than 2 are called non-uniform patterns and are more likely to be seen in HAG patients. For every pixel in the scar region of a CMR slice, we find its LBP string and label it as either non-uniform or one of the $P+1$ uniform patterns, according to its representative pattern. Having a total of $P+2$ different bins, we build a histogram of LBP strings using all of the pixels within the scar volume. Each
Fig. 4. (a) A symmetric circular neighborhood of radius $R=1$ with $P=8$ samples around the center pixel. The gray level values of the pixels are shown inside the circles. (b) Thresholding with respect to the center pixel. The rightmost (least significant) digit in the binary LBP string is the bit to the right of the center pixel. Other places in the string are filled in turn by moving along the clockwise direction. (c) Equivalent decimal representation can be found as the sum of the values inside the circles ($\text{LBP}_{8,1} = 248$).

Fig. 5. Construction of the LBP histogram ($R=1, P=8$). For all pixels within the scar volume, the binary LBP strings are computed and each pixel is accordingly labeled as one of the $P+1$ uniform patterns or the other non-uniform patterns. The number of pixels in each bin is then divided by the total number of pixels. The $P+1$ uniform bins are shown in cyan. All of the non-uniform patterns are accumulated in the last bin which is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

The histogram is then normalized to have a total area of 1. Fig. 5 summarizes the process of forming LBP histograms from CMR image slices.

3.5. Chi-square features

After creating the LBP histograms for all patients (sample histograms), it is possible to use a non-parametric classification scheme to separate HAG and LAG patients. In order to do so, we need two model histograms (one per each class) and a distance metric measuring the similarity between two given histograms. This way, a test patient will be assigned to the class that has a smaller distance when compared to the model histograms of the HAG and LAG patients. Learning the model histogram for a certain class can simply be done by taking the average of the training histograms in that class. Several distance measures have been proposed in the literature [18], out of which the chi-square distance proves to work better for smaller datasets. The chi-square distance between a sample histogram $F$ and a model histogram $G$ is defined as:

$$
\chi^2(F, G) = \sum_{b=1}^{B} \frac{(F_b - G_b)^2}{F_b + G_b},
$$

where $F_b$ and $G_b$ denote the frequencies of the same bin, $b$, in the sample and model histograms, respectively, and $B$ is the total number of bins. A scalar feature, $\chi_{ip,R}$, that later will be used in k-NN and SVM classifiers, is defined as:

$$
\chi_{ip,R} = \frac{\chi^2(F, G_H)}{\chi^2(F, G_F)},
$$

where $F$ is the patient’s LBP histogram, and $G_H$ and $G_F$ are the model histograms for HAG and LAG patients, respectively. $P$ and $R$ in Eq. (7) refer to the LBP parameters used to construct the histograms. According to the way $\chi_{ip,R}$ is defined, we expect high values for this feature to correspond to the patients at high risk of arrhythmias.

4. Experiments

The CMR image sequences of 20 high-risk and 34 low-risk patients were used to evaluate the discriminative strength of the features proposed in this paper. The assessment of LV volumes on these images was performed on full short axis datasets in a random blinded fashion using ViewForum Software (Philips Medical Systems, Best, Netherlands). Indices for LV volumes were obtained by correcting for body surface area. To be later used for feature extraction, the myocardial tissue was segmented into scarred and healthy regions, with the consensus of two expert cardiologists. 17 potential features were then extracted from the segmented scarred myocardium, as described in Section 3, namely: RSV, six localization features ($\mu_\phi, \sigma_\phi, \eta_\phi, \mu_\lambda, \sigma_\lambda, \eta_\lambda$), three scar gradient features with different kernel sizes ($\text{grad}_{\text{car}}, \text{grad}_{\text{car}}, \text{grad}_{\text{car}}$), and seven chi-square features computed for different values of $(P, R)$ for $P, R = 1, 8, 1, 8, 1, 8, 1, 8, 1, 8, 2, 16, 2, 16, 3, 24, 3, 24, 4$). In Table 1, each of the 17 features are serially numbered from 1 to 17. These feature numbers are used hereafter in the result tables. Experiment 1 and Experiment 2 were conducted using built-in routines in matrix laboratory (MATLAB) and Waikato environment for knowledge analysis (WEKA), respectively, and these experiments are described in Sections 4.1 and 4.2.

