

# A medical image retrieval scheme with relevance feedback through a medical social network

Mouhamed Gaith Ayadi<sup>1</sup> · Riadh Bouslimi<sup>1</sup> · Jalel Akaichi<sup>1</sup>

Received: 5 December 2015 / Revised: 12 July 2016 / Accepted: 17 July 2016 / Published online: 29 July 2016  
© Springer-Verlag Wien 2016

**Abstract** Medical social networking sites enabled multimedia content sharing in large volumes, by allowing physicians and patients to upload their medical images. Moreover, it is necessary to employ new techniques in order to effectively handle and benefit from them. This huge volume of images needs to formulate new types of queries that pose complex questions to medical social network databases. Content-based image retrieval (CBIR) stills an active and efficient research topic to manipulate medical images. In order to palliate this situation, we propose in this paper the integration of a content-based medical image retrieval method through a medical social network, based on an efficient fusion of low-level visual image features (color, shape and texture features), which offers an efficient and flexible precision. A clear application of our CBIR system consists of providing stored images that are visually similar to a new (undiagnosed) one, allowing specialist and patients to check past examination diagnoses from comments and other physicians' annotations, and to establish, therefore, a new diagnostic or to prepare a new report of an image's examination. To scale up the performance of the integrated CBIR system, we implement a relevance feedback method. It is an effective method to bridge the semantic gap between low-level visual features and high-level semantic meanings.

Experiments show that the proposed medical image retrieval scheme achieves better performance and accuracy in retrieving images. However, we need also to verify whether our approach is considered by the specialists as a potential aid in a real environment. To do so, we evaluate our methodology's impact in the user's decision, inquiring the specialists about the degree of confidence in the retrieval system. By analyzing the obtained results, we can argue that the proposed methodology presented a high acceptance regarding the specialists' interests in the clinical practice domain and can improve the decision-making process during analysis.

**Keywords** Medical social network · Content-based medical image retrieval (CBMIR) · Feature extraction · Relevance feedback

## 1 Introduction

Online social networking is attracting more and more people in today's Internet, where users can share and consume all kinds of multimedia contents (Zhi et al. 2013). Like most people, health care professionals use mainstream social media networks to connect with friends and family. But, almost one-third of them also join social networks focused exclusively on health care, where health care professionals can collaborate and share resources online, and patients can access more than information (Doganay 2014). According to Doganay (2014), patient-focused networks, often built around a particular condition or disease, give individuals and their families' supportive communities where they receive comfort, insights, and potential leads on new treatments. The data mining practices of sites like Facebook and Twitter make some patients

✉ Mouhamed Gaith Ayadi  
mouhamed.gaith.ayadi@gmail.com

Riadh Bouslimi  
bouslimi.riadh@gmail.com

Jalel Akaichi  
jalel.akeichi@isg.rnu.tn

<sup>1</sup> Department of Computer Sciences, ISG, BESTMOD, Tunis, Tunisia

and providers leery of posting questions or comments, while many health care organizations use Facebook, Twitter, LinkedIn, Instagram, and other social tools to communicate with constituents, individuals often worry about posting information in the wrong place. By sharing data on specialized sites, health care, professionals and other users can feel safer about expressing their thoughts (Doganay 2014). Franklin and Greene (2007) consider that participation in the health care management can render patients longer health conscious. According to Grenier (2003), the main objective behind medical networks is to foster collaboration between medical actors and to place the patient at the heart of the health system. In reality, the fact of making important decisions, related to medical images, individually, can lead the physician to commit errors leading to malpractices and consequently to unexpected damages. This fact is justified by a study done by The Institute of Medicine of the National Sciences Academy<sup>1</sup> (IMNAS) in USA. This institute published a study estimating that up to 98,000 hospital deaths each year can be attributed to medical malpractice (Messaoudi et al. 2013). In order to minimize medical errors, a medical social network, as a first contribution, destined to present patients' medical images and physicians' interpretations expressing their medical reviews and advices presents the solution to support collaboration between physicians and patients. This will, obviously, help to save time and better serve the patients about their situations.

But, the need, to index and retrieve medical images content shared through a social network, becomes obviously a requirement due to the large volume of images uploaded by physicians and patients. It is necessary to employ new techniques in order to effectively handle and benefit from images and to perform similarity queries and support decision making. The diversity of the available content inspired users to demand and formulate more complicated queries. Generally, image databases are text-annotated, whereby image retrieval is based on keyword searching. Such an approach has many disadvantages, keyword annotation is very subjective. Moreover, keyword-based image retrieval is not appropriate because there is no fixed set of words that describes image content. The content-based image retrieval (CBIR) was presented to address the difficulties related to the text-based image retrieval (Zeyad et al. 2014). The CBIR system practices image content to search and retrieve digital images. CBIR is a method, in which various visual contents have been considered to search and retrieve images on the basis of automatically derived features such as color, texture, and shape (Singh et al. 2012), from large scale of image databases based on the user's requests in the form of a query

image (Nandagopalan et al. 2008). A major problem with CBIR systems is the so-called semantic gap, which refers to the difficulty of translation of user's intentions into similarities among low-level features (Arevalillo-Herráez et al. 2013). To avoid the semantic gap, it is important to have the user interacting and telling what are the truly relevant images from those retrieved. This way of interaction between the user and the system is called relevance feedback (RF). Relevance feedback is a framework inherited from traditional information retrieval that has extensively been used to increase the efficiency of CBIR systems. Relevance feedback is one of the mechanisms, of increasing the accuracy of the retrievals (Bugatti et al. 2011).

In this paper, we propose the integration of a medical CBIR system with different stages of the relevance feedback process through the implemented medical social network site based on an efficient fusion of color, shape, and texture features. We, first, introduce the architecture of the integrated CBIR system, and then describe in detail each component of the system. The RF approach was proposed to learn the user's intentions by asking the specialist to quantify the relevance of each medical image retrieved by a query, allowing the system to automatically adjust future query results (Bugatti et al. 2011). The idea is that the system executes a similarity query using the undiagnosed image as the query center and the system displays the retrieved images that are similar to the given one, allowing the specialist to take advantage of previous analyses of them from user's comments, helping the improvement of the decision-making process and whose goal also is learning among medical residents and medical students that are currently, in the course of their training.

However, the study of the social network site effect and our methodology impact, in the decision-making process, become very important task. Indeed, the great majority of medical social network communities do not verify whether the methodology of existing CBIR system in a medical social network site is actually considered by the specialists and also patients as a potential aid in a real environment and to analyze his effectiveness for supporting physicians and patients in clinical routine. For this reason, we propose an additional contribution consisting of the evaluation of our method's impact in the user's decision, inquiring the specialists about their degree of confidence in the integrated CBIR system and in our social network, in generally.

