J Digit Imaging (2015) 28:160–178
DOI 10.1007/s10278-014-9730-z

Optimal Query-Based Relevance Feedback in Medical Image Retrieval Using Score Fusion-Based Classification

Mohammad Behnam · Hossein Pourghassem

Published online: 23 September 2014

Abstract In this paper, a new content-based medical image retrieval (CBMIR) framework using an effective classification method and a novel relevance feedback (RF) approach are proposed. For a large-scale database with diverse collection of different modalities, query image classification is inevitable due to firstly, reducing the computational complexity and secondly, increasing influence of data fusion by removing unimportant data and focus on the more valuable information. Hence, we find probability distribution of classes in the database using Gaussian mixture model (GMM) for each feature descriptor and then using the fusion of obtained scores from the dependency probabilities, the most relevant clusters are identified for a given query. Afterwards, visual similarity of query image and images in relevant clusters are calculated. This method is performed separately on all feature descriptors, and then the results are fused together using feature similarity ranking level fusion algorithm. In the RF level, we propose a new approach to find the optimal queries based on relevant images. The main idea is based on density function estimation of positive images and strategy of moving toward the aggregation of estimated density function. The proposed framework has been evaluated on ImageCLEF 2005 database consisting of 10,000 medical X-ray images of 57 semantic classes. The experimental results show that compared with the existing CBMIR systems, our framework obtains the acceptable performance both in the image classification and in the image retrieval by RF.

Keywords Content-based medical image retrieval (CBMIR) · Visual features · Semantic concepts · Score fusion · Relevance feedback · Nonparametric estimation · Optimal queries

M. Behnam · H. Pourghassem (✉)
Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, 517, Najafabad, Isfahan, Iran
e-mail: h_pourghasem@iaun.ac.ir

Introduction

The huge volume of multimedia products especially digital images in different domains such as education, medical, online shopping, and entertainment has led to increasing demands in powerful indexing and retrieval tools. In image retrieval domain, the early retrieval systems carried out based on annotating text to images in database, but this method did not provide user satisfaction, because the annotation process not only is subjective task but also is tedious work and time-consuming [1]. Moreover, a series of words do not fully describe the image content. Due to these reasons, efforts were directed toward content-based image retrieval (CBIR) [2–4]. In medical applications, with the rapid growth of technology and appearance of different imaging systems, and consequently increase in medical images, utilization of content-based medical image retrieval (CBMIR) systems became unavoidable due to effective and accurate diagnosis, research, and medical education [5].

In general, a CBIR system extracts low-level visual features and the images are retrieved based on these features [4]. The performance of CBIR systems based on only low-level contents is low because these visual features often fail to describe the high-level semantic concepts in user's mind. Liu et al. [6] identified five major categories of the state-of-the-art techniques in narrowing down this gap which applying machine learning techniques to associate low-level features with query concepts are most commonly used in the literature. These techniques reduce the semantic gap to some extent, but the use of these techniques alone does not fully meet user needs. Thus, applying the combination of techniques is essential to achieve the high-performance retrieval.

In a large database of medical images with diverse collection of different modalities, query image classification in the first step is inevitable due to firstly, reducing the computational complexity and secondly, increasing influence of data

fusion by removing unimportant data and focus on the more valuable information. Hence, by filtering out the images of irrelevant classes and reducing the search space, the performance of CBIR systems is improved significantly [7–10]. Pourghassem and Ghassemian [7] proposed a hierarchical medical image classification method using a merging scheme and multilayer perceptron classifiers. In this system, semantic classes are created from overlapping classes in the database. Fesharaki and Pourghassem [8] proposed a classification based on Bayesian rule. Also, Rahman et al. [9] proposed a medical image retrieval system with probabilistic multiclass support vector machine (SVM) classifiers and adaptive similarity fusion approach based on a linear combination of individual feature level similarities. In this system, the feature weights are calculated by considering both the precision and the rank of top retrieved relevant images as predictor by SVMs. Ultimately, Greenspan and Pinhas [10] proposed a continuous and probabilistic image representation scheme using Gaussian mixture modeling along with information-theoretic image matching via the Kullback-Leibler measure. This system was used to match and categorize X-ray images by body regions.

In this paper, we find probability distribution of classes in the database using Gaussian mixture model (GMM) for each feature descriptor [11, 12]. Since the within-class variabilities in medical image database is high, the GMM can properly express this diversity using a probability density function with multiple Gaussian functions. Since each extracted feature from images just characterizes certain aspect of image content, we employ multiple features. As regards, some features are more effective in the image content description of each class; thus, a fusion algorithm should determine the influence of these features in the image classification. Then, by using the fusion of obtained scores from the dependency probabilities, the most relevant clusters are identified for a given query. For this purpose, we propose two types of negativity and positivity scores of query belonging to each class and then calculate total scores based on these two score types. Afterwards, using an appropriate score threshold value, the most relevant clusters are identified.

In general case, there are two main approaches to fusion: early fusion, where multiple image descriptors are composed to form a new one before indexing, and late fusion, where ranking results from individual descriptors are fused during image retrieval [13]. In the image retrieval application, some commonly used methods for rank-based fusion which originates from social theory in voting are Borda count (BC), Inverse Rank Position (IRP), and Leave Out (LO) algorithms for fusing the similarity scores of image feature [14]. In the image retrieval phase, we use the IRP method to retrieve the relevant images.

Due to involve the users' intention in the retrieval procedure, many algorithms have been presented to integrate relevance feedback (RF) to retrieval algorithm in the literature [15–23]. The query refinement strategies are one of the most commonly used methods which modifies the query image in the RF phase. Su et al. [15] made use of the three kinds of query refinement strategies including the following: query point movement (QPM), query reweighting (QR), and query expansion (QEX), to converge the search space toward the user's intention effectively. The well-known QPM approach proposed by Rocchio [16] move the current query point to relevant images and is far away from irrelevant images.

The QR firstly proposed by Rui et al. [17, 18], which convert image feature vectors to weighted term vectors (information retrieval domain). Since the QR and QPM cannot completely cover the user's interest, the QEX has been presented in the solution space of RF. For example, the modified version of MARS [19] groups the similar relevant points into several clusters and selects good representative points from these clusters to construct the multipoint query.

In medical image application, Rahman et al. [20] introduced a CBMIR system using a probabilistic multiclass support vector machine (SVM) and fuzzy c-mean clustering for categorization and updating the parameter of the query class based on the positive feedback images. Pourghasssem and Daneshvar [21] proposed a framework for the image retrieval of a large-scale database of medical X-ray images. This system is designed based on query image classification into several pre-specified homogeneous classes using a merging scheme and an iterative classification. Also, this system proposed an RF mechanism using a synthetic distance measure based on the weighted Euclidean distance measure and GMM-based dependency probability similarity measure. In the study of Rahman et al. [22], a learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM has been introduced. This system applied a RF-based similarity fusion technique which feature weights are updated at each iteration by considering both the precision and rank order information of relevant images. Yang et al. [23] proposed a boosting framework for distance metric learning to medical image retrieval that aims to preserve both visual and semantic similarities. The boosting framework firstly learned a binary representation using side information, in the form of labeled pairs, and then computed the distance as a weighted hamming distance using the learned binary representation.

In this paper, we propose a novel RF algorithm based on reweighting the relevant image and finding optimal queries. The proposed algorithm is based on finding new queries which are closer to the user intention (optimal query) using probability density function of relevant images. For this purpose, the optimal queries are obtained using a nonparametric estimation of density function associated with positive (relevant) images and an objective function in optimization procedure. In other words, we use the combination of Parzen window as an appropriate kernel density estimators [24] and

genetic algorithm (GA) [25] as an optimization method. The strategy of moving the relevant images toward global maximums of estimated density function is an idea to find the optimal queries in a multidimensional space. Unlike QPM [16], which the query refinement was performed by averaging method, we use the probability density function of relevant images for refining the current queries and also expanding the optimal queries. Hence, in this paper, we make an appropriate fitness function based on estimated density function and use GA for our optimization task. Accordingly, the optimal queries are found by applying a simple scenario on the optimized relevant points. However, in the image retrieval domain, several techniques based on GA have been presented in the literature, but these papers have mostly focused on feature selection and dimensionality reduction [26, 27]. In the study of Steji´c et al. [28], a GA-based RF has been presented to infer the (sub-)optimal assignment of region and feature weights, which maximizes the similarity between the query image and the set of relevant images. In the study of Lai and Chen [29], a user-oriented mechanism for CBIR method based on an interactive genetic algorithm (IGA) has been proposed. To reduce the gap between the retrieval results and the users' expectation, Lai and Chen [29] employed the IGA to help the users in image identification.

The rest of this paper is organized as follows. In the section "Proposed Content-Based Image Retrieval Framework," the proposed content-based retrieval framework is presented. In this section, first image feature descriptors are explained. Then, the details of our score fusion-based classification algorithm are presented. The section "Optimal Query-Based Relevance Feedback Algorithm" presents the proposed optimal query-based RF algorithm, followed by the "Experimental Results" section. The "Discussion and Comparison with Other Works" section provides a discussion on the experimental results. Finally, the "Conclusion" section concludes the paper with some conclusive remarks.

Proposed Content-Based Image Retrieval Framework

Figure 1 shows the block diagram of proposed framework for content-based medical image retrieval using our score fusion-based classification algorithm. In this framework, visual features are extracted independently from images in database and then are stored in repositories. To obtain uncorrelated feature vectors and reduce the space and time complexities for GMM-based distribution function estimation of classes, we use principal component analysis (PCA), which is one of the most popular techniques for dimensionality reduction. For a query image, the feature vector is provided to calculate the dependency probability of each feature descriptor to classes. Afterwards, using a new technique for scoring classes based on fusion of dependency probabilities, the most relevant clusters for the given query are identified. In the following, details of each module in our proposed content-based image retrieval framework are described.

Feature Descriptors

We use several techniques based on local binary pattern (LBP) such as uniform LBP, local binary count (LBC), and complete LBC (CLBC) to extract texture features. Moreover, one-dimensional Fourier descriptors (1-D FDs) and 2-D modified generic Fourier descriptor (MGFD) are used to extract the shape features from the medical images.

Texture Features

LBP features are considered as one of the best texture features as they are invariant to monotonic changes in grayscale and can be extracted faster compared to other texture features. The LBP coding strategy can be described as follows:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

where g_c represents the grayscale of center pixel, $g_p (p=0,\cdots, P-1)$ denotes the grayscale of neighbor pixel on a circle of radius R, and P is the total number of neighbors. In this paper, we use $P=8$, $R=1$.