Please cite this article in press as: Kotu LP, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. Artif Intell Med (2015), http://dx.doi.org/10.1016/j.artmed.2015.06.001
Firstly, in Experiment 1 the features are investigated one by one in a cross-validated classification setup to evaluate the features discriminative power individually. Thereafter, the best performing features are combined to make feature vectors of different dimensions. These feature vectors are tested with an exhaustive search, and the best results after a leave-one-out cross-validation (CV) are presented. Experiment 1 is done in MATLAB by the authors. To investigate the discriminative power of the features from a different angle, another experimental approach is added, Experiment 2. A wrapper based feature selection method is used in Experiment 2 with the help of in-built functions in WEKA. Our data set is limited, thus we do not have the possibility of doing the feature selection on a training set, and the final performance evaluation on a validation set. Using the same data first for feature selection, and thereafter for CV with the selected features will be optimistically biased. Thus, a nested CV approach is adopted. Some experiments are conducted to address the imbalance in the dataset using synthetic minority over-sampling technique (SMOTE) [19]. More details on the two experimental setups are given in the following sections.

4.1. Experiment 1: systematic testing of individual features and feature combinations

In Experiment 1, we used MATLAB routines developed by Ojala et al. to find the LBP labels of the pixels within the scar volume. In order to find the most discriminating features in the classification of HAG and LAG patients, different combinations of the 17 aforementioned scalar features were used as inputs to the widely used k-NN and SVM classifiers. The leave-one-out CV scheme was used to evaluate and compare the classification performance of different feature combinations and classifiers [20]. The training dataset was normalized across the patients such that each feature had zero mean and unit standard deviation. The overall performance of the classification task does not only depend on the features used, but also on the classifier’s parameters. These parameters include the number of neighbors and the distance measure for k-NN and the kernel function for SVM. Different numbers of neighbors (N = 2, 5, 7, 8, 11), distance metrics (Euclidean, city block, correlation), and search rules (nearest, random, consensus) in the k-NN classifier were explored [20]. The parameters were chosen based on the classification accuracy. In our experiments, city block distance with nearest search rule performed better with medium neighborhoods. The parameters of SVM were tuned using grid search [21] to get the optimal parameters. The C parameter of SVM was varied to find best classification results. The following C values were used (0.005, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6) in the experiments. Different kernel functions (multilayer perception kernel, polynomial, Gaussian or radial basis function) were tried to find the best set of classification results. RBF and polynomial kernel function gave good classification accuracy with the SVM classifier, and the classification performance was tested further for various sigma values and polynomial orders. The results obtained with RBF and polynomial kernels for different sigma values and orders, respectively, were presented in Section 5.1. Available k-NN and SVM classification routines in MATLAB R2014b were used in this work. In this manuscript, the good classification results obtained with k-NN and SVM are tabulated and discussed in Section 5.1.

4.2. Experiment 2: wrapper based feature selection and nested cross validation

Additional classification experiments were conducted using WEKA data mining software developed by Machine Learning Group at the University of Waikato [22]. The CMR image dataset is slightly unbalanced with 20 HAG patients and 34 LAG patients. Some classification schemes are sensitive to unbalanced datasets, like k-NN and tree-based classifiers whereas others are not that sensitive, like SVM. To do a fair comparison of the different classifiers, and a fair presentation of final results, the results of Experiment 2 are reported with and without the use of SMOTE [19], a technique used to ameliorate the imbalance between the groups by generating synthetic data. In all SMOTE experiments, the smallest class is increased to the same size as the largest class, i.e. the HAG group is increased with 70%. The number of nearest neighbors are set to 5 in all SMOTE experiments. In short, the gap is found as the difference between an original feature vector and one, randomly chosen, of the five nearest neighbors to this feature vector. This gap is multiplied by a random number between 0 and 1, and added to the original feature vector, producing a synthetic feature vector.