The remainder of this paper is structured as follows. Section 2 presents the background needed to follow our methodology. Section 3 describes the design of our social network. Section 4 details the proposed methodology, shows experimental measurements, discusses the experiments, and analyzes the achieved results. Section 5

<sup>1</sup> <http://iom.nationalacademies.org/>.

presents the evaluation of our methodology and details the experiments and results achieved. Finally, Sect. 6 draws some conclusions and highlights future research directions.

## 2 Related work

### 2.1 Medical social networks

Today, social networks have the ability to connect people with just about everything. The influence of social network and those using social networks grows and changes daily, generating a profound impact on society. Furthermore, a growing majority of modern patients, particularly those with chronic conditions are seeking out social network and other online sources to acquire health information, connect with others affected by similar conditions, and play a more active role in their health care decisions (Daniel et al. 2013). In 2011, more than 80 % of adults reported using the Internet as a resource for health care quality information and more than half of patients (57 %) said they were more likely to select hospitals based on their social media presence (Feldman 2012). Indeed, research shows that 81 % of consumers believe that if hospitals have a strong media presence, they are likely to be more innovative than other hospitals. According to the Centers for Disease Control and Prevention (CDC), “Using social network tools has become an effective way to expand reach, foster engagement and increase access to credible, science-based health messages.” (Feldman 2012) Medicals networks have several forms which can be networks of hospital management (internal coordination and fragmentation within the hospital by specialty), resource networks (shared resources such as scanners), information networks (data collection to adjust policy information), and among many others care networks (Grenier 2003), which offers the members suffering from diseases an opportunity to change their lives, connect with others, and share problems. Indeed, research shows that 81 % of patients believe that if hospitals have a strong media presence, they are likely to be more innovative than other hospitals.<sup>2</sup>

According to Feldman (2012), hospitals are increasingly adopting the use of social network for a variety of key tasks, including: education and wellness programs, crisis communication, staff and volunteer training, employee and volunteer recruitment, information sharing, clinical trial recruitment and other research, public relations, and marketing. Since the beginning of 2011 alone, the growth in social network use for hospitals has been staggering. Ed Bennett, the manager of web operations at the University

Medical Center in Baltimore, has been tracking hospital social media on his private site since 2008. He reports that as of October 2011, nearly 4000 social media sites were owned by US hospitals (Feldman 2012). Many examples of medical social networks were presented in relation to different medical activities: SoberCircle<sup>3</sup> is intended for alcoholics and drug addicts in need for support and encouragement by others. SparkPeople,<sup>4</sup> Fitocracy,<sup>5</sup> and Dacadoo<sup>6</sup> do share workouts and exercises in order to sustain them during weight loss. Asklepios,<sup>7</sup> exclusively for Canadian doctors, is meant to exchange the best practices ever and to learn from each other. CardioSource,<sup>8</sup> Cardiothoracic Surgery Network, concerns cardiothoracic surgery. Diabospace<sup>9</sup> is intended for diabetics. “Parlons Cancer”<sup>10</sup> is dedicated to cancer patients and their families. Renaloo<sup>11</sup> deals with kidney disease, dialysis, and transplantation. RxSpace<sup>12</sup> is dedicated to pharmacy students, pharmacists, pharmacy owners, and academia to interact with each other.

Besides, we can summarize that the top 25 health and medical social network sites are organized in three. Firstly, social networks for doctors which offer great opportunities to confer find support and provide their own expertise, such as, Sermo (<http://www.sermo.com/>), Ozmosis (<https://ozmosis.org/>), MomMD (<http://www.mommd.com/>), Doctor Network (<http://doctor-network.com/>), etc.

Secondly, social networks for nurses which can help them to connect with others who understand what is happening in the field by asking and answering questions and learning more about their profession, such as, Nursing Link (<http://nursinglink.monster.com/>), Ultimate Nurse (<http://www.ultimatenurse.com/>), Nurse Zone (<http://nursezone.com/>), etc.

Finally, social networks for all health and medical careers, such as, MedicalMingle (<http://www.medicalmingle.com/>), Clinically Psyched (<http://clinicallypsyched.com/>), PatientsLikeMe (<http://www.patientslikeme.com/>), Radiolopolis (<http://radiolopolis.com/>), Docadoc (<http://docadoc.com/>), Carenity (<http://www.carenity.com/>), etc. These different kinds of social networking sites offer to their members an opportunity to

<sup>2</sup> <http://corp.yougov.com/healthcare/consumers-use-preference-expectations-hospital-social-media>.

<sup>3</sup> <http://www.sobercircle.com>.

<sup>4</sup> <http://www.sparkpeople.com/>.

<sup>5</sup> <https://www.fitocracy.com/>.

<sup>6</sup> <https://www.dacadoo.com/>.

<sup>7</sup> <http://www.asklepios.com/>.

<sup>8</sup> <http://www.acc.org/>.

<sup>9</sup> <http://www.diabospace.com/>.

<sup>10</sup> <http://www.parlonscancer.ca/>.

<sup>11</sup> <http://www.renaloo.com/>.

<sup>12</sup> <http://www.rxspace.com/>.

be connected with others and share experiences and knowledge.

Our first goal is to design a social network dedicated to physicians and patients. The basic model of the targeted social media should take into account the management of a:

- Set of patients which provide personal information on their health care profile.
- Set of physicians providing information enabling their identifications.
- Set of mechanisms permitting to patients to upload the medical images related to their diseases.
- Set of mechanisms permitting to physicians to comment the uploaded medical images.
- Set of search functions by which patients and physicians can locate easy and efficient information about medical images.
- Site operator who controls the site and triggers a set of mechanisms permitting to collect medical images in order to process them for various purposes such as medical images' indexation.

Like in (Feldman 2012), our social network, addressed to physicians and patients, can connect millions of voices to:

- Increase the timely dissemination and potential impact of health and safety information;
- Leverage audience networks to facilitate information sharing;
- Personalize and reinforce health messages that can be more easily tailored or targeted to particular audiences;
- Empower people to make safer and healthier decisions;
- Facilitate interactive communication, connection, and public engagement;
- Updates patients about changes in physician's practice;
- Keeps patients informed about upcoming appointments, tests, and immunizations;
- Engages patients in discussions about key health issues;
- Answers patients' medical questions;
- Communicates with family members, other caregivers;
- Grows physician's practice;
- etc.

In our work, we respect that our social network needs to contain all of these features, situated above.

## 2.2 Content-based image retrieval

### 2.2.1 An overview

Content-based image retrieval (CBIR) is an image search technique designed to find images that are most similar to a given query. It complements text-based retrieval by using

quantifiable and objective image features as the search criteria (Smeulders et al. 2000). Essentially, CBIR measures the similarity of two images based on the similarity of the properties of their visual components, which can include the color, texture, shape, and spatial arrangement of regions of interest (ROIs) (Kumar et al. 2013).