Ojala et al. [30] proposed multiresolution approach to grayscale and rotation invariant texture classification based on LBP. They presented a uniformity measure U, which corresponds to the number of spatial transitions (bitwise 0/1 changes) in the pattern as follows:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s\left(g_p - g_c\right) & if \quad U\left(LBP_{P,R}\right) \leq 2 \\ P+1 & otherwise \end{cases} \tag{2}$$

where

$$U\left(LBP_{P,R}\right) = \left|s(g_{P-1} - g_c) - s(g_0 - g_c)\right| + \sum_{p=1}^{P-1} \left|s\left(g_p - g_c\right) - s\left(g_{p-1} - g_c\right)\right| \tag{3}$$

Superscript ^{riu2} reflects the use of rotation invariant "uniform" patterns that have U value of at most 2. By definition, exactly $P+1$ uniform binary pattern can occur in a circularly symmetric neighbor set of P pixels. Equation 2 assigns a unique label to each of them, corresponding to the number of "1" bits in the pattern ($0 \rightarrow P$), while the "nonuniform" patterns have ($P+1$) label. After the LBP^{riu2} code of each

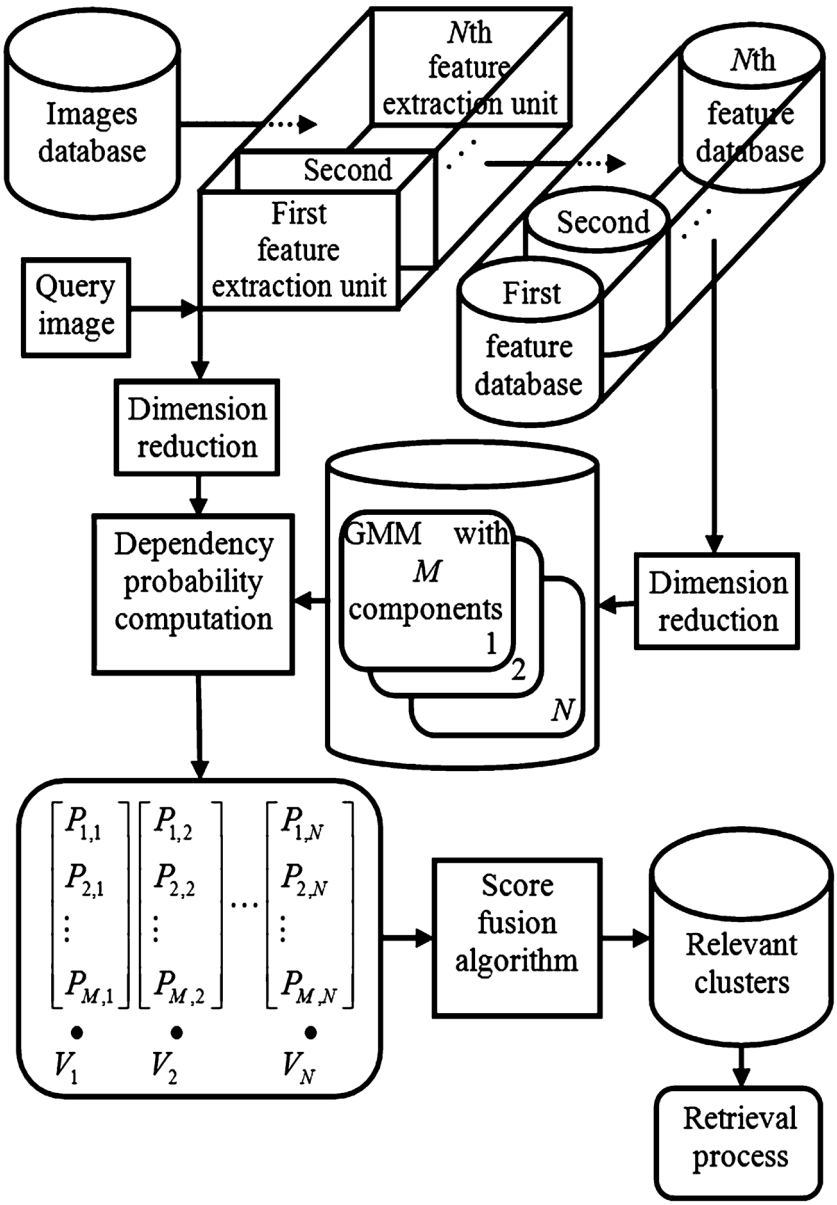

Fig. 1 Block diagram of proposed content-based image retrieval framework

pixel is defined, a histogram will be built to represent the texture feature.

Zhao et al. [31] proposed LBC which only count the number of value 1's in the binary neighbor sets instead of encoding them. The computing process for the LBC can be defined as follows:

$$LBC_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right), \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4}$$

where g_c, g_p, P, and R are the variables, as defined in Eq. 1. Zhao et al. [31] believed that although the LBC codes do not represent visual microstructures, these codes will cause that macroscopic textural structure and can be characterized by statistics of the microstructures. Guo et al. [32] proposed the completed LBP (CLBP) by combining the conventional LBP with the measures of local intensity difference and central grayscale. They proposed CLBP descriptor by decomposing the image local differences into two complementary components, i.e., the signs and magnitudes. Similar to the CLBP, Zhao et al. [31] proposed CLBC to extract completed local textural information, including CLBC-sign (CLBC_S) and CLBC-magnitude (CLBC_M). Generally, the CLBC_S is equal to the original LBC described in Eq. 4. In order to code the CLBC_M in a consistent format with the CLBC_S, the CLBC_M can be defined as follows:

$$CLBC\_M_{P,R} = \sum_{p=0}^{P-1} s\left(m_p - c\right), m_p = \left|g_p - g_c\right| \tag{5}$$

where g_c, g_p, and $s(x)$ are defined as the parameters of Eq. 1, and c denotes the mean value of m_p in the whole image.

Shape Features

The methods for shape description can be categorized into either boundary-based or region-based methods. 1-D FDs [33, 34] are most popular and wildly used as contour-based shape descriptors. Generally, 1-D FDs are obtained through Fourier transform (FT) on a shape signature function derived from shape boundary coordinates. Typical shape signature functions are centroid distance function, complex coordinate function, and cumulative angular function. However, finding the accurate and meticulous boundary is still a complex and time-consuming problem and requires the use of different segmentation techniques, but our intention in this paper is not to focus on these methods. Since, applying a single method on all images in database does not provide the desired results due to changes in the brightness and contrast even for images related to a specific class. Hence, the shortcomings of this issue and consequently obtaining features with less precision for some images will be compensated by using data fusion techniques and RF. Therefore, we use simple image processing techniques such as adaptive local histogram [8], edge detection filters, adaptive thresholding method [35], and morphological operators to extract the boundary.

Suppose that $(x(t),y(t))$ are a sequence that represent the coordinates of boundary pixels of a 2-D object where $t=0, 1,\cdots,N_p-1$ and N_p is the total number of pixels on the boundary. The centroid distance function is given by the distance of boundary points from the centroid (x_c,y_c) of the shape:

$$r(t) = \left([x(t)-x_c]^2 + [y(t)-y_c]^2\right)^{1/2}, \quad t = 0,\ 1,\ \ldots, N_p-1 \tag{6}$$

where

$$x_c = \frac{1}{N_p}\sum_{t=0}^{N_p-1} x(t)\,, \quad y_c = \frac{1}{N_p}\sum_{t=0}^{N_p-1} y(t) \tag{7}$$

The complex coordinate function is obtained by representing the horizontal coordinates of the boundary pixels as real part and vertical coordinates as imaginary part of complex numbers:

$$c(t) = (x(t)-x_c) + j(y(t)-y_c) \tag{8}$$

The angular function measures the angular direction of tangent line as a function of arc length. The main problem is that the angular function has discontinuities even for smooth curves. In the study of Zahn and Roskies [34], this problem is eliminated by considering a normalized form of the cumulative angular function φ^* as follows:

$$\varphi^*(t) = \varphi\left(\frac{Lt}{2\pi}\right) + t \tag{9}$$

where $\varphi^*(0)$ and $\varphi^*(2\pi)$ is identical to zero, and ϕ^* is invariant under translations, rotations, and changes of perimeter of shape L. The cumulative angular function $\phi(l)$ in Eq. 9 is defined as the net amount of angular bend between starting point and point l, $0\le l\le L$, i.e.,

$$\varphi(l) = \theta(l)-\theta(0) \tag{10}$$

where $\theta(0)$ and $\theta(l)$ are the angular directions at starting point and point l, respectively. An example of signature functions of hand image shape is shown in Fig. 2.

Here, we first implemented 1-D FT on boundary and reconstructed the shape from lower order Fourier component. We re-sample the boundary of each object to 256 samples by a uniform sampling function for equalizing the length of extracted shape features of all objects in our database.

To achieve rotation invariance, the phase of components is discarded and only the amplitudes of complex coefficients are used. To achieve scale invariance, the magnitudes of coefficients is normalized by the magnitude of DC component or the first nonzero coefficient. The translation invariance and start point invariance are obtained directly from the contour representation and FT, respectively. As a result, the following feature vectors for $r(t)$, $c(t)$, and $\varphi^*(t)$ are obtained, respectively:

$$f_r = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \ldots, \frac{|F_{m/2}|}{|F_0|}\right] \tag{11}$$

$$f_c = \left[\frac{|F_{-(m/2-1)}|}{|F_1|}, \ldots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \ldots, \frac{|F_{m/2}|}{|F_1|}\right] \tag{12}$$

$$f_\phi = \left[|F_1|, |F_2|, \ldots, |F_{m/2}|\right] \tag{13}$$

where F_i denotes the ith component of FT coefficients. In Eqs. 11 and 13, only the positive frequency axes are considered because the $r(t)$ and $\varphi^*(t)$ are real and, therefore, their FTs exhibit symmetry. Also by definition, $\varphi^*(t)$ is the scale invariance and does not need to be normalized. In Eq. 12, the first nonzero frequency component (F_1) was used for normalizing

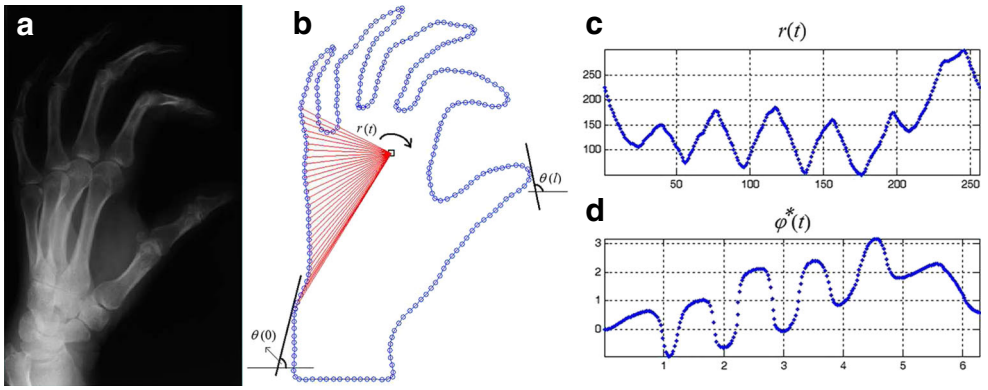

Fig. 2 **a** Original image, class 47: hand, other orientation. **b** Extracted contour after smoothing and down-sampling; $\theta(0)$ is the absolute angular direction at the starting point and $\theta(l)$ is angular direction at point l, $r(t)$ is centroid distance function at point t, the *square marker* represents the centroid of shape. **c** Representation of centroid distance function with respect to t. **d** Representation of normalized form of the cumulative angular function with respect to t

because the DC coefficient (F_0) is dependent on the position of a shape and therefore is discarded.