In Experiment 1, leave-one-out CV was used. In Experiment 2, nested CV was used to assess the classification performance of different classifiers. Nested CV is used for classifier model selection and assessment, and it is capable of finding unbiased estimate of true error [23]. Nested CV consists of two steps: (i) n-external CV and (ii) internal CV. The purpose of the internal CV is to select a feature subset via a wrapper method with forward (best first) selection. In our WEKA experiments, the n-cross external CV was performed using 10 folds in the external loop and the number of folds of the internal CV was varied. In the 10-fold external CV, the dataset was partitioned into 10 equally sized subsets or folds. For each fold of the external CV, the training set (90%) was again divided into non-overlapping sets depending on the number of folds in the internal CV loop. In nested CV, the internal loop is done firstly to select the feature subset of that particular fold of the external loop. Thereafter the external loop trains the classifier using the subset of features found by the internal loop, and tests it on the last 10% of the data. This procedure is repeated for all of the folds of the external loop. Different classifiers were tested, but the same type of classifier was used in the internal and external loop in all the experiments. The different classifiers that were tested in the WEKA experiments were: k-NN, sequential minimal optimization (SMO) SVM [24], alternating decision (AD) tree [25] and random forest [26] classifiers. In Experiment 2, the same classifier parameters were used for k-NN and SVM classifiers as in Experiment 1. For AD tree and random forest classifiers, boosting iterations and number of trees, respectively were varied to find the best average AUC.

Note that the feature set may vary with each external fold of the CV scheme. Thus, using a nested CV gives a measure of how well the method works for that dataset. The method includes the feature subset selector and the choice of classifier. The results of the WEKA experiments are tabulated in Section 5.2.

Table 1

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Feature number</th>
<th>RSV</th>
<th>grad0</th>
<th>grad1</th>
<th>grad2</th>
<th>μ_A</th>
<th>σ_A</th>
<th>η_A</th>
<th>μ_φ</th>
<th>σ_φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature number</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Feature name</td>
<td>η_φ</td>
<td>chi_1.1</td>
<td>chi_1.2</td>
<td>chi_2.1</td>
<td>chi_2.2</td>
<td>chi_16.2</td>
<td>chi_16.3</td>
<td>chi_24.3</td>
<td>chi_24.4</td>
<td></td>
</tr>
</tbody>
</table>
5. Results and discussion

Classification accuracy, sensitivity, specificity, and the area under the curve (AUC) are used in this study to evaluate classification performance. Using leave-one-out CV, the number of true HAGs ($t^H$, high-risk patient classified as HAG), true LAGs ($t^L$, low-risk patient classified as LAG), false HAGs ($f^H$, low-risk patient classified as HAG), and false LAGs ($f^L$, high-risk patients classified as LAG) were calculated after each classification round, and used to find the evaluation metrics defined below.

\[
\text{Classification Accuracy} = \frac{t^H + t^L}{t^H + f^H + t^L + f^L},
\]

\[
\text{Sensitivity} = \frac{t^H}{t^H + f^H}, \quad \text{Specificity} = \frac{t^L}{t^L + f^L}.
\]

AUC value is the area under the curve of the sensitivity and 1-specificity coordinates resulting from a receiver operating characteristics (ROC) analysis [19]. The performance of a classifier is evaluated with help of AUC value which expresses the discriminative ability of a classifier and varies from 0 to 1. The ground truth, as mentioned in Section 2, was provided by medical professionals. The 95% confidence intervals for classification accuracy were also found as suggested in [27].

5.1. Results from Experiment 1

Each one of the 17 features described in Section 4 were separately used in both k-NN and SVM classifiers. These features were then ranked in terms of their AUC values. However, if two feature combinations have the same AUC, the one with higher classification accuracy is selected. The seven best performing ones (RSV, grad1, grad2, \(\mu_L\), \(\sigma_L\), \(\chi_{16,3}\), \(\chi_{24,3}\), \(\chi_{24,4}\)) were saved for further experimentation. These seven scalar features were combined to form higher dimensional features (2- to 7-dimensional feature vectors) and the classification performance metrics were evaluated for every possible feature combination. Figs. 6 and 7 represent the HAG vs. LAG classification accuracy for the seven feature combinations, and the average AUC using k-NN and SVM classifiers, respectively. The results are sorted in ascending order for the seven best single-dimensional features and all of their multi-dimensional combinations.