The color is one of the most reliable visual features that are also easier to implement in image retrieval systems. Color is a property, i.e., light reflection aids information processing from medical image. Color is extensively a visual feature used in CBIR that moderates the robust and simple processing. Color sensitivity and color space were, respectively, proposed in finding color similar to how humans perceive color (Zeyad et al. 2014; Komali et al. 2012). Texture features are also intended to capture the granularity and repetitive patterns of surfaces within an image. Their role in domain-specific image retrieval is particularly vital due to their close relation to the underlying semantics in these cases, such as in aerial imagery and medical imaging. Texture features, such as gray-level co-occurrence matrix (GLCM), Markov random field (MRF) model, simultaneous auto-regressive (SAR) model, edge histogram descriptor (EHD), etc., have long been studied in image processing, computer vision, and computer graphics (Xiang-Yang et al. 2012).

The shape is also a visual feature that describes the contours of objects in an image, which are usually extracted from segmenting the image into meaningful regions or objects. It is known to be an important cue for human to identify and recognize real-world objects. However, since it is difficult to achieve such image segmentation for natural images, the use of shape features in image retrieval has been limited to special applications where the extraction of object contours is readily available such as in trademark images (Zhang and Ye 2010). Zhang and Lu (2004) broadly classify shape extraction techniques into two major groups: contour-based and region-based methods. Contour-based methods calculate shape features only from the boundary of the shape, while region-based methods extract features from the entire region. Because contour-based techniques use only a portion of the region, they are more sensitive to noise than region-based techniques, as small changes in the shape significantly affect the shape contour. Therefore, color image retrieval usually employs region-based shape features (Zhang et al. 2012). The choice of features is a critical task when designing a CBIR system because it is closely related to the definition of similarity. Features fall into several categories. Application-specific features are tuned to a particular problem and describe characteristics unique to a particular problem domain; they are semantic features intended to encode a specific meaning (Smeulders et al. 2000).

The major challenges for CBIR include the application-specific definition of similarity (based on users' criterion), extraction of image features that are relevant to this definition of similarity, and organizing these features into indices for fast retrieval from large repositories (Smeulders et al. 2000; Lew et al. 2006; Rui et al. 1999; Datta et al. 2008). According to Kumar et al. (2013), the similarity of image features can be measured in a number of ways. When the features are represented as a vector, distance metrics such as the Euclidean distance can be used. The notion of elastic deformation can be used to define similarity when subtle geometric differences between images are important. Graph matching enables the comparison of images based upon a combination of image features and the arrangement of objects in the images (or the relationships between them). Finally, statistical classifiers can be trained to categorize the query image into known classes (Kumar et al. 2013). A detailed discussion of various similarity measures can be found in (Akgül et al. 2011).

An underlying assumption of most CBIR systems is that the chosen image features used are sufficient to describe the image accurately. The choice of image features must, therefore, be made to minimize two major limitations: the sensory gap and the semantic gap (Smeulders et al. 2000). The sensory gap is the difference between the object in the world and the features derived from the image. It arises when an image is noisy, has low illumination, or includes objects that are partially occluded by other objects. The sensory gap is further compounded when 2D images of physical 3D objects are considered; some information is lost as the choice of viewpoint means an object may occlude part of itself (Kumar et al. 2013). The semantic gap is the conflict between the intent of the user and the images retrieved by the algorithm. It occurs because CBIR systems are unable to interpret images; they do not understand the "meaning" in the images in the same way that a human does. Retrieval is performed on the basis of image features not image interpretations (Kumar et al. 2013).

In order to overcome the semantic gap problem and to improve CBIR systems performance, relevance feedback (RF) is the process of automatically adjusting an existing query using the information feedback by the user about the relevance of previously retrieved objects such that the adjusted query (Shanmugapriya and Nallusamy 2014). The key issue in relevance feedback is how to effectively utilize the feedback information to improve the retrieval performance (Xin and Jin 2004). After obtaining the retrieval results, user provides the feedback as to whether the results are relevant or non-relevant. If the results are non-relevant, the feedback loop is repeated many times until the user is satisfied. Patil and Kokare (2011) provide an overview of the technical achievements in the research area of relevance feedback (RF) in content-based image retrieval (CBIR). It

also covers the current state of the art of the research of relevance feedback in CBIR, various relevance feedback techniques, and issues in relevance feedback. Hence many classical machine learning schemes may be applied to the RF, which include decision tree learning (MacArthur et al. 2000), Bayesian learning (Cox et al. 2000; Su et al. 2003), support vector machines (Wang and Hua 2011; Hoi et al. 2009; Wang et al. 2014), boosting (Tieu and Viola 2003) and so on. According to the strategy employed, RF techniques can be divided into two main categories: query point movement, and re-weighting schemes of the similarity measure. The query point movement techniques consider that a query is represented by a single query center. Therefore, at each user interaction cycle, the strategy estimates an ideal query center in the query space, moving the query center toward the relevant examples and away from the irrelevant ones. On the other hand, the re-weighting techniques usually focus on adjusting weights to each dimension of the feature vector emphasizing some dimensions and diminishing the influence of others (Doulamis and Doulamis 2006). In the proposed method, relevance feedback technique can be done using decision trees.

Many CBIR systems have been proposed in recent decades, including IBM's query by image content (QBIC) system (Flickner et al. 1995), which was used to search for famous artworks; others include the Virage framework (Bach et al. 1996), Photobook (Pentland et al. 1996). More recently, google search by image used the points, colors, lines, and textures in images uploaded by users to find similar images (Chechik et al. 2010). These recent developments mean that CBIR is a technology that is available to the masses. Other systems have been also presented like for examples: MARS, NeTra, PicHunter, Blobworld, VisualSEEK and SIMPLicity (Wang et al. 2012; Hu et al. 2011).

To evaluate CBIR systems, precision and recall are generally two quality measures defined to calculate the accuracy of an approximate search paradigm. Precision refers to the proportion of retrieved images that are relevant, i.e., the proportion of all retrieved images that the user was expecting. Recall is the proportion of all relevant images that were retrieved, i.e., the proportion of similar images in the dataset that were actually retrieved. The ideal case would be a retrieval system that achieves 100 % precision and 100 % recall (Kumar et al. 2013). The reality is that most current algorithms fail to find all similar images, and many of the retrieved images contain dissimilar images (false positives).

### 2.2.2 Content-based image retrieval in medical applications

Our focus in this section is on techniques adapted to the medical field. Several studies have already reported on the



potential clinical benefits of CBIR in clinical applications. The ASSERT CBIR system used for high-resolution CT (HRCT) lung images (Shyu et al. 1999) showed an improvement in the accuracy of the diagnosis made by physicians (Aisen et al. 2003). Another study for liver CT concluded that CBIR could provide real-time decision support (Napel et al. 2010). CBIR was also shown to have benefits when used as part of a radiology teaching system (Müller et al. 2005).