However, mentioned Fourier descriptors assume the knowledge of shape boundary information which may not be available in general situations; 1-D FDs cannot capture shape interior content which is important for shape discrimination. To overcome the drawbacks of existing shape representation techniques, MGFD has been proposed [36, 37]. Applying 2-D FT on a shape image in Cartesian space directly to derive FD is not practical because the features captured by 2-D FT are not rotation invariant. Therefore, a modified polar FT (MPFT) [36] has been derived by treating the polar image in polar space as a normal two-dimensional rectangular image in Cartesian space. Figure 3 demonstrates the rectangular polar image. Figure 3a is the original shape image in polar space, and Fig. 3b is the rectangular polar image plotted into Cartesian space.

Therefore, if we apply 2-D FT on this rectangular image, the polar FT has the similar form to the normal 2-D discrete FT in Cartesian space. Consequently, the MPFT is obtained as follows:

$$MPFT(\rho,\varphi)=\sum_r\sum_i f(r,\theta_i)\exp\left[j2\pi\left(\frac{r}{R_1}\rho+\frac{2\pi i}{R_2}\varphi\right)\right] \tag{14}$$

where $0\le r\le R_1$ and $\theta_i=i(2\pi/R_2)(0\le i\le R_2)$; $0\le\rho\le R_1$, $0\le\varphi\le R_2$. R_1 and R_2 are the radial and angular resolutions. (r,θ) is the polar coordinates in image plane, and (ρ,φ) is the polar coordinates in frequency plane. The acquired Fourier coefficients are translation invariant. The rotation and scaling invariance in MGFD are achieved by the following:

$$\text{MGFD}=\left\{\frac{|pf(0,0)|}{\text{area}},\frac{|pf(0,1)|}{|pf(0,0)|},\ldots,\frac{|pf(0,n)|}{|pf(0,0)|},\ldots,\frac{|pf(m,0)|}{|pf(0,0)|},\ldots,\frac{|pf(m,n)|}{|pf(0,0)|}\right\} \tag{15}$$

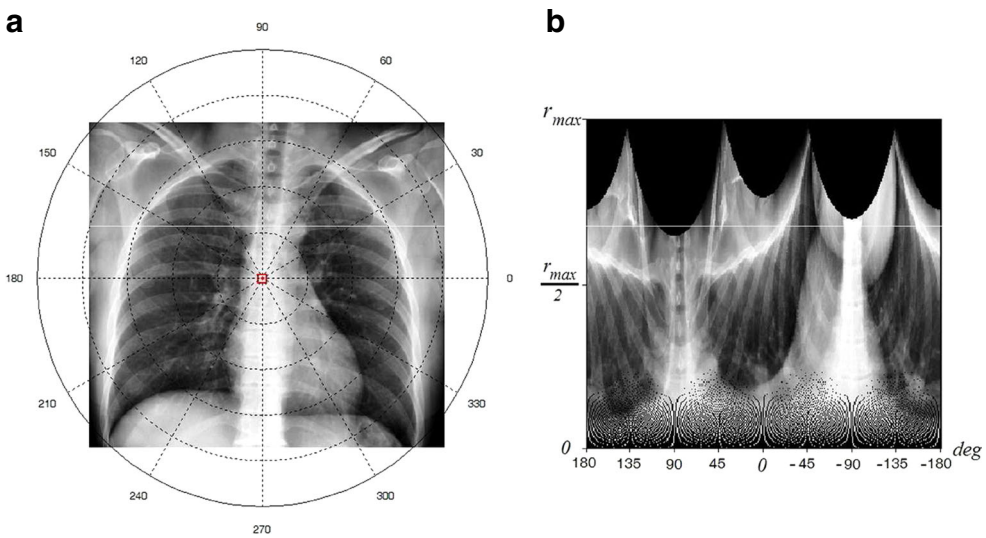

Fig. 3 **a** Original image (class 12: chest) in polar space with centroid as origin, the *square marker* represents the centroid of shape. **b** Polar image of **a** plotted into Cartesian space

where area is the area of bounding circle in which the polar image resides and for simplicity MPFT replaced with *pf*. The *m* is the maximum number of radial frequencies selected, and *n* is the maximum number of selected angular frequencies. In this work, we use 49 MGFD features from equal radial and angular frequencies.

Score Fusion-Based Classification

In this section, the details of our algorithm for filtering out the images of irrelevant classes and reducing the search space in order to optimize the retrieval process are described. For this purpose, we first find GMM of classes in the database for each extracted feature descriptor. Since, each one of the feature descriptors describe various aspects of images, the Gaussian model for each feature vectors differs from each other and have distinctive distributions. Then, scores of query belonging to each class are calculated using distribution function-based dependency probabilities of feature descriptors. In the first step, reducing the dimensions of feature descriptors to find distribution function is inevitable. Thus, the dimensions of feature descriptors are reduced by PCA.

Dimensional Reduction

To reduce the dimension of feature vectors, we use PCA [38] that is defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximized. In our application, PCA is applied on each feature (with dimension $d>1$) of N extracted features separately so that a well-conditioned covariance matrix obtained in the reduction process.

Distribution Function-Based Dependency Probability

In a general manner, the GMM distribution can be written as a linear superposition of Gaussians [39] in the form

$$p(x)=\sum_{m=1}^{M}\ \pi_m \frac{1}{\sqrt{(2\pi)^d|\Sigma_m|}} e^{-1/2(x-\mu_m)^T\Sigma_m^{-1}(x-\mu_m)} \tag{16}$$

where M is the number of components in the model, and $x\in R^d$ is a d-dimensional feature vector. Each Gaussian density is called a component of the mixture and has its own mean μ_m and covariance Σ_m matrix. The parameters π_m in Eq. 16 are called mixing coefficients where $0\le\pi_m\le1$ and $\sum_{m=0}^{M}\pi_m=1$.

In a conventional way an iterative algorithm such as expectation maximization (EM) is employed to estimate the parameters of a GMM [11], but in our application, these parameters simply has been calculated in a supervised manner from predefined classes of training data. Accordingly, having M classes and N feature descriptors, we use M Gaussian components with the same π_m and specified means and covariance matrixes for each feature descriptor. Afterwards, posterior probabilities of query belonging to each class for each feature descriptor are calculated. Suppose that $P_{m,n}$ is the dependency probability of nth feature descriptor to the class m for an unknown pattern (query image). We define this probability as

$$P_{m,n}=\pi_m \frac{1}{\sqrt{(2\pi)^d|\Sigma_{m,n}|}} e^{-1/2\left(x_n-\mu_{m,n}\right)^T\Sigma_{m,n}^{-1}\left(x_n-\mu_{m,n}\right)} \tag{17}$$

where x_n is the query feature vector based on nth feature descriptor ($n=1,\cdots,N$), $\mu_{m,n}$ and $\Sigma_{m,n}$ are the mean vector and covariance matrix of mth class ($m=1,\ldots,M$) based on nth feature descriptor, respectively, and π_m is the mixing coefficient of mth component (class) and we set to $1/M$. The $P_{m,n}$ in Eq. 17 is the posterior probability of query belonging to each class based on each feature descriptor which we have named it as dependency probability. As a result, by calculating the dependency probabilities for all classes and all feature descriptors, N vectors with M elements are obtained as follows:

$$V_1=\begin{bmatrix}P_{1,1}\\ \vdots\\ P_{M,1}\end{bmatrix},\ V_2=\begin{bmatrix}P_{1,2}\\ \vdots\\ P_{M,2}\end{bmatrix},\ \ldots,\ V_N=\begin{bmatrix}P_{1,N}\\ \vdots\\ P_{M,N}\end{bmatrix} \tag{18}$$

Score Fusion Algorithm

In this section, for a given query, the most relevant classes are identified using a new scheme based on fusion of dependency probabilities. Our fusion scheme is based on the scoring of each class according to the sign of differences between priori and posteriori probabilities.

Suppose that P_m is a priori probability of query belonging to the class m and is equal to $1/M$ for all classes. If calculated posterior probabilities are less than $1/M(P_{m,n}<P_m)$, the scores of each class are defined as follows:

$$Score_m^-=\sum_{n=1}^{N} Score_{m,n}^- \tag{19}$$

where

$$Score_{m,n}^-=\log\left(-\left(\frac{1}{P_{m,n}-P_m}+\frac{1}{P_m}\right)+\varepsilon_0\right) \tag{20}$$

where $m=1,\cdots,M$ is index of class, and $n=1,\cdots,N$ is index of feature descriptor. The $Score^{-}_{m,n}$ is a score for query belonging to mth class when nth feature descriptor is used for the distribution function estimation. The ε_0 in Eq. 20 is a very small positive number that is used to avoid of infinity. Subsequently, $Score^{-}_{m}$ is obtained from the total negativity score for query belonging to mth class. According to Eq. 20, it is clear that if $P_{m,n}$ is closer to P_m, then $Score^{-}_{m}$ take a larger value, and if $P_{m,n}$ is much farther away from the P_m, then lower score is added to $Score^{-}_{m}$. Note, negative value of $Score^{-}_{m,n}$ is replaced with zero score. Figure 4a shows the behavior of the negativity scoring function with respect to $P_{m,n}$.

On the other side, if the posterior probabilities are greater than $1/M(P_{m,n}\geq P_m)$, we calculate the score for each class as

$$Score^{+}_{m}=\sum_{n=1}^{N} Score^{+}_{m,n} \tag{21}$$

where

$$Score^{+}_{m,n}=10^{\left(\frac{1}{1-P_{m,n}}\right)} \tag{22}$$

where $Score^{+}_{m,n}$ is a score for query belonging to mth class when nth feature descriptor is used for the distribution function estimation. Also, $Score^{+}_{m}$ is the total positivity score for query belonging to mth class.