Fig. 6 shows that neighborhood sizes \(N = 8, 11\) in k-NN outperform smaller ones \(N = 2, 5\). The AUC values for the corresponding neighborhoods are shown in the figure. From the plot, for 3D and 4D, city block distance with 8 neighbors performed better than other neighborhoods, and the feature combinations that gave good classification accuracy and AUC were presented in Table 2. Fig. 7 shows the classification accuracy, and average AUC values using polynomial and RBF kernel functions in the case of the SVM classifier for the best performing \(C\) parameter values. The AUC values for the RBF kernel are high compared to the polynomial kernel function as classification performance with the RBF kernel for different feature dimensions does not vary much. However, the highest classification accuracies were obtained with the polynomial kernel. The average AUC values in Figs. 6 and 7 indicate the superiority of SVM over k-NN. We were interested in the features and the feature sets that gave high classification accuracy for the k-NN and SVM classifiers. Therefore, the classification results of k-NN with city block distance with 8 neighbors and SVM with third order polynomial

Please cite this article in press as: Kotu LP, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. Artif Intell Med (2015), http://dx.doi.org/10.1016/j.artmed.2015.06.001
performing and feature combinations accuracy kernel

Table 2
HAG vs. LAG classification performance in Experiment 1 using different feature combinations in the k-NN classifier (city block distance, 8 neighbors); average accuracy = 74.3%.

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy (%)</th>
<th>Confidence interval (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D features</td>
<td>1</td>
<td>59.2</td>
<td>45.9–71.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>70.4</td>
<td>64.5–85.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>68.5</td>
<td>55.2–79.3</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>66.6</td>
<td>53.3–77.7</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>72.2</td>
<td>59.1–82.3</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>74.1</td>
<td>61.0–83.8</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>79.6</td>
<td>67.1–88.2</td>
</tr>
<tr>
<td>2D features</td>
<td>4, 15</td>
<td>83.3</td>
<td>71.2–90.9</td>
</tr>
<tr>
<td></td>
<td>4, 16</td>
<td>85.1</td>
<td>73.4–92.3</td>
</tr>
<tr>
<td></td>
<td>4, 17</td>
<td>83.3</td>
<td>71.2–90.9</td>
</tr>
<tr>
<td>3D features</td>
<td>1, 4, 17</td>
<td>87.0</td>
<td>75.6–93.6</td>
</tr>
<tr>
<td></td>
<td>4, 16, 17</td>
<td>81.5</td>
<td>69.1–89.6</td>
</tr>
<tr>
<td></td>
<td>4, 5, 17</td>
<td>83.3</td>
<td>71.2–90.9</td>
</tr>
<tr>
<td>4D features</td>
<td>4, 5, 15, 17</td>
<td>83.3</td>
<td>71.2–90.9</td>
</tr>
<tr>
<td>5D features</td>
<td>4, 5, 6, 15, 16</td>
<td>85.1</td>
<td>73.4–92.3</td>
</tr>
<tr>
<td>6D features</td>
<td>4, 5, 6, 15, 16, 17</td>
<td>83.3</td>
<td>71.2–90.9</td>
</tr>
</tbody>
</table>

kernel function are summarized and tabulated in Tables 2 and 3, respectively. For multi-dimensional features, only the combinations that resulted in a classification accuracy of at least 80%, and an AUC above 0.8 are shown.

Fig. 7. Classification accuracy of the seven best discriminating scalar features and their possible combinations using SVM classifier with different kernel functions for best performing C parameter values. The average AUC values are shown for different kernel functions and C values in the figure. The vertical axis represents classification accuracy, and the horizontal axis represents the total number of feature combinations for the seven selected features in each dimension. Results are sorted in ascending order for each feature space dimensionality. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 3
HAG vs. LAG classification performance in Experiment 1 using different feature combinations in the SVM classifier (3rd order polynomial kernel and C = 0.5); average accuracy = 74.4%.