The image retrieval in medical applications (IRMA)<sup>13</sup> project has been a sustained effort in the CBIR of X-ray images for medical diagnosis systems. The IRMA approach is divided into seven interdependent steps (Keyers et al. 2003): (1) categorization based on global features, (2) registration using geometry and contrast, (3) local feature extraction, (4) category-dependent and query-dependent feature selection, (5) multi-scale indexing, (6) identification of semantic knowledge, and (7) retrieval on the basis of the previous steps. The IRMA method classifies images into anatomical areas, modalities, and viewpoints and provides a generic framework (Güld et al. 2007) that allows the derivation of flexible implementations that are optimized for specific applications.

A number of papers (Antani et al. 2004; Antani et al. 2003; Lee et al. 2003; Xu et al. 2008; Hsu et al. 2009; Lee et al. 2009; Qian et al. 2010) have described investigations into every component of CBIR for spine X-ray retrieval, including feature extraction (Antani et al. 2003; Lee et al. 2003; Qian et al. 2010), indexing (Qian et al. 2010), similarity measurement (Qian et al. 2010), and visualization and refinement.

For the purpose of this study, it is essential to integrate or benefit from the use of the following works such as Agma et al. (2007), Singh et al. (2009), Bugatti et al. (2011), and Harishchandra et al. (2014). Starting from the proposal of Agma et al. (2007), this study introduced the extraction of texture features, by statistical features and color features, by the gray-level histogram, for the indexing and retrieval of images, in the context of a comparative study between descriptors. The first step of the method is a segmentation process based on Markov random fields. After the segmentation process, features are extracted from each cluster. Euclidean distance was applied to measure the similarity between images' features.

Greenspan (2007) indexed the images by the Gaussian mixture modeling (GMM) as a texture descriptor to construct vectors of features. Kullback–Leibler (KL) measure was used to calculate the similarity between features. In Selvarani and Annadurai (2007), an image retrieval system is provided to retrieve the images using as texture

representation, statistics measures (mean and variance). They presented a method of texture analysis based on the use of Gabor filters. Euclidean distance was used to measure the similarity between images index. Zhang et al. (2008) showed also good performance with a new approach based on directional texture extracted from images. Direction histogram is computed from the enhanced Fourier spectrum. Gray-level co-occurrence matrix extracts texture characteristics based on the set. Euclidean distance is used, also, to evaluate the probability of similarity between images.

In Wu and Tai (2009), the texture features were extracted from images by using the gray-level variation histogram. Earth mover's distance was used to measure the image similarity. Singh et al. (2009) proposed also a medical CBIR retrieval framework, a method of texture analysis for image retrieval, based on the use of Gabor filters. The similarity used for comparing images index was the Euclidean distance. Moreover, Seng and Mirisaei (2009) have experimented various similarity measures comparison. This study was based on the extraction of texture features by the wavelet transform and color features by global color histogram. In the same context, Bugatti et al. (2009) have extracted texture features by the gray-level co-occurrence matrix and color features by the gray-level histogram for the image indexing and retrieval. As similarity measure, they used Minkowski and Canberra distances. Other approaches for radiograph retrieval have tried to group features into semantically meaningful patterns. In one such study (Iakovidis et al. 2009), multi-scale statistical features were extracted from images by a 2D discrete wavelet transform. These features were then clustered into small patterns; images were represented as complex patterns consisting of sets of these smaller patterns. Experimental results revealed that the method had significantly higher precision and recall compared to two conventional approaches: local and global gray-level histograms.

Furthermore, Sidong et al. (2010) have extracted, for their medical CBIR system, texture feature based on curvelet transform, Gabor filters, and wavelet transform. Euclidean distance is used to measure similarity between the images. Bueno et al. (2010) have proposed a texture retrieval algorithm for images, based on the extraction of the Haralick's texture features and Zernike moments. The measure used for similarity between images index is Manhattan distance. De Oliveira et al. (2010) have proposed a breast density feature extraction for the image retrieval, to help the radiologists in their diagnosis. Bugatti et al. (2011) have proposed also a method based on k-nearest neighbor classification algorithm for MRI medical image classification with color descriptors, by the gray-level histogram, as extracted features. In the same context,

<sup>13</sup> IRMA Homepage (English): [http://www.irma-project.org/index\\_en.php](http://www.irma-project.org/index_en.php).

Bhattacharjee and Parekh (2011) have experimented various similarity measures (Manhattan distance and Euclidean Distance) based on the extraction of texture features by the gray-level co-occurrence matrix.

In order to present a comparative study between features, Ashish and Manpreet (2012) indexed the images by embedding the shapes features extracted by the Fourier descriptors and the Haar wavelet transform. John and Kazunori (2012) used other kinds of features: SIFT scale invariant feature transform with various similarity measures (Minkowski and standard measures, statistical measures and divergence measures). Ramamurthy and Chandran (2012) have proposed an efficient image retrieval tool namely, “Content Based Medical Image Retrieval with Texture Content using gray-level co-occurrence matrix (GLCM) and *k*-Means Clustering algorithms.” This image retrieval tool is capable of retrieving images based on the texture feature of the image, and it takes into account the preprocessing, feature extraction, classification and retrieval steps in order to construct an efficient retrieval tool. The main feature of this tool is used of GLCM of the extracting texture pattern of the image and *k*-means clustering algorithm for image classification in order to improve retrieval efficiency. The proposed image retrieval system consists of three stages, i.e., segmentation, texture feature extraction and clustering process. The result indicates that the tool gives better performance in terms of percentage for all the 1000 real-time medical images from which the scalable performance of the system has been proved. In other side, Ramamurthy et al. (2012) develop an efficient visual-content-based technique by the extraction of texture feature by the local binary pattern algorithm (LBP) and intensity feature. The extracted feature images are combined to form a single feature vector value, and it is then compared using Euclidean distance method to retrieve similar images from the database.

As a part of a comparative study, a number of statistic texture features have been proposed in (Swarnambiga and Vasuki 2013) with various similarity measures, in order to have higher retrieval accuracy. Rajakumar and Muttan (2013) presented a CBIR system based on wavelet transform. They used the wavelet-based texture features. For retrieval phase, they used the Mahalanobis distance. Yogapriya and Ila (2013) have introduced another medical CBIR system based on texture features by using local binary patterns (LBP) and the Euclidean distance.

Harishchandra et al. (2014) proposed a new technique of the CBIR using the combination between color descriptors, extracted by color auto-correlogram, and texture descriptors, extracted by co-occurrence matrix. A modified Minkowski distance was used to measure the similarity between the query image and the database images.