In Eq. 22, the closer values of $P_{m,n}$ to P_m give the lower scores. However, with increasing the posterior probabilities, the scores are the larger than value. The behavior of positivity scoring function with respect to $P_{m,n}$ has been shown in Fig. 4b. In this case, if the dependency probability $P_{m,n}$ is far more than 0.5, then a very high positivity score is given to mth class which indicates that the probability of query belonging to mth class is very high. Ultimately, the total scores for query belonging to each class are calculated as follows:

$$Score_m = Score^{-}_{m} + Score^{+}_{m} \tag{23}$$

If the calculated score in Eq. 23 is greater than a given threshold, then the class m will be selected as one of the clusters associated with the query image. The proposed score fusion algorithm to calculate the scores of each class is summarized as follows:

```
for each class m
    P_m=1/M
    counter_1=0, counter_2=0
    Score^-=0, Score^+=0
    for each feature descriptor n
        if (P_m,n-P_m<0)
            counter_1=counter_1+1
            if (log(-(1/(P_m,n-P_m)+1/P_m)+ _0)<0)
                Score^-(counter_1)=0
            else
                Score^-(counter_1)=log(-(1/(P_m,n-P_m)+1/P_m)+ _0)
            end
        else
            counter_2=counter_2+1
            Score^+(counter_2)=10^(1/(1-P_m,n))
        end
    end
    Score(m)=ΣScore^-+ΣScore^+
end
```

Retrieval Process

In retrieval phase, the similarity between query image and images in the relevant clusters must be measured. For this purpose, Euclidean distance measure [40] is used separately on all feature descriptors to calculate

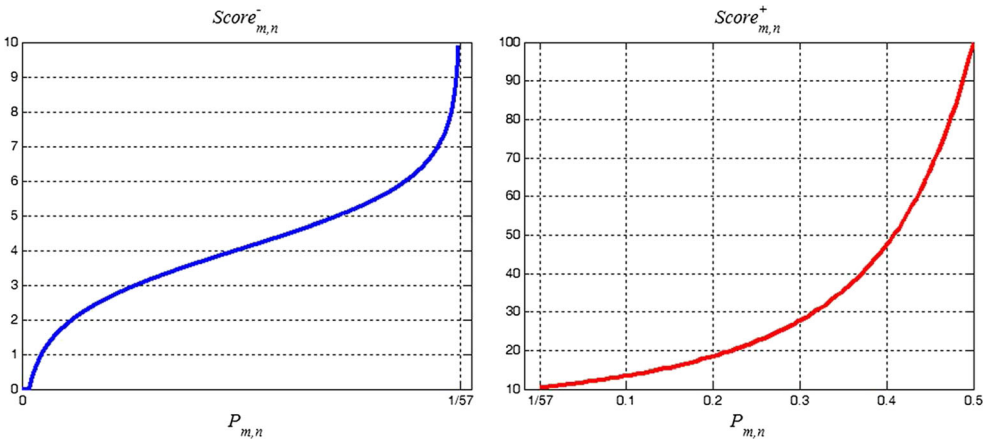

Fig. 4 The behaviors of negativity and positivity scoring functions with respect to $P_{m,n}$

the similarity of images. Accordingly, the retrieval results are several lists of images ranked by their similarities with the query image. After finding these image similarity ranking lists, the feature similarity ranking level fusion is applied to obtain the final visual similarity. In this work, the IRP algorithm [14] is used to fuse the ranking level of feature similarities. The IRP for jth image ($j=1,\cdots,N_t$) with respect to query image (q) is defined as follows:

$$IRP(q,j)=\frac{1}{\sum_{n=1}^{N}\frac{1}{rank\ \ position_n}} \tag{24}$$

where N_t is the total number of images in relevant clusters, and $n=1,\cdots,N$ is nth feature descriptor. After calculating IRP for all j and sorting it in ascending order, the k top images in the final ranking list are displayed to the user as the retrieval results.

Optimal Query-Based Relevance Feedback Algorithm

In this section, our RF approach based on finding optimal queries is presented. The main idea is based on density function estimation of relevant images and the strategy of moving toward the aggregation of estimated density function using an optimization algorithm. These optimal queries are found based on the value of optimized relevant images in the global maximums of the estimated density function.

The block diagram of proposed RF algorithm are shown in Fig. 5. In this block diagram, firstly in the retrieval process, the distance between query image and images in the relevant clusters which was found in the previous section is calculated. As mentioned in the previous section, Euclidean distance measure is used to measure the similarity of images and then IRP algorithm is used for feature similarity ranking level fusion. Consequently, the retrieval results are presented to the user and then he or she determines the relevant and irrelevant images. In the first step of our RF algorithm, the classes of irrelevant images are eliminated from our search space in dataset. With this procedure, the irrelevant classes which is very close to query image are filtered, and thus, the search space becomes smaller and smaller in each RF iteration. In the second step, optimal queries must be found using the relevant images so that more desirable results are obtained in the next retrieval process. For this purpose, the probability density function of relevant images must be estimated. Hence, in this work, we use the Parzen window technique as a nonparametric density estimator [24].

Probability Density Function Estimation

Assume that we have d-dimensional relevant samples $r_1, r_2, \ldots, r_{N_r}$ where N_r is the number of relevant images. The estimated density function at any new sample x is given by

$$p(x)=\frac{K/N_r}{V} \tag{25}$$

where K is the number of relevant samples inside a d-dimentional region R_d, and V is the volume of R_d. Consider that R_d is a hypercube centered at x, and let h be the length of the hypercube edge. Then, the total number of samples (K) which falling within the region R_d is given by

$$K=\sum_{i=1}^{N_r}\psi\left(\frac{x-r_i}{h}\right) \tag{26}$$

where $\psi(.)$ is the kernel function which is also called Parzen window and indicates whether r_i is inside the hypercube or not. If r_i falls within the hypercube, then $\psi((x-r_i)/h)$ is one and otherwise is zero. As a result, the Parzen probability density estimation for a hypercube as the kernel is given by

$$p(x)=\frac{1}{N_r}\sum_{i=1}^{N_r}\frac{1}{V}\psi\left(\frac{x-r_i}{h}\right) \tag{27}$$

By generalization, this idea and using a Gaussian window as a kernel function, we can obtain a smoother density model:

$$\psi(x-r_i)=\frac{1}{\sqrt{2\pi|\Sigma|}}e^{-\frac{1}{2}(x-r_i)^t\Sigma^{-1}(x-r_i)} \tag{28}$$

and then

$$p(x)=\frac{1}{N_r}\sum_{i=1}^{N_r}\frac{1}{\sqrt{2\pi|\Sigma|}}e^{-\frac{1}{2}(x-r_i)^t\Sigma^{-1}(x-r_i)} \tag{29}$$

where r_i is the ith relevant sample (image), x is an unknown sample, N_r is the number of relevant samples, and Σ is the covariance matrix of relevant samples.

Thus, our density model is obtained by placing a Gaussian window over each relevant data point (each image in the feature space) and then adding up the contributions over the whole dataset, and then dividing by N_r so that the density is correctly normalized. In order to make a real sense, suppose that ten relevant samples have been obtained in 2-D space. The estimated density function for this example is shown in Fig. 6.

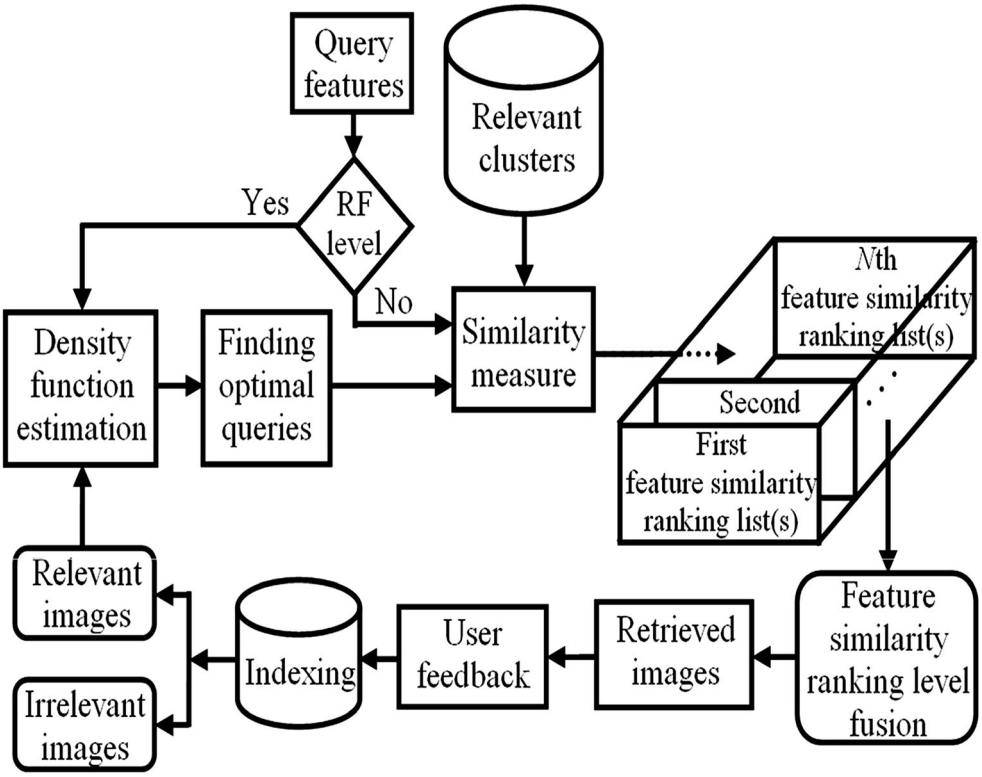

Fig. 5 Block diagram of proposed RF algorithm

According to Eq. 29, the width of the kernel windows varies according to the choice of covariance matrix Σ. Assuming that the Σ is diagonal with the same variances, i.e., $\Sigma=\sigma^2 I$, where I is the unit matrix. Since, more than 99.7 % of values drawn from a normal distribution are within three standard deviations σ away from the mean (i.e., $\mu\pm3\sigma$); hence, in this work, the covariance matrix of the windows are calculated as follows:

$$\sum = \left(\frac{(dist_{\max} + dist_{\min})/2}{6}\right)^2 I \tag{30}$$

where $dist_{\max}$ and $dist_{\min}$ are the maximum and minimum distances between relevant samples, respectively.

Optimization Algorithm

After obtaining the density function, the optimal queries should be found using an optimization algorithm. The strategy of moving relevant points toward the global maximums of estimated density function is an idea to find the optimal queries in a multidimensional space. Since, in evolutionary algorithms, GA [25] can deal with large search spaces and has less chance to get local optimal solution than other algorithms, so it is a good choice for our optimization task.