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy (%)</th>
<th>Confidence interval (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D features</td>
<td>1</td>
<td>66.7</td>
<td>53.4–77.8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>70.4</td>
<td>64.5–85.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>55.5</td>
<td>42.4–68.0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>70.4</td>
<td>64.5–85.5</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>74.1</td>
<td>61.0–83.8</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>74.1</td>
<td>61.0–83.8</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>79.6</td>
<td>67.1–88.2</td>
</tr>
<tr>
<td>2D features</td>
<td>4, 15</td>
<td>88.9</td>
<td>77.8–94.8</td>
</tr>
<tr>
<td></td>
<td>4, 16</td>
<td>87.0</td>
<td>75.5–93.5</td>
</tr>
<tr>
<td></td>
<td>4, 5, 17</td>
<td>88.9</td>
<td>77.8–94.8</td>
</tr>
<tr>
<td>3D features</td>
<td>4, 5, 16, 17</td>
<td>89.9</td>
<td>77.8–94.8</td>
</tr>
<tr>
<td>4D features</td>
<td>4, 5, 6, 15, 16</td>
<td>94.4</td>
<td>84.9–98.1</td>
</tr>
<tr>
<td>5D features</td>
<td>4, 5, 6, 15, 16, 17</td>
<td>90.7</td>
<td>80.0–96.0</td>
</tr>
<tr>
<td>6D features</td>
<td>1, 4, 5, 6, 15, 16</td>
<td>94.4</td>
<td>84.9–98.1</td>
</tr>
</tbody>
</table>

We applied the methodology proposed in our previous work [14] to the dataset used in this study, resulting in a classification accuracy of 77.8%. As can be seen, there is a significant increase in all of the performance metrics, which is due to the introduction of

Please cite this article in press as: Kotu LP, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. Artif Intell Med (2015), http://dx.doi.org/10.1016/j.artmed.2015.06.001
new features and the use of the SVM classifier (94.4%). Our group’s previous work [10] reported an AUC of 0.92 for classifying HAG and LAG patients using Bayes classifier with two texture features and LVEF. In our study, the texture feature combinations like (4, 15) itself gave an AUC of 0.854 and 0.818 with SVM and k-NN, respectively. Localization features alone did not perform well. However, when added with the texture features they gave good classification performance. Without using LVEF feature, we got an AUC of 0.97 with the SVM classifier for a combination of texture and localization features (4, 5, 6, 15, 16, 17).

5.2. Results from Experiment 2

The classification results from Experiment 2, obtained in WEKA, are tabulated in Table 5. The first column in Table 5 gives the classifier names, and the parameters used. The second column shows if SMOTE is used or not. The third and forth columns show the average accuracy and the AUC values. The last column shows the proposed feature set if all the data were used as training set. Different classifiers were used in different parameter settings, and the best results for different number of folds in the internal CVs are tabulated in this manuscript. The external CV is always ten folds.

In Table 5, without SMOTE, k-NN and SVM classifiers gave the highest average accuracy with 83.3% and 92.6%, respectively. The different trails of k-NN and SVM show that the texture features are important as well as localization features. In the second row of the table, k-NN with 8 neighbors and city block distance gave the average accuracy with 75.9% (AUC = 0.788, no SMOTE). This result is consistent with the results from Experiment 1, presented in Table 2 which provided an average accuracy of 74.3% (AUC = 0.758). The slight difference in the average accuracy and AUC is due to the difference in selected feature set for Experiment 1 (4, 5, 6, 15, 16, 17) and Experiment 2, where the feature set can be different for each fold. The method used on all the training data provides the set: (1, 2, 3, 4, 11, 17). Similarly, the SVM classifier with a third order polynomial kernel without SMOTE gave an average AUC of 0.776 and 0.921 for Experiment 1 and Experiment 2, respectively. The difference in AUC values is due to the fixed feature set used in Experiment 1, whereas in Experiment 2 the feature set can vary over the folds. Since, the selected feature set has been fixed in Experiment 1 the AUC values did not increase beyond a level with varying C parameters for the different kernel functions tested in the SVM classifier. In Experiment 2, SVM classifier with varying C parameter gave high average accuracy and AUC because of wrapper based feature selection. The AD tree and random classifiers without SMOTE for different feature combinations gave an AUC of 0.76 and 0.885, respectively.