## 2.3 Discussion

We have made a comprehensive review on the state of the art of medical and non-medical CBIR techniques. Most approaches found, in the literature, focused on two major aspects of CBIR: feature extraction and similarity measures. In terms of features, several types of features have been used. These features can be extracted either locally or globally. Most of these systems have a very similar architecture for browsing and archiving/indexing images comprising tools for the extraction of visual features, for the storage and efficient retrieval of these features, for distance measurements or similarity calculation and a type of graphical user interface (GUI). However, no contribution is at the present time about the integration of CBIR systems through medical social networks. An attempt to index and retrieve medical image content shared through the social network, using techniques of the content-based image retrieval community, becomes an important task due the huge volumes of uploaded images, in order to analyze and to benefit from our medical social network content. But, the main question we address is: Which features are suitable for the task of medical image retrieval? Currently, all existing features have limitations of describing medical images, according to our study, and none of existing features is powerful enough to represent the large variety of images. Common practice is to combine several types of features to represent as many images as possible. Indeed, it is hard to attain satisfactory retrieval results by using a single feature because, in general, an image contains various visual characteristics. Moreover, user interaction and relevance feedback are two other techniques that need to be more integrated into retrieval systems as this can help to lead to much better results. Image retrieval needs to be interactive and all the interaction needs to be exploited for delivering the best possible results. Our present study's approach is about the integration of a robust medical image retrieval scheme with relevance feedback through a social network for collaboration, whose goals are teaching, learning (among medical residents and students), and diagnostics. In what follows, it is vital to explicitly describe this approach.

## 3 Social network description and implementation

Social network is the study of social entities, as a group of people linked to one another by one or more common attributes, and their interactions and relationships. The interactions and relationships can be represented with a network or graph, where each node represents an actor and each link represents a relationship. A social network plays an important role as a support for the dissemination of

information, ideas, and impact among its members. It may be a major selling point for patients looking for physicians. In fact, social networking has been proven effective in sharing knowledge and establishing communication among patients and physicians.

Our proposed social network, as many others described in the literature Gong and Sun (2011) and Almansoori et al. (2011), aims to allow patients and physicians to connect with each other by eliminating all geographical and time frontiers. Both users exploit it to seek advices and to share experiences related to medical images' interpretations and diseases' analysis.

Like in Facebook, patients and physicians need to be registered by creating profiles about themselves (detailed information). This step is very important in order to use our medical social network. Our social network site is a virtual place where registered patients upload their medical images to be commented by various registered physicians. Not only registered patients can upload their medical images, but also registered physicians have this right, especially medical students, in the objective to learn from this collaboration. Uploading and sharing images are not the only functionalities. Users can also share status, medical events, and any information related to medical activities.

The main objective, behind using our medical social network, is to enable patients and physicians to exchange information, share knowledge and experiences. It, also, gives to its members the opportunity to connect and communicate with others in need of support and encouragement related to different situations and health care problems. Consequently, the patient's condition represents the heart of the health system justified by experts' interpretations and analysis expressed by comments. Patients frequently place trust in peer recommendations on social network site making them key platforms to influence change. We used PHP (Hypertext Preprocessor), in the development of our social network, and MySQL for the database management. PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. It ensures many features that allow creating and managing an entire social network website. Pages were written in HTML, PHP, Ajax, and JavaScript, and they were designed using Dreamweaver.

In order to better explain our medical social functionalities, we describe in the following some interfaces of the social network:

- The first interface to use is the identification interface. Patients and physicians need to provide the registration process, as new users. The registration interface

dedicated to physicians has more specificity because profiles have different structures.

- After the identification, the basic functions and services of the social network are displayed in the main interface. Patients and physicians can perform different activities, especially, get access to posting, commenting functions, uploading images, etc.
- In the image retrieval interface (the user interface components), patients and/or physicians try make a diagnosis for a new medical image, by searching to most similar image to a given query, according to the visual extracted features. Both need to complete a form for the visual features, in relation by a link to an online tool for features extraction. The same process will be treated by back-office functions implementing our approach.
- To answer urgent questions and/or advices, synchronous communications between members (patients and physicians, physicians and physicians) are performed in the communication interface. Mixed posts performed by different physicians and patients, according to their opinions and questions, are displayed in the posting interface.

The different interfaces of our social network allow an easier mapping for users, their intent and targeted functions. The most important that these interfaces describe the way of interaction between patients and physicians with the social network site and the way of interaction between patients and physicians with each other's by exploiting existing functions. The social network content blocks are visually separated. This separation will help to organize pages content, and each element is well defined and presented separately, to succeed the interaction between patients and physicians. In fact, this separation makes the content understandable, easy to recognize and makes the content more reachable.

In order to improve the interaction between users (patients and physicians, physicians and physicians), our social network contains an advanced search function, allowing the organization of the connections between patients and physicians having common interests related to diseases and their analysis, interpretations, and treatments. This function supports users to rapidly find the content and contacts they are searching for.

Considering the recommendations of a design expert, the created interfaces are simple in terms of color scheme and graphics, for example in Facebook. The idea consists of using a few colors and the background is generally white. This management of color scheme helps the exploitation of our social network site by physicians and patients, makes the content well presented, and comments



better seen and readable. Therefore, buttons and links are placed almost on every page of the social network. Some links are related to the navigation processes, and some others permit to users to regulate specific functions. Buttons are used to associate users to actions and to navigate among different pages; they are clear and more remarkable.

When incoming comments or messages appear on the social network sites, physicians and patients need to react to them in real time. For this reason, they should consider establishing their own social network presence. So, we assure our social network with a synchronous communication, in order to establishing interaction and sharing information. We, also, equipped our social network with a real-time update feature ensuring the delivery of updates (medical images uploading, physicians' annotations, etc.) as soon as they are submitted. In order to encourage patients and physicians to exploit our social network site, many actions could be performed such as suggesting new friends, preferably in relation to medical activities, interests, events, and groups. This is performed to extend their social circles allowing implicitly the extension of the shared knowledge about medical images and the associated diseases.

The extraction of more knowledge, from various posts or comments, is possible by using mining tools. For this reason, our social network is equipped by mining tools such those presented in Xie et al. (2013), where authors consider the emergence and pervasivity of online social networks that have enhanced web data with developing interactions and communities both at large scale and in real time.

The successful building of our social network needs to respect security aspects and to protect users' privacy.