GA generates successive populations of alternative solutions and searches for solutions to optimize problem by evolving better and better solutions. For the breeding cycle, the search process creates new and hopefully fitter individuals based on an appropriate fitness function. In this paper, we use the estimated density $p(x)$ as fitness function which higher

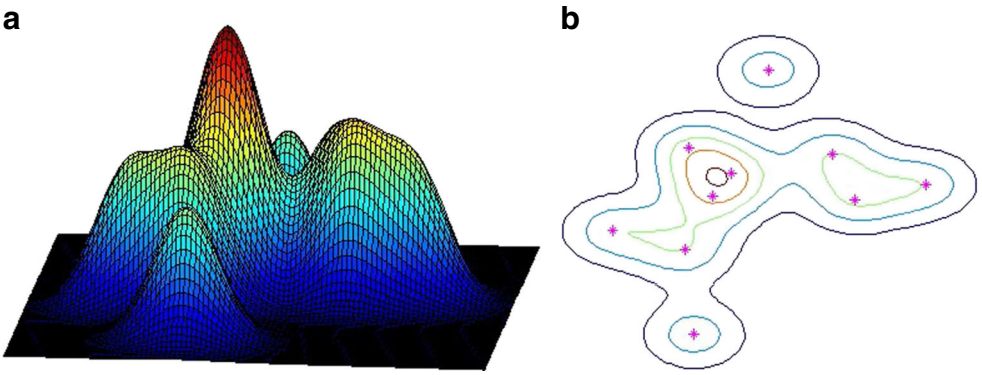

Fig. 6 **a** Estimated density function of ten relevant samples in 2-D space. **b** Contour graph of **a**; the *star markers* represent the relevant samples

values corresponding to individuals indicate the higher fitness. Accordingly, by finding the optimal coefficients for relevant images, the optimized relevant images or briefly optimized points are obtained.

After finding optimized points, with the selection of two appropriate thresholds (Th_M,Th_D), optimal queries are obtained. For this purpose, firstly, the optimized point with the largest value in estimated density function is selected as the optimal query. Then, the rest of optimized points are chosen based on two parameters Th_M and Th_D that is defined as

$$Th_M = \alpha \times p_{\max} \tag{31}$$

$$Th_D = \beta \times \sigma \tag{32}$$

where α and β are two arbitrary coefficients that their values can influence the number of optimal queries. The $p_{\max}$ is the largest value in estimated density function, and σ is the standard deviation of Parzen window which is obtained according to the definition of $\sum=\sigma^2 I$ in Eq. 30. Accordingly, if the value of other optimized points are greater than Th_M and the distance between them is also greater than Th_D, these points are also selected as the optimal queries. The reasonable ranges for α and β can be in [0.6–1] and [1–3] intervals, respectively. For example, in Fig. 6, two optimal queries are obtained considering α=0.75 and β=1.

Experimental Results

In this study, to evaluate the performance of our proposed framework, we use ImageCLEF 2005 database [41] consisting of 10,000 medical X-ray images of 57 classes. An image sample of each class has been shown in Fig. 7, and the details of image properties per each category are included in Table 1. This database has been partitioned into two datasets consisting of 9,000 images as training dataset and 1,000 images as test dataset. The number of images in each class of training database is given in Table 1.

Evaluation Measure

To evaluate the performance of our retrieval framework, two parameter precision and recall have been introduced [42]. Let A be the set of relevant images and B be the set of retrieved images. Then, recall and precision are defined as

$$\text{Pr}ecision = P(A|B) = \frac{\text{number of relevant retrieved images}}{\text{number of retrieved images}} \tag{33}$$

$$\text{Re}call = P(B|A) = \frac{\text{number of relevant retrieved images}}{\text{number of relevant images in database}} \tag{34}$$

Moreover, Pr(Re=0.5), the precision value at the point where the recall value is 0.5; Pr(Re=Pr), the precision value where the recall and precision values are equal; Pr−Re area, the area under the Pr−Re curve; and Pr(N_R), the precision after N_R images are retrieved are used in the evaluation of the previous presented systems [43].

Feature Space

In the training phase, by applying the mentioned feature extraction methods, each image is represented by N feature vectors with D_n elements (D_n,n=1,⋯,N). In this work, we have used N=7 including MGFD with D_1=49, centroid distance FDs with D_2=128, CLBC_S with D_3=9, CLBC_M with D_4=9, uniform LBP with D_5=10, complex coordinate distance FDs with D_6=256, and angular function FDs with D_7=128. As mentioned in the section "Proposed Content-Based Image Retrieval Framework," to obtain uncorrelated feature vectors and reduce the space and time complexities for distribution function estimation of feature descriptors, we used PCA technique. In this work, the LBP-based features are uncorrelated. Therefore, PCA is not applied on it. Therefore, PCA are applied only on 1-D FDs and 2-D MGFD so that a well-conditioned covariance matrix is obtained. As a result, the dimensional of D_1, D_2, D_6, and D_7 are reduced to 9. This means that these features are projected onto a space having dimensionality 9 where in this case the total transferred energy by these eigen-vectors are 95, 98.5, 96, and 99.9 %, respectively.

Retrieval Results

As previously mentioned, after extracting the feature vectors and applying dimensional reduction algorithm, Gaussian model of total classes was found using GMM for each feature descriptor based on the training database. The number of components in our mixture model is considered equal to the number of classes (i.e., M=57), and each component represents the distribution of one class. The mixing coefficients are considered to 1/M.

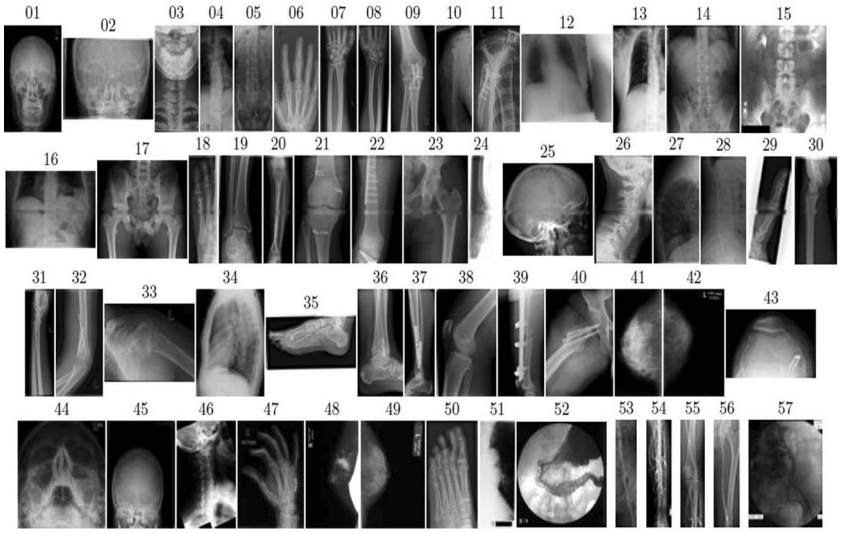

Fig. 7 Example images of 57-class database [41]

Table 1 X-ray image classes

Class	Number	Anatomic	Direction	Class	Number	Anatomic	Direction
01	336	Cranium, musculoskeletal system	Coronal	30	59	Radiocarpal joint, musculoskeletal system	Sagittal
02	32	Facial cranium, musculoskeletal system	Coronal	31	60	Hand, forearm, musculoskeletal system	Sagittal
03	215	Cervical spine, musculoskeletal system	Coronal	32	78	Elbow, musculoskeletal system	Sagittal
04	102	Thoracic spine, musculoskeletal system	Coronal	33	62	Shoulder, musculoskeletal system	Sagittal
05	225	Lumbar spine, musculoskeletal system	Coronal	34	880	Chest, unspecified	Sagittal
06	576	Hand, musculoskeletal system	Coronal	35	18	Foot, musculoskeletal system	Sagittal
07	77	Radiocarpal joint, musculoskeletal system	Coronal	36	94	Ankle joint, musculoskeletal system	Sagittal
08	48	Hand, forearm, musculoskeletal system	Coronal	37	22	Lower leg, musculoskeletal system	Sagittal
09	69	Elbow, musculoskeletal system	Coronal	38	116	Knee, musculoskeletal system	Sagittal
10	32	Upper arm, musculoskeletal system	Coronal	39	38	Upper leg, musculoskeletal system	Sagittal
11	108	Shoulder, musculoskeletal system	Coronal	40	51	Hip, musculoskeletal system	Sagittal
12	2,563	Chest, unspecified	Coronal	41	65	Right breast, reproductive system	Axial
13	93	Bones, musculoskeletal system	Coronal	42	74	Left breast, reproductive system	Axial
14	152	Abdomen, gastrointestinal system	Coronal	43	98	Knee, musculoskeletal system	Axial
15	15	Abdomen, uropoietic system	Coronal	44	193	Facial cranium, musculoskeletal system	other orientation
16	23	Upper abdomen, gastrointestinal system	Coronal	45	35	Neurocranium, musculoskeletal system	other orientation
17	217	Pelvis, musculoskeletal system	Coronal	46	30	Cervical spine, musculoskeletal system	other orientation
18	205	Foot, musculoskeletal system	Coronal	47	147	Hand, musculoskeletal system	other orientation
19	137	Ankle joint, musculoskeletal system	Coronal	48	79	Right breast, reproductive system	other orientation
20	31	Lower leg, musculoskeletal system	Coronal	49	78	Left breast, reproductive system	other orientation
21	194	Knee, musculoskeletal system	Coronal	50	91	Foot, musculoskeletal system	other orientation
22	48	Upper leg, musculoskeletal system	Coronal	51	9	Hilum, respiratory system	Coronal
23	79	Hip, musculoskeletal system	Coronal	52	9	Upper abdomen, gastrointestinal system	Coronal
24	17	Facial cranium, musculoskeletal system	Sagittal	53	15	Pelvis, cardiovascular system	Coronal
25	284	Neurocranium, musculoskeletal system	Sagittal	54	46	Lower leg, cardiovascular system	Coronal
26	170	Cervical spine, musculoskeletal system	Sagittal	55	10	Knee, cardiovascular system	Coronal
27	109	Thoracic spine, musculoskeletal system	Sagittal	56	15	Upper leg, cardiovascular system	Coronal
28	228	Lumbar spine, musculoskeletal system	Sagittal	57	57	Pelvis, cardiovascular system	Coronal
29	86	Hand, musculoskeletal system	Sagittal				

In the test phase, the posterior probabilities of query belonging to each component given that the corresponding feature vector or briefly dependency probabilities are calculated. Therefore, we have 7 dependency probability vectors with 57 elements (according to 7 feature descriptors and 57 classes) which the sum of elements of each vector is equal to 1. Afterwards, as mentioned in the section "Proposed Content-Based Image Retrieval Framework," using our proposed fusion scheme, the scores of query belonging to each class are calculated and then using an appropriate score threshold, the most relevant clusters are identified for the given query. As shown in Table 1, the number of images in each class of training database is not uniformly distributed; nevertheless, we have considered an equal priori probability for all classes (i.e., P_m=1/57, m=1,⋯,M). Figure 8 shows the performance of our proposed classification algorithm for different values of the score threshold.

Figure 8a shows the accuracy of our proposed fusion-based classification algorithm based on different score thresholds while Fig. 8b shows the average number of clusters based on different score thresholds. According to these graphs, for the score threshold values of 2.5, 5, 10, and 15, the accuracy of 99.8, 99.6, 99.1, and 98.5 % are obtained, respectively. Also, for these score thresholds, 33.6, 27.9, 26.1, and 16.8 clusters are obtained by averaging on the test dataset. For more details, when the score threshold is equal to 5, the search space is reduced to 27 clusters, and only 4 images from 1,000 images of test database has been incorrectly clustered that have been shown in Fig. 9.