In Experiment 2, it is evident that localization features are also important as well as the texture features. All the experiments presented in this paper showed that the information present in the CMR image-based features can be useful for distinguishing the HAG from LAG patient groups. From the AUC values without SMOTE, we can infer that the SVM classifier gave the highest AUC value with 0.921, followed by random forest with an AUC value of 0.885. From Table 5 it can be seen that using SMOTE for re-sampling improves the results using k-NN, AD tree and, random forest, as expected. The results using SVM, however, are not affected by the re-sampling. With or without the use of SMOTE, the best results are achieved with SVM classifier.

5.3. Left ventricular ejection fraction and scar size

Section 2 explained that the process of categorizing patients into the HAG and LAG classes is based on LVEF as well as cardiac history (previous ventricular tachycardia), and possibly other biomarkers. The scar size calculated in reference to the body surface area (with the unit of grams per m²) is also a known risk marker.

Table 5

<table>
<thead>
<tr>
<th>Classifier and parameters</th>
<th>SMOTE or no SMOTE</th>
<th>Average accuracy (%)</th>
<th>AUC</th>
<th>Proposed feature set</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-NN, 10N, KD tree search,</td>
<td>No SMOTE</td>
<td>83.3</td>
<td>0.836</td>
<td>2, 3, 4, 17</td>
</tr>
<tr>
<td>Euclidean distance, internal CV-5</td>
<td>SMOTE</td>
<td>80.9</td>
<td>0.814</td>
<td>2, 3, 4, 17</td>
</tr>
<tr>
<td>k-NN, 8N, linear NN search, city block distance, internal CV-8</td>
<td>No SMOTE</td>
<td>75.9</td>
<td>0.788</td>
<td>1, 2, 3, 4, 11, 17</td>
</tr>
<tr>
<td>k-NN, 10N, linear NN search, kNN and SVM, 2nd order polynomial, kernel C = 5, internal CV-8</td>
<td>No SMOTE</td>
<td>80.9</td>
<td>0.843</td>
<td>2, 3, 4, 17</td>
</tr>
<tr>
<td>SMO SVM, 3rd order polynomial, kernel C = 5, internal CV-5</td>
<td>No SMOTE</td>
<td>81.5</td>
<td>0.857</td>
<td>2, 3, 4, 16, 17</td>
</tr>
<tr>
<td>Euclidean distance, internal CV-7</td>
<td>SMOTE</td>
<td>85.3</td>
<td>0.869</td>
<td>3, 4, 6, 7, 10, 11, 16, 17</td>
</tr>
<tr>
<td>SMO SVM, 4th order polynomial, kernel C = 2, internal CV-5</td>
<td>No SMOTE</td>
<td>88.9</td>
<td>0.881</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10</td>
</tr>
<tr>
<td>AD tree, boosting iterations-12, internal CV-8</td>
<td>SMOTE</td>
<td>85.3</td>
<td>0.853</td>
<td>1, 2, 4, 5, 6, 8, 9</td>
</tr>
<tr>
<td>Random forest, no. of trees = 100, internal CV-8</td>
<td>No SMOTE</td>
<td>77.8</td>
<td>0.76</td>
<td>2, 4, 8, 12, 17</td>
</tr>
<tr>
<td>Random forest, no. of trees = 200, internal CV-5</td>
<td>SMOTE</td>
<td>80.9</td>
<td>0.86</td>
<td>3, 4, 6, 9, 10</td>
</tr>
<tr>
<td>SVM, polynomial, order = 2, C = 1.25</td>
<td>81.5</td>
<td>0.885</td>
<td>2, 3, 4, 6, 8, 10, 16, 17</td>
<td></td>
</tr>
<tr>
<td>SVM, polynomial, order = 2, C = 1.25</td>
<td>79.6</td>
<td>0.853</td>
<td>2, 3, 4, 5, 6, 8, 10, 16, 17</td>
<td></td>
</tr>
<tr>
<td>SVM, polynomial, order = 2, C = 1.25</td>
<td>90.7</td>
<td>80.0–96.0</td>
<td>0.941</td>
<td></td>
</tr>
<tr>
<td>SVM, polynomial, order = 2, C = 1.25</td>
<td>87.0</td>
<td>75.6–93.6</td>
<td>0.928</td>
<td></td>
</tr>
</tbody>
</table>
Note that this is different from the relative scar volume introduced in Section 3.1. The LVEF and scar size of the dataset are available; thus, we can analyze the discriminative value of these factors for comparison. When we study LVEF with the corresponding class label across all patients in the data material, and use logistic regression to do a receiver operating characteristic (ROC) analysis, the resulting area under the receiver operating curve (AUC) is found to be 0.937. The high value is understandable since the LVEF played an (the most) important role in distinguishing the classes. A similar analysis is done on the scar size, giving AUC at 0.841. Thereafter the features are used in a leave-one-out CV scheme with k-NN and SVM classifiers. Table 4 shows the classification results of the LVEF and scar size as separate 1D features, and together as a 2D feature vector. As expected, the combination of the two features gives better classification performance (accuracy: 90.7%, AUC: 0.941).