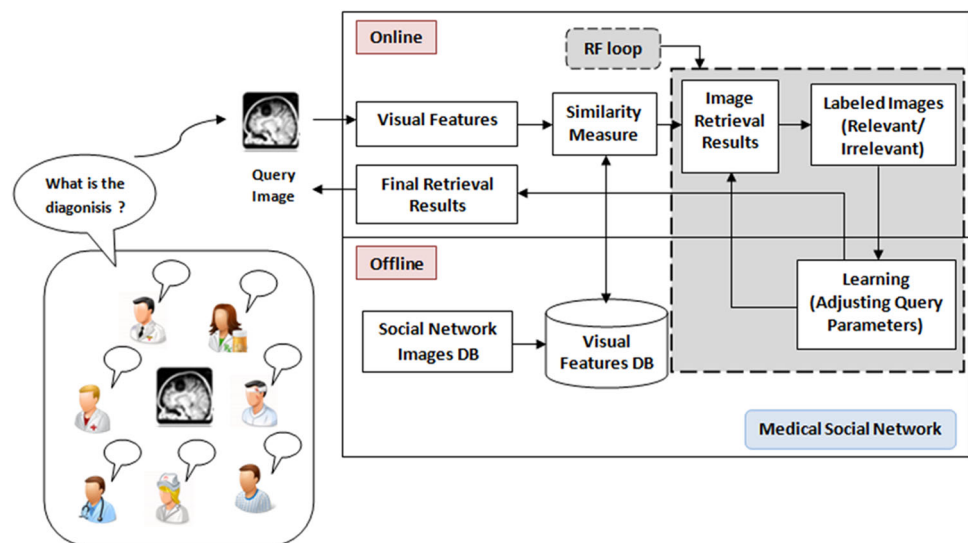
Physicians must protect their own privacy, as well as that of their patients. To do it, we use privacy settings on our social network site to keep out everyone except patients or fellow physicians. We will also consider security aspects such those presented in Li (2014), where the author proposed methods to secure health care social networking sites providing users with tools and services to easily establish contact with each other around shared problems and utilize the wisdom of masses to outbreak disease.

## 4 The proposed methodology

### 4.1 An overview

Figure 1 shows a structure of the conceptual design for the CBMIR method through the social network site. From Fig. 1, we can see that our methodology mechanism has five main components: query unit, extraction unit, retrieval, labeling, and learning unit. Compared with the traditional CBMIR without feedback, there are two more components which are labeling and learning, and they are the key contributions in the relevance feedback system. During the offline phase, it first extracts three kinds of visual features for each image in the image social network database *SNDB*, which are color, texture and shape features, then, stores these features into the feature database (the extraction component). Note that there are a large amount of methods for these three kinds of features' extractions. Accordingly, it is a very important to select efficient features and to choose effective methods to extract them, in order to perform image retrieval. Indeed, the combination

**Fig. 1** Proposed methodology pipeline



between visual features, in order to construct a single feature's vector, will give more precision to the retrieval process.

During the image retrieval, as shown in Fig. 1, three visual features of a query image are first extracted by using three feature extraction methods which are determined in the offline phase (the query component). The query image represents the undiagnosed image proposed by physicians, specialists, or patients. They decide whether they want to perform a similarity query to support their diagnosis or to improve their certainty about an image analysis.

Then, in the matching process, the system executes a similarity query using the undiagnosed image as the query center. The similarity, between the query image and each image in *SNDB*, can be computed by using a corresponding distance formula which is decided among those existing in the literature (in our case the Euclidian distance). Finally, the CBMIR method outputs several most similar images to the given one according to the above similarity (the retrieval component), allowing physicians or patients to take advantage of previous analysis of them based on existing comments and posts.

In order to satisfy user, a strategy employed, in our CBMIR system, to obtain a better approximation of the user's expectations and preferences is the relevance feedback (RF) (labeling and learning components). Relevance feedback technique allows the user to judge the returned answers, informing which images are relevant according to the image query. Thus, it is a real-time learning strategy that adapts the answer from a retrieval system exploring the user interaction, leading to higher precision and improved query refinement. The user (physician or patient) indicates the irrelevant images, which are eliminated from the interface, from the relevant ones. If he is not satisfied with the results (the initial images retrieved) and wants to refine the search, it can be done through a relevance feedback iteration (RF loop). The users were stimulated to try a relevance feedback method [the Rocchio algorithm in this experiment (Rocchio 1971)] to perform a refined query and analyze the new result set. They could use as many relevance feedback cycles, until they have satisfactory results, as needed to improve the accuracy and the certainty of their diagnosis. Finally, users finalize search. Users have, also, the choice to finalize search from the beginning if they are really satisfied. The algorithm, describing the steps of the proposed method in general, is as follows:

#### CBMIR WITH RF ALGORITHM

##### INPUT

$M_i$ : Medical Image/The query image

$M_{i_{snb}}$ : Medical Images in the social network database (*SNDB*)

##### OUTPUT

$n$  images similar to the input image

##### VAR

$FV_{qi}$ : Feature's Vector/The query image

$FV_{idb}$ : Feature's Vector for images in the *SNDB*

Similarity: real [The distance between images]

$nbf$ : number of features vectors in The DB

$j$ : integer

##### 0.BEGIN

//Features' extraction for medical images in database

1. FOR EACH  $M_{i_{snb}}$  IN The *SNDB* DO

2. BEGIN

3. EXTRACT Color Features;

4. EXTRACT Texture Features;

5. EXTRACT Shape Features;

6.  $FV_{idb} \leftarrow$  CONSTRUCT(Combination between color, texture and shape features );

7. STORE  $FV_{idb}$  IN The features' DB

8. ENDFOR;

//Features' extraction for query image

9. FOR EACH  $M_i$  DO

10. BEGIN

11. EXTRACT Color Features;

12. EXTRACT Texture Features;

13. EXTRACT Shape Features;

14.  $FV_{qi} \leftarrow$  CONSTRUCT(Combination between color, texture and shape features );

15. ENDFOR;

// Retrieve the images

16. Similarity  $\leftarrow 0$

17. FOR EACH  $FV_{idb}$  IN The features' DB DO

18. FOR ( $j=0$ ;  $j \leq nbf$ ;  $j=j+1$ );

19. Similarity  $j \leftarrow$  DISTANCE ( $(FV_{idb}), (FV_{qi})$ );

20. ENDFOR;

21. ENDFOR;

22. SWITCH (Similarity )

23. RETRIEVE the first  $n$  similar images;

24. END;

25. DISPLAY the first  $n$  similar images;

// Relevance Feedback

26. IF user ISNOT satisfied THEN

27. REPEAT UNTIL user satisfaction

28. Providing judgment on the currently displayed images;

29. User indicates if they are relevant or irrelevant to her/his request;

30. Machine learns the judgment of the user and again search for the images according to user query;

31. DISPLAY new results;

32. FINALIZE search;

33. END REPEAT;

34. ELSE

35. FINALIZE search;

36. ENDFOR;

37. END.

## 4.2 Feature extraction

The first issue, in image retrieval scheme, is to extract the features of the image efficiently and, then, represent them in a particular form to be used effectively in the matching of images. As an unstructured array of pixels, the first step in semantic understanding of images is to extract efficient and effective visual features from these pixels. Appropriate feature representation significantly improves the performance of the semantic learning techniques (Zhang et al.