As a result, according to Figs. 8 and 9, by increasing the score threshold, two issues are remarkable: First, the search space becomes smaller and consequently, it is more efficient in the image retrieval; second, the queries must be selected more carefully and the extent possible must be had more quality. Thus, the choice of score threshold can be performed as a compromise between computational complexity and performance of system. On the other hand, there is no significant changes in the results if the score threshold is more than 100 (Fig. 8b). Hence, it is not necessary to choose a score threshold much larger than 100.

Figure 10 shows precision-recall curves for different values of the score threshold. In Fig. 10a, the score threshold value has been considered equal to zero and in this case the search process is carried out in the whole database without any filtering. In Fig. 10b–d, the score threshold values have been set to 15, 30, and 60, respectively.

In the graphs of Fig. 10, by increasing the score threshold value, the values of precision and recall become greater and on the other hand, the maximum values of precision and recall are reduced. In other words, firstly by increasing the score threshold value, the search space becomes smaller and secondly, more images from test database are incorrectly clustered. For example, when score threshold value is set to 60, the maximum values of precision and recall are reduced to 0.91. For the score threshold value of 60, the search space is averagely reduced to three clusters but the accuracy of 91 % is obtained. This means that 9 % (90 images) from 1,000 images of test dataset are incorrectly clustered and therefore their precision and recall are zero. The significant point in Fig. 10 is the improvement of precision and recall values with applying the RF iterations. In other words, in each RF iteration, the irrelevant classes are filtered and thus the search space becomes smaller and smaller. This condition grows more rapidly by increasing the score threshold value.

For example, Table 2 presents the process of reducing the search space and finding the most relevant clusters to a given query. In this example, an arbitrary query (class 38) has been given to the system and the score threshold value has been set to 30. After calculating dependency probabilities for each of the seven feature descriptors and scores for all of the classes, the search space is reduced to eight classes. The relevant classes are class 05, class 07, class 08, class 09, class 11, class 19, class 21, and class 38, respectively. In Table 2, the scores of relevant classes are greater than the score threshold of 30 which these scores are obtained based on fusion of the dependency probabilities $P_{m',n}$ (m'=5,7,8,9,11,19,21,38 and n=1,⋯,7) of 7 feature descriptors. The largest dependency probability is related to $P_{11,7}$ where is equal to 0.9781, and it means that conditional probability of query belonging to class 11 using the seventh feature descriptor (angular function FDs) is

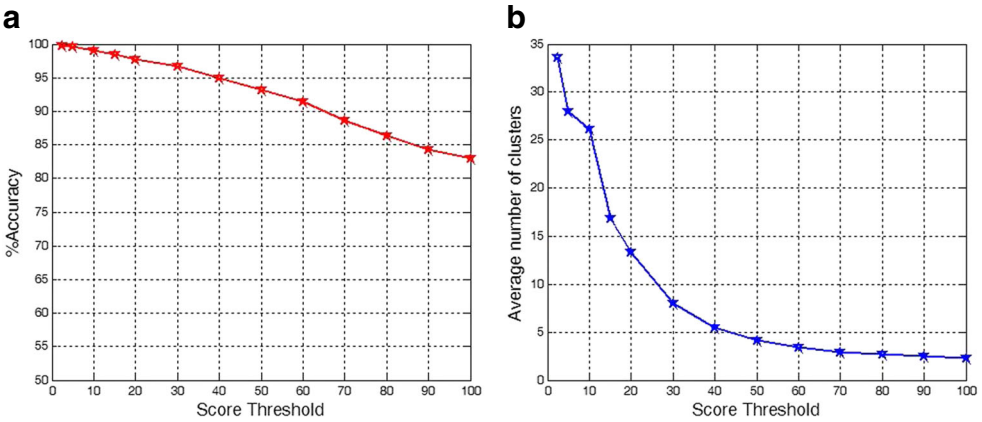

Fig. 8 Performance of our proposed classification algorithm for different values of the score threshold. **a** The accuracy of proposed fusion-based classification based on different score thresholds. **b** The average number of clusters based on different score thresholds

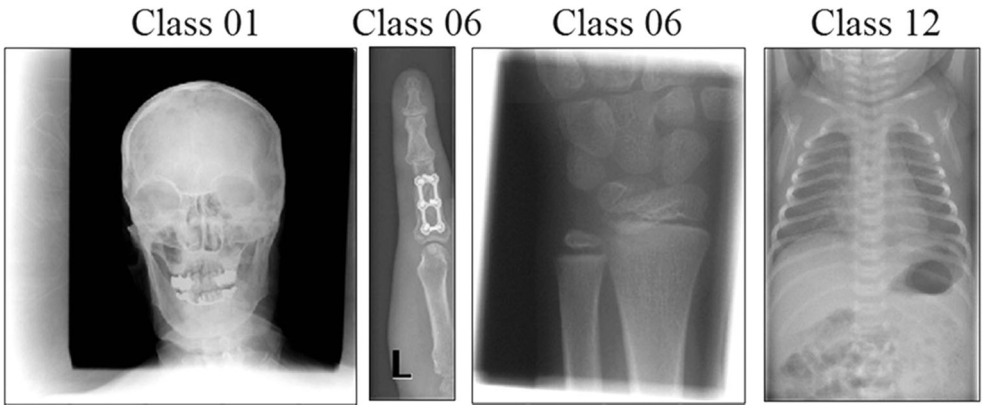

Fig. 9 Four images of test database that have been incorrectly clustered when score threshold is set to 5

0.9781 which leads to a very high score. Figure 11 shows the retrieval results of 20 images for this query.

According to Figs. 7 and 11 and Table 2, it is clear that class 11 has the highest score, but no images have been retrieved from this class. This is due to the fact that in our feature similarity ranking level fusion algorithm, the ranks are not measured using only one feature descriptor. Therefore, as shown in Fig. 11, the most images have been retrieved from the class 21 since this class has more values of posterior probabilities compared with other classes. In this example, only four relevant images of the class 38 have been retrieved (ranks 1, 2, 8, and 12). Note, according to Table 2, if the value of score threshold is set to 50, the search space will be reduced to three classes and more favorable results will be obtained. The whole retrieval process and representation of results takes only 3.29 s.

However, after determining the relevant and irrelevant images, our RF algorithm is employed. In the example of Fig. 11, we have the four relevant images which their density functions must be estimated. As mentioned in section "Optimal Query-Based Relevance Feedback Algorithm," to find the optimal queries in the multidimensional space, GA is applied. For this purpose, we apply the GA to find the optimal weights for relevant images and each feature descriptor to obtain the optimized relevant images. In this work, we use $N=7$ fitness functions based on the estimated density function of relevant images

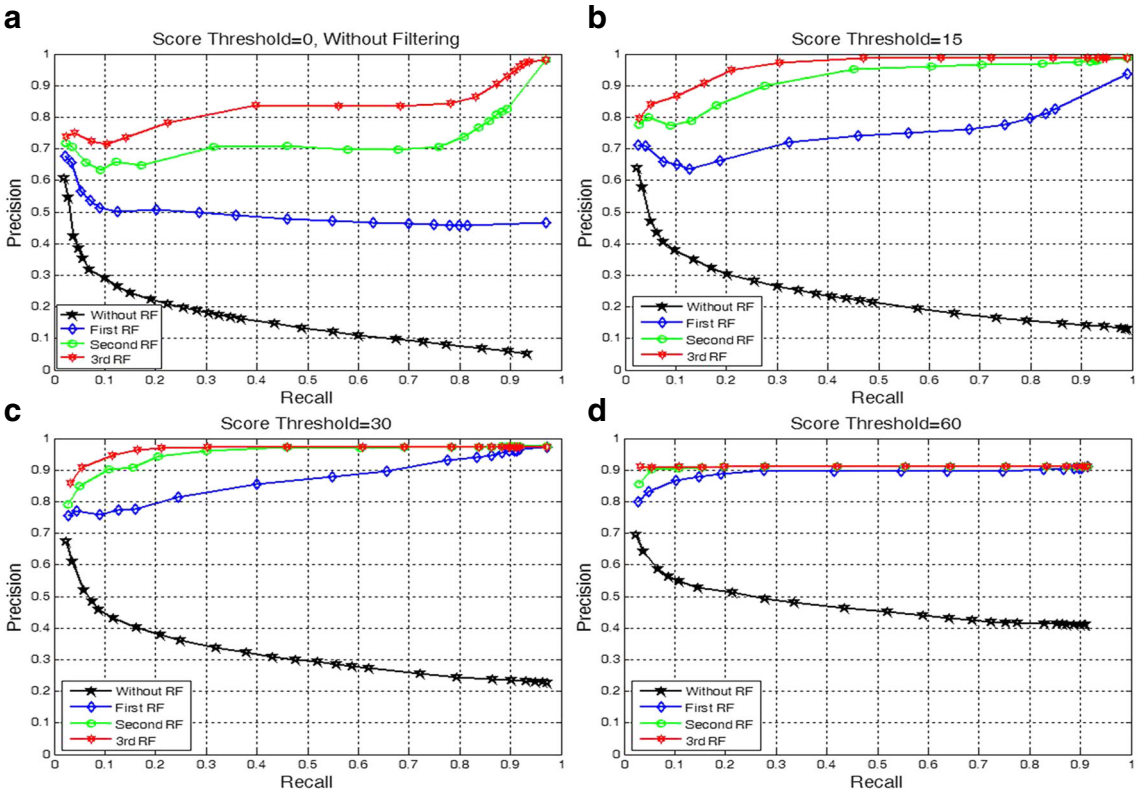

Fig. 10 Precision-recall curves for different values of the score threshold. **a** The score threshold value has been considered equal to zero; the search process is performed without any filtering. **b–d** The score threshold values have been set to 15, 30, and 60, respectively

Table 2 An example of classification process: the score threshold value has been set to 30 and so the search space has been reduced to eight classes (m'= 5,7,8,9,11,19,21,38)

Query	$P_{m',n}$	Class 05	Class 07	Class 08	Class 09	Class 11	Class 19	Class 21	Class 38
	$P_{m',1}$	0.0234	0.0035	0.0277	0.2287	2.2538×10^{-05}	0.0554	0.1865	0.2423
	$P_{m',2}$	0.2480	0.1429	0.2361	1.4394×10^{-17}	1.7977×10^{-05}	0.0637	1.2285×10^{-05}	0.0046
	$P_{m',3}$	9.4316×10^{-14}	9.1279×10^{-04}	6.5456×10^{-07}	0.0139	3.8557×10^{-04}	8.3302×10^{-04}	07728	01789
	$P_{m',4}$	5.3768×10^{-25}	0.0070	4.6307×10^{-16}	0.0339	0.0019	0.0068	0.3748	0.3738
	$P_{m',5}$	4.0726×10^{-16}	0.0061	8.2335×10^{-06}	0.0060	0.0037	1.8116×10^{-04}	0.4987	0.4491
	$P_{m',6}$	0.1835	0.0677	3.5447×10^{-04}	6.6935×10^{-06}	1.8029×10^{-04}	0.0703	9.6310×10^{-04}	0.0869
	$P_{m',7}$	4.0430×10^{-04}	1.4814×10^{-46}	2.1363×10^{-137}	1.3172×10^{-35}	0.9781	3.9476×10^{-19}	3.7652×10^{-04}	6.4645×10^{-04}
	Score	48.8433	31.1978	31.1240	34.4463	4.3348×10^{45}	37.0523	2.5404×10^{04}	156.3732

of each feature descriptor. In our application, elite count and crossover fraction parameters are set to 2 and 0.8, respectively. The nonuniform mutation is applied to Gaussian distribution centered on zero that scale and shrink parameters are set to 1 while uniform mutation is used with mutation rate 0.15. The initial population is randomly generated in range [0–1] with 30 chromosomes. The length of chromosomes is determined based on the dimension of feature descriptors. The GA will be stop if there is no improvement in the fitness function for 40 consecutive generations or the sum of deviations among individuals becomes smaller than 1e−6 or ultimately the maximum number of iterations (i.e., 100 iterations) is carried out.