It should be noted that our “ground-truth” (i.e. the categorization of HAG and LAG patients) is partly based upon these features. Thus, it is not fair/correct to include them in a proper test. Table 4 is solely intended for curiosity purposes or for a comparison with what we can accomplish using only the information from the CMR images. Note that the classification results obtained by using the proposed CMRI-based features are not much lower than the classification results using LVEF and scar size. These results indicate the potential of the CMRI-based features, and lead us to believe that features calculated from the CMR images can be used to identify HAG patients who subsequently require ICD implantation.

6. Conclusion and future work

This work is an attempt to identify post-MI patients with a high risk of experiencing fatal arrhythmias, using a combination of machine learning tools and image processing techniques. We proposed CMR image-based features that capture the size, location, and heterogeneity of the scarred myocardium to classify high- and low-risk patients. Of all of the different combinations of the proposed features, texture features based on scatter gradient and LBPs along with the localization features demonstrate good discriminative power. The fact that the classification performance does not drastically change across the different classifiers (which use different classification philosophies) in Experiment 1 and Experiment 2, suggests that these features are robust to the choice of classifier and can be used in medical decision support systems. Classification performance for the HAG and LAG using LVEF and scar size (accuracy: 90.7%, AUC: 0.941) is comparable to the proposed classification scheme in Experiment 1 (accuracy: 94.44%, AUC: 0.965) and Experiment 2 (accuracy: 92.6%, AUC: 0.921). It is important to note that features extracted from the CMR images are not currently used when categorizing patients as HAG or LAG. The results presented in this paper should serve as a strong indication that such features should be included in future arrhythmia risk stratification guidelines or decision-making processes.

The useful CMR image-based features can be combined with recognized markers like LVEF and scar size in the future classification systems. The proposed features need to be tested on a new and larger dataset. Our proposed scheme requires a human input (of a cardiologist) in the loop to mark the scar boundaries. Our group, as well as others, have worked on constructing systems for automatic segmentation of the myocardium [28-31] as well as the scar areas [32-34]. In future work, we will investigate the performance of the proposed high-risk/low-risk classification when the regions are based on automatic segmentation, which would make the entire process fully automated. In addition we want to further explore the scarred myocardium in order to improve these features and extract new features from the spectral domain using Gabor filters and local phase quantization (LPQ) [35].

Conflict of interests

The authors declare that they have no conflict of interests.

Acknowledgements

The University of Stavanger financed the Ph.D. work of Lasya Priya Kotu and this manuscript is her part of Ph.D.

References


Please cite this article in press as: Kotu LP, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. Artif Intell Med (2015), http://dx.doi.org/10.1016/j.artmed.2015.06.001