2012). In the following, we present, in detail, various feature extraction techniques used in this paper. Both global and region-based image representations are used in the existing image retrieval scheme.

#### 4.2.1 Color features

Color-based image retrieval is the most basic and most important method for CBIR. Color features are the most intuitive and most obvious image features. It is also an important feature of perception. Comparing with other image features such as texture and shape, etc., color features are very stable and robust. It is not sensitive to rotation, translation, and scale changes (Yue et al. 2010). Moreover, the color features calculation is relatively simple. In the proposed image retrieval method, we capture the characteristics of an image's color contents by using the global gray-level histogram and color moments. These features are calculated by converting each pixel to gray scale using the following formula (Yue et al. 2010):

$$\text{Grey} = 0.299 \text{ Red} + 0.587 \text{ Green} + 0.114 \text{ Blue} \quad (1)$$

**4.2.1.1 Global gray-level histogram** The gray-level histogram describes the gray-level distribution of an image (Goh et al. 2005). The gray-level space needs to be divided into several small ranges in order to calculate the gray-level histogram. Each interval is regarded as a bin. Thus, the gray-level is quantized. The gray-level histogram can be calculated through counting pixels where the gray-level fall into each interval (Yue et al. 2010). It quantizes a gray-level space into different bins and counts the frequency of pixels belonging to each gray-level bin. According to the study proposed by Zhang et al. (2012), this feature is robust to translation and rotation changes. However, a gray-level histogram does not tell pixels' spatial information. In addition, the dimension of a histogram is usually very high. The gray-level histogram defined the mathematical equations given in Eqs. 1 and 2 (Zeyad et al. 2014), where  $I(x, y)$  represents the probability of pixel falling into the bin:

$$hc(m) = \frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I(x, y) \quad (2)$$

$$I(x, y) = \begin{cases} 1 & \text{if } I(x, y) \text{ in bin } m \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

**4.2.1.2 Color moments** Color moments are one of the simplest features. They are used in many retrieval systems (Fazal and Baharum 2013; Afifi and Ashour 2012). The common moments are mean, standard deviation, skewness, and kurtosis. They have been proved to be efficient and effective in representing color distributions of images. In our case, we are interested in representing gray-value

distributions of images. These features are useful when they are calculated for region or object. According to Zhang et al. (2012), these features are compact and robust. However, the moments are not enough to represent all the color information of an image. If the value of the  $i$ th gray-value channel at the  $j$ th image pixel is  $p_{ij}$ , then the gray-value moments are as follows:

$$\text{Mean} : m_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad (4)$$

$$\text{Standard deviation} : \sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{ij} - m_i)^2} \quad (5)$$

$$\text{Skewness} : s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (p_{ij} - m_i)^3} \quad (6)$$

$$\text{Kurtosis} : k_i = \sqrt[4]{\frac{1}{N} \sum_{j=1}^N (p_{ij} - m_i)^4} \quad (7)$$

#### 4.2.2 Texture features

Texture is another important image feature. While color is usually a pixel property, texture can only be measured from a group of pixels. Texture has been well studied in image processing and computer vision area (Zhang et al. 2012). A number of techniques have been proposed to extract texture features. For our work, the texture feature is extracted by using spatial texture feature extraction method. Statistical texture features (Haralick features) characterize texture as a measure of low-level statistics of gray-level images. Haralick features are energy, entropy, contrast, correlation, and inverse difference moment. These features are derived from the gray-level co-occurrence matrix (GLCM). According to the study of Zhang et al. (2012), statistical features are intuitive, compact, and robust because they are derived from large support. However, they are not sufficient to describe the large variety of textures.

According to Zeyad et al. (2014), the co-occurrence matrix  $C(i, j)$  is used to compute the co-occurrence of pixels with gray values  $i$  and  $j$  as a given distance that is the polar coordinates  $(d, \theta)$  with discrete length and orientation. In practice,  $\theta$  takes the values  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $225^\circ$ ,  $270^\circ$ , and  $315^\circ$ , respectively. The co-occurrence matrix  $C(i, j)$  is defined as follows according to Eq. 7.

$$C_{\theta^d}(i, j) = \# \left\{ \begin{array}{l} 1, \text{ if } I(x_1, x_2) = i, \text{ and } I(x_1 + d \cos \theta, y_1 + d \sin \theta) = j \\ 0, \text{ otherwise} \end{array} \right\} \quad (8)$$

where  $\#$  denotes the number of elements in the set,  $d$  represents the distance between gray-level value  $i$  and  $j$  of

an image,  $i, j = 0.255$  the number of possible gray levels in the image, while the co-occurrence matrix  $C(i, j)$  dimension is  $M \times N$ . The co-occurrence matrix measures can be calculated as follows:

$$\text{Energy} : E_{ne} = \sum_i \sum_j C_{\theta^0, d}^2(i, j) \quad (9)$$

$$\text{Entropy} : E_{nt} = - \sum_i \sum_j C_{\theta^0, d}(i, j) \log C_{\theta^0, d}(i, j) \quad (10)$$

$$\text{Contrast} : C_{on} = \sum_i \sum_j (i - j)^2 C_{\theta^0, d}(i, j) \quad (11)$$

$$\text{Correlation} : C_{or} = \left( \sum_i \sum_j C_{\theta^0, d}(i, j) - \mu_x \mu_y \right) / \sigma_x \sigma_y \quad (12)$$

Inverse difference moment :

$$I_{dm} = \sum_i \sum_j (C_{\theta^0, d}(i, j) / |i - j|^2), \quad i \neq j \quad (13)$$

where,  $\mu_x \mu_y$  represents the means and  $\sigma_x \sigma_y$  represents the standard deviations. They are defined as follows:

$$\mu_x = \sum_i i \sum_j C(i, j) \quad (14)$$

$$\mu_y = \sum_j j \sum_i C(i, j) \quad (15)$$

$$\sigma_x = \sum_i (i - \mu_x)^2 \sum_j C(i, j) \quad (16)$$

$$\sigma_y = \sum_j (j - \mu_y)^2 \sum_i C(i, j) \quad (17)$$

### 4.2.3 Shape features

Shape is an important and most powerful feature used for images' indexing and retrieval. Shape features have been employed for image retrieval in many applications. A number of simple and statistical region shape descriptors are commonly used in our work, including, area, circularity, eccentricity, roundness, and convexity. The area-based descriptor is used in a number of works (Mezaris et al. 2005; Goh et al. 2005). Circularity is used in Song 2012; Stojmenovic et al. 2013. Circularity measures the ratio of area to boundary. In Stojmenovic and Nayak 2008, eccentricity or elongation is also used in addition to area. Eccentricity is the ratio of the length of the major axis to that of minor axis. Roundness and convexity are used in Song 2012. Roundness measures the ratio of area to the major axis, and convexity is the ratio of the area to the convex area. According to the study of Zhang et al. (2012), Individual simple shape descriptors are not robust. Therefore, they are normally combined to create a more effective shape descriptor. These features can be calculated as follows:

$$\text{Circularity (compactness)} : C_{irc} = 4\pi \times \left( \frac{A}{P^2} \right) \quad (18)$$

$$\text{Elongation (aspect ratio)} : E_{lon} = \frac{M_A}{m_A} \quad (19)$$

$$\text{Roundness} : R_{oun} = \frac{4A}{\pi M_A^2} \quad (20)$$

$$\text{Solidity (convexity)} : C_{onv} = \frac{A}{C_A} \quad (21)$$

where  $A$  represents the area,  $P$  represents the perimeter,  $M_A$  is the major axis,  $m_A$  is the minor axis and  $C_A$  is the convex.

After the calculation of these features (color, texture and shape features), these values are combined to get a feature vector. The feature vectors ( $FV_{s_{idb}}$ ) of all the images are constructed and stored to create features database. The feature vector  $FV_{qi}$  of the user query is also constructed in the same way and compared with the feature vectors of the database for similarity and searching of relevant images. The combination between features represents a solution to exceed limits of each single feature. The algorithm, describing the indexation phase of the proposed method with more specification about the extracted feature, is as follows:

#### INDEXATION ALGORITHM

##### INPUT

$M_{iq}$  : Medical Image[The query image]

$M_{s_{idb}}$  : Medical Images in the social network database (SNDB)

##### OUTPUT

$FV_{qi}$  : Feature's Vector[The query image]

$FV_{s_{idb}}$  : Feature's Vector for images in the SNDB

##### 0.BEGIN

//Features' extraction for medical images in database

1. **FOR EACH**  $M_{s_{idb}}$  **IN** The SNDB **DO**

2. **BEGIN**

3. **EXTRACT** Grey Value Moments Features (GVMF);

4. **GENERATE** The Grey-Level Histogram(GLH) **AND COUNT** each feature value;

5. **EXTRACT** Haralick Features By The GLCM(HF);

6. **EXTRACT** Statistical Shape Features(SSF);

7.  $FV_{s_{idb}} \leftarrow \text{CONSTRUCT}$ (Combination between GVMF, GLH, HF and SSF);

8. **STORE**  $FV_{s_{idb}}$  **IN** The features' DB

9. **ENDFOR**;

//Features' extraction for query image

10. **FOR EACH**  $M_{iq}$  **DO**

11. **BEGIN**

12. **EXTRACT** Gray Value Moments Features (GVMF);

13. **GENERATE** The Grey-Level Histogram(GLH) **AND COUNT** each feature value;

14. **EXTRACT** Haralick Features By The GLCM(HF);

15. **EXTRACT** Statistical Shape Features(SSF);

16.  $FV_{qi} \leftarrow \text{CONSTRUCT}$ (Combination between GVMF, GLH, HF and SSF);

17. **ENDFOR**;

18. **END**.

### 4.3 Similarity measure

Once features database ( $FV_{s_{idb}}$ ) of stored images is created, and the feature vector of the query image ( $FV_{qi}$ ) is computed, as shown in Fig. 1, similarity measurement presents the second issue. To measure the similarity, the difference is calculated by using one of distance metrics. We used the



Euclidean distance. Euclidean Distance is used to measure the similarity between two images with N-dimensional feature vector. The vectors of the images with small distances are most similar to the query images (the top 10 retrieved images in our case). We suppose that  $FV_{qi} = (f_{q0}, f_{q1}, \dots, f_{qN-1})$  and  $FV_{idb} = (f_{db0}, f_{db1}, \dots, f_{dbN-1})$ , so the Euclidean distance is calculated by using the following formula (Fazal and Baharum 2013; Swarnambiga and Vasuki 2013):

$$\text{Euclidean distance} = \sqrt{\sum (f_{qi} - f_{dbi})^2}. \quad (22)$$

The algorithm, describing the retrieval phase of the proposed method with more specification about the metric used to calculate the similarity, is as follows:

---

#### RETRIEVAL ALGORITHM

---

##### INPUT

$FV_{qi}$  : Feature's Vector/The query image/

$FV_{idb}$  : Feature's Vector for images in the SNDB

##### OUTPUT

The top 10 images similar to the input image

##### VAR

Similarity : real [The distance between images]

nbf: number of features vectors in The DB

j: integer

##### 0.BEGIN

1. Similarity  $\leftarrow 0$

2. **FOR EACH**  $FV_{idb}$  **IN** The features' DB **DO**

3.   **FOR** ( $j = 0; j \leq \text{nbf}; j = j + 1$ );

// Application of the Euclidean distance

4.   Similarity  $j \leftarrow \text{SQRT} ((FV_{qi} - (FV_{idb}))^2)$ ;

5.   **ENDFOR**;

6. **ENDFOR**;

7. **SWITCH** (Similarity )

8.   **RETRIEVE** the first 10 similar images;

9. **END**;

10. **DISPLAY** the first 10 similar images;

11. **END**.

## 4.4 Relevance feedback

Basically, the relevance feedback technique is composed of three steps. In the first step, the system retrieves the most similar images according to the initial query. During the second step, the user (patient or physician) guides the search process, judging the returned images, and weighing the returned images based on a relevance degree (relevant or irrelevant). In the third step, the system captures the user's expectation based on the performed feedback and automatically adjusts the further queries based on the user's informed relevance. The second and third steps are repeated until the user is satisfied with the results. As the system captures the user's intention when a new query is performed, the resulting set of images can be continually improved until the gain flattens, according to the iterative learning process (Liu et al. 2007). We performed the Rocchio algorithm in this experiment (Rocchio 1971). We use the query point movement (QPM) (Doulamis and

Doulamis 2006) technique to perform RF. Rocchio's formula is the mostly used technique to iteratively improve this estimation:

$$Q' = \alpha Q + \beta \left( \frac{1}{N_{R'}} \sum_{i \in D'_{R'}} D_i \right) - \gamma \left( \frac{1}{N_{N'}} \sum_{i \in D'_{N'}} D_i \right) \quad (23)$$

where  $Q$  is the original query (the initial feature vector),  $Q'$  is the updated query (the resultant feature vector),  $D'_{R'}$  and  $D'_{N'}$  are respectively the positive and negative examples given by the user,  $N_{R'}$  and  $N_{N'}$  are the number of relevant and irrelevant examples in  $D'_{R'}$  and  $D'_{N'}$ , and finally  $\alpha$ ,  $\beta$ , and  $\gamma$  are weighting factors, experimentally obtained, to tune the relevance feedback factors. The scenario for the relevance feedback steps in our methodology is as follows:

---

#### RELEVANCE