After finding the optimized relevant images, the largest peak in the estimated density function is selected as the optimal query and the others are chosen based on two thresholds Th_M and Th_D. In this work, α and β in Eqs. 31 and 32 are considered equal to 0.75 and 1, respectively. So, we have $Th_M=0.75\times p_{\max}$ and $Th_D=\sigma$. This means that if other candidate points are greater than 75 % of the largest peak and the distance between them also is greater than Th_D, these maximums are also selected as the optimal queries.

After finding the optimal queries, the similarity between the optimal queries and images of the relevant classes is measured corresponding to each feature descriptor. Then, the IRP algorithm is used for the feature similarity ranking level fusion and ultimately k top images in the final ranking list are displayed to the user as the retrieval results. This process is

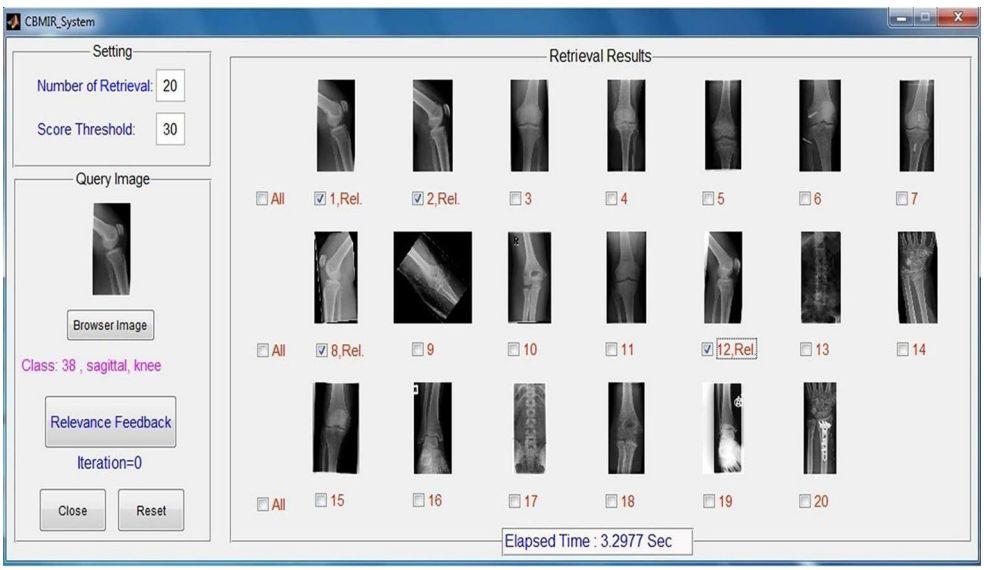

Fig. 11 The retrieval results of 20 images for a given query; the value of score threshold has been set to 30, and only four relevant images (*Rel.*) have been retrieved (ranks 1, 2, 8, and 12) which are specified by the user, whole retrieval process, and representation of results have been taken only 3.29 s

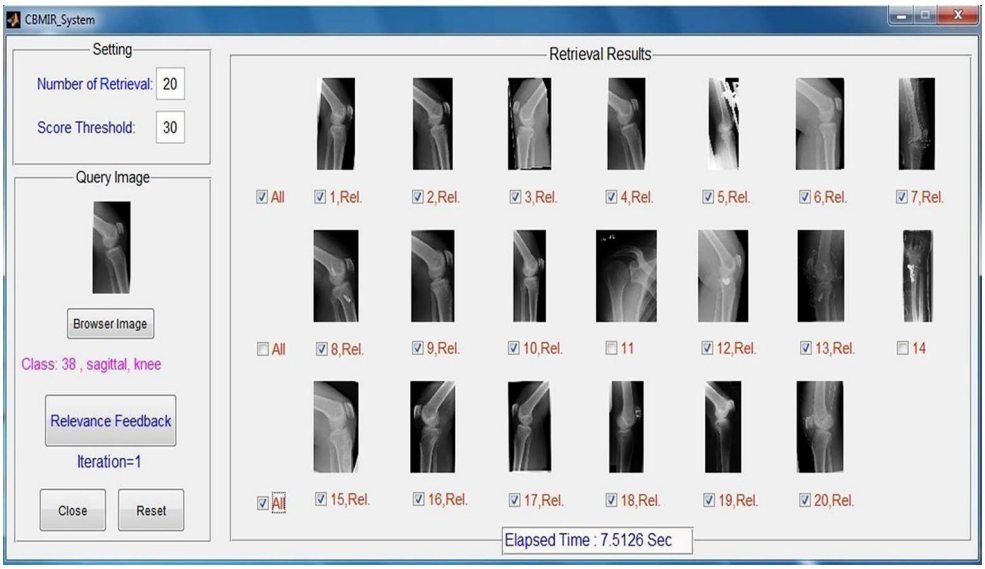

Fig. 12 Retrieval results in the first iteration of RF

repeated until the user is satisfied with the result. Figures 12 and 13 show the retrieval results after applying of the first and second iterations of RF, respectively. The RF processing time in the first iteration is 7.51 s while this time is 21.25 s for the second iteration. In Fig. 12, only two irrelevant images have been retrieved in the first RF (ranks 11 and 14), and in the second iteration of RF, there is no any irrelevant images in the retrieval results (Fig. 13).

Discussion and Comparison with Other Works

The retrieval results of our proposed framework were obtained based on dataset of ImageCLEF 2005 consisting of 10,000 medical X-ray images of 57 semantic classes (9,000 images as training dataset and 1,000 images as test dataset). The performance of this framework was evaluated based on four criteria, Pr(20), Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area where the last three criteria were extracted from precision-recall curve. Table 3 presents the retrieval results of our proposed frameworks based on these four criteria for different values of the score threshold. In Table 3, the best retrieval results in our framework are 0.95, 0.97, 0.97, and 0.96 for Pr(20), Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively.

To evaluate results of our work, we obtain a comparison between our proposed framework and other presented retrieval techniques in the literature. The presented classification-based image retrieval framework in [9] had been evaluated on ImageCLEFmed 2006 database consisting of a diverse medical image collection of 11,000 images of 116 categories (10,000 images as training dataset and 1,000 images as test

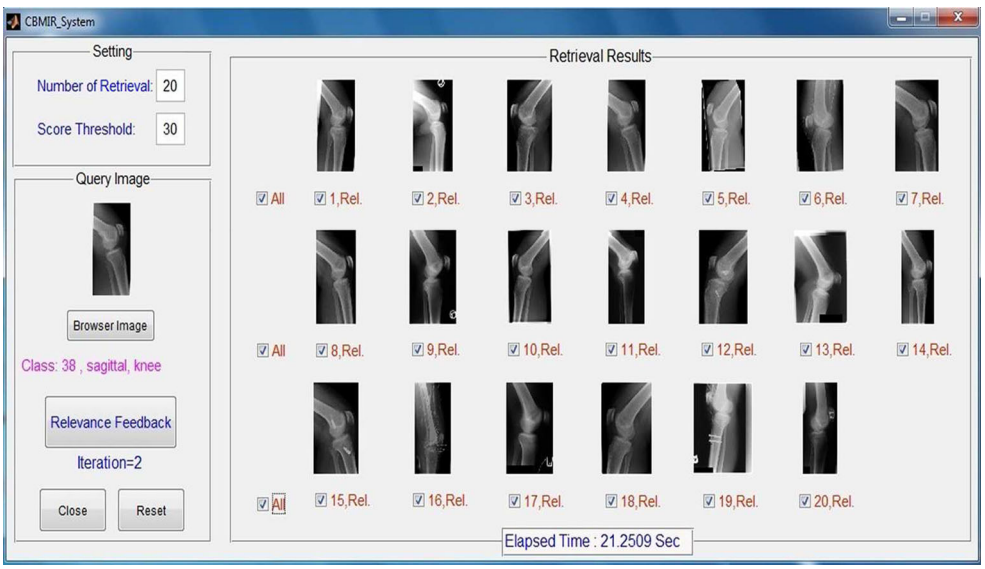

Fig. 13 Retrieval results in the second iteration of RF

Table 3 The results of proposed framework for different values of the score threshold

RF iteration	Score threshold=0				Score threshold=15			
	Pr(20)	Pr(Re=Pr)	Pr(Re=0.5)	Pr-Re area	Pr(20)	Pr(Re=Pr)	Pr(Re=0.5)	Pr-Re area
Without RF	0.35	0.21	0.13	0.17	0.41	0.27	0.21	0.23
First RF	0.51	0.46	0.47	0.47	0.64	0.79	0.74	0.75
Second RF	0.66	0.95	0.71	0.71	0.84	0.97	0.96	0.92
Third RF	0.73	0.96	0.84	0.80	0.95	0.97	0.97	0.96
	Score threshold=30				Score threshold=60			
Without RF	0.46	0.32	0.30	0.32	0.55	0.46	0.45	0.44
First RF	0.78	0.97	0.87	0.84	0.89	0.91	0.89	0.79
Second RF	0.94	0.97	0.97	0.94	0.91	0.91	0.91	0.80
Third RF	0.95	0.97	0.97	0.96	0.91	0.91	0.91	0.81

dataset). The best reported retrieval result was 0.63, 0.64, and 0.58 for Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively (these values have been approximately calculated from precision-recall curve in [9]). The best reported retrieval result in [10] on a dataset of 1,501 radiological images of 17 classes was 0.62, 0.67, and 0.66 for Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively (these values have been approximately calculated from precision-recall curve in [10]). The best performance of the presented CBIR system in [20] based on a database consisting of 5,000 images of 20 classes was 0.68, 0.82, and 0.72 for Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively (these values have been approximately calculated from precision-recall curve in [20]). The best performance of the presented iterative classification-based image retrieval framework in [21] based on dataset of ImageCLEF 2005 consisting of 10,000 medical X-ray images of 57 classes (9,000 images as training dataset and 1,000 images as test dataset) was 0.915, 0.88, 0.71, and 0.67 for Pr(20), Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively. The best performance of the proposed CBIR system in [22] based on a database consisting of 5,000 images of 30 categories was 0.52, 0.53, and 0.50 for Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively (these values have been approximately calculated from precision-recall curve in [22]). The best performance of the proposed CBIR system in [23] based on a database consisting of 2,785 images of 15 categories was 0.77, 0.78, and 0.76 for Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively (these values have been approximately calculated from precision-recall curve in [23]). Table 4 provides a comparison between our retrieval results and other works.

Table 4 Comparison between our proposed RF algorithm and the previous works

Approach	Retrieval measures			
	Pr(20)	Pr(Re=Pr)	Pr(Re=0.5)	Pr-Re area
Rahman et al. [9]	–	0.63	0.64	0.58
Greenspan and Pinhas [10]	–	0.62	0.67	0.66
Rahman et al. [20]	–	0.68	0.82	0.72
Pourghassem and Daneshvar [21]	0.915	0.88	0.71	0.67
Rahman et al. [22]	–	0.52	0.53	0.50
Yang et al. [23]	–	0.77	0.78	0.76
Proposed algorithm (third RF)	0.95	0.97	0.97	0.96

Conclusion

In this paper, we proposed a novel content-based medical image retrieval framework based on a score fusion algorithm for query image classification and a new query expansion technique in the RF level. In the classification phase, the scores of query belonging to each class were calculated using distribution function-based dependency probabilities of feature descriptors, and then the most relevant clusters were identified based on an appropriate score threshold value for each given query. Since each feature descriptor describes various aspects of images and has distinctive distributions, we calculated dependency probabilities for each feature descriptor and then using fusion of obtained scores from these dependency probabilities, the total scores of query belonging to each class were calculated. In the retrieval phase, we applied feature similarity ranking level fusion algorithm to obtain the final visual similarities. In the RF level, a novel approach was proposed to find the optimal queries based on the relevant images. Our idea in RF is based on density function estimation of relevant images using Parzen technique and then the strategy of moving relevant points toward the global maximums of estimated density function in a multidimensional space using GA as an optimization method. Our proposed CBMIR framework was evaluated on a database

consisting of 10,000 medical X-ray images of 57 semantic classes. The best retrieval results in our framework were acquired 0.95, 0.97, 0.97, and 0.96 for Pr(20), Pr(Re=Pr), Pr(Re=0.5), and Pr−Re area, respectively where the effectiveness of our proposed framework was demonstrated compared with other presented retrieval algorithms in the literature.

Acknowledgments The authors would like to thank the IRMA Group, Aachen, Germany, for making the database available for the experiments.

References

1. Tamura H, Yokoya N: Image database systems: a survey. Pattern Recognit 17(1):29–43, 1984
2. Chang SK, Hsu A: Image information systems: where do we go from here? IEEE Trans Knowl Data Eng 5(5):431–442, 1992
3. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R: Content-based image retrieval at the end of the early years, IEEE Trans. Pattern Anal Mach Intell 22(12):1349–1380, 2000
4. Rui Y, Huang TS, Chang S-F: Image retrieval: current techniques, promising directions, and open issues. J Vis Commun Image Represent 10(4):39–62, 1999
5. Jafari Fesharaki N, Pourghassem H: Medical x-ray image hierarchical classification using a merging and splitting scheme in feature space. J Med Signals Sensors 3(3):150–163, 2013
6. Liu Y, Zhanga D, Lua G, Ma WY: A survey of content-based image retrieval with high-level semantics. Pattern Recognit 40(1):262–282, 2007
7. Pourghassem H, Ghassemian H: Content-based medical image classification using a new hierarchical merging scheme. J Comput Med Imaging Graph 32(8):651–661, 2008
8. N. Jafari Fesharaki, H. Pourghassem: Medical X-ray Images Classification Based on Shape Features and Bayesian Rule, Fourth International Conference on Computational Intelligence and Communication Networks (CICN2012): Mathura, Uttar Pradesh, India 3–5 Nov. 2012, pp 369-373
9. Rahman MM, Desai BC, Bhattacharya P: Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion. J Comput Med Imaging Graph 32(2):95–108, 2008
10. Greenspan H, Pinhas AT: Medical image categorization and retrieval for PACS using the GMM-KL framework. IEEE Trans Inf Tech Bio 11(2):190–202, 2007
11. Marakakis A, Siolas G, Galatsanos N, Likas A, Stafylopatis A: Relevance feedback approach for image retrieval combining support vector machines and adapted Gaussian mixture models. IET Image Process 5(6):531–540, 2011
12. Xie Z, Wang S, Hu D: New insight at level set & Gaussian mixture model for natural image segmentation. SIViP 7(3):521–536, 2013
13. Chatzichristofis SA, Zagoris K, Boutalis Y, Arampatzis A: A Fuzzy Rank-Based Late Fusion Method for Image Retrieval. 18th International Conference, MMM(2012). Springer Berlin Heidelberg, Klagenfurt, Austria, 2012, pp 463–472
14. M. Jović, Y. Hatakeyama, F. Dong, K. Hirota: Image Retrieval Based on Similarity Score Fusion from Feature Similarity Ranking Lists, Third International Conference, FSKD 2006, Xi'an: China, Sep. 2006, pp 461–470
15. Su JH, Huang WH, Yu PS, Tseng VS: Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. IEEE Trans Knowl Data Eng 23(3):360–372, 2011
16. J.J. Rocchio: Relevance Feedback in Information Retrieval: The SMART Retrieval System—Experiments in Automatic Document Processing, 1971, pp 313–323
17. Y. Rui, T. S. Huang, and S. Mehrotra: Content-Based Image Retrieval with Relevance Feedback in MARS: Proc. IEEE Int'l Conf. Image Processing, Oct. 1997, pp 815–818
18. Rui Y, Huang TS, Ortega M, Mehrotra S: Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Trans Circ Syst Video Techno 8(5):644–655, 1998
19. K. Porkaew, K. Chakrabarti, and S. Mehrotra: Query Refinement for Multimedia Similarity Retrieval in MARS: Proc. ACM Int'l Multimedia Conf. (ACMMM), 1999, pp 235–238
20. Rahman MM, Bhattacharya P, Desai BC: A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Trans Inf Tech Bio 11(1):58–69, 2007
21. Pourghassem H, Daneshvar S: A framework for medical image retrieval using merging-based classification with dependency probability-based relevance feedback. Turk J Electr Eng Comput Sci 21(3):882–896, 2013
22. Rahman MM, Antani SK, Thoma GR: A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback. IEEE Trans Inf Technol Biomed 15(4):640–646, 2011
23. Yang L, Jin R, Mummert L, Sukthankar R, Goode A, Zheng B, Hoi SCH, Satyanarayanan M: A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. IEEE Trans Pattern Anal Mach Intell 32(1):30–44, 2010
24. Duda RO, Hart PE, Stork DG: Pattern Classification, 2nd edition. John Wiley & Sons, New York, NY, 2001
25. Nurmohamadi M, Pourghassem H: Clavulanic acid production estimation based on color and structural features of streptomyces clavuligerus bacteria using self-organizing map and genetic algorithm. Comput Methods Programs Biomed 114(3):337–348, 2014
26. M. Behnam, H. Pourghassem: Feature Descriptor Optimization in Medical Image Retrieval Based on Genetic Algorithm: 20th Iranian Conference on Biomedical Engineering (ICBME2013), Tehran, Iran: 18–20 Dec, 2013, pp 280–285
27. da Silva SF, Ribeiro MX, Neto JB, TrainaJr C, Traina AJM: Improving the ranking quality of medical image retrieval using a genetic feature selection method. Decis Support Syst 51(4):810–820, 2011
28. Steji'c Z, Takama Y, Hirota K: Genetic algorithms for a family of image similarity models incorporated in the relevance feedback mechanism. Appl Soft Comput 2(4):306–327, 2003
29. Lai CC, Chen YC: A user-oriented image retrieval system based on interactive genetic algorithm. IEEE Trans Instrum Meas 60(10): 3318–3325, 2011
30. Ojala T, Pietikäinen M, Mäenpää T: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987, 2002
31. Zhao Y, Huang DS, Jia W: Completed local binary count for rotation invariant texture classification. IEEE Trans Image Process 21(10): 4492–4497, 2012
32. Guo ZH, Zhang L, Zhang D: A completed modeling of local binary pattern operator for texture classification. IEEE Trans Image Process 19(6):1657–1663, 2010
33. Persoon E, Fu K: Shape discrimination using fourier descriptors. IEEE Trans Pattern Anal Mach Intell 8(3):388–397, 1977
34. Zahn CT, Roskies RZ: Fourier descriptors for plane closed curves. IEEE Trans Comput 21(3):269–281, 1972
35. Tahmasebi A, Pourghassem H: A novel intra-class distance-based signature identification algorithm using weighted gabor features and dynamic characteristics. Arab J Sci Eng 38(11):3019–3029, 2013
36. Zhang DS, Lu G: Generic fourier descriptor for shape-based image retrieval. IEEE Int Conf Multimed Expo 1:425–428, 2002

37. D. S. Zhang and G. Lu: Enhanced generic Fourier descriptors for object-based image retrieval: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2002), 2002, pp 3668_3671
38. Shahbeig S, Pourghassem H: A fast and automatic algorithm for optic disc extraction in retinal images using PCA-based pre-processing and curvelet transform. J Opt Soc Am A 30(1):13–21, 2013
39. Pourghassem H: A novel material detection algorithm based on 2D GMM-based power density function and image detail addition scheme in dual energy x-ray images. J X-Ray Sci Technol (IOS press) 20(2):213–228, 2012
40. Pourghassem H: A relevance feedback approach based on modification of similarity measure using particle swarm optimization in a medical x-ray image retrieval system. Majlesi J Electrical Eng 4(2): 9–17, 2010
41. P. Clough, H. Muller, T. Deselaers, M. Grubinger, T. M. Lehmann, J. Jensen, W. Hersh: The CLEF 2005 Cross–Language Image Retrieval Track: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria: 21–23, 2006, September, 2005, pp. 535–557
42. Smith JR, Chang S: Tools and techniques for color image retrieval. Storage Retrieval Image Video Databases IV 2670:1–12, 1996
43. Deselaers T, Keysers D, Ney H: Classification error rate for quantitative evaluation of content-based image retrieval systems. Proc 17th International Conf Pattern Recognit 2:505–508, 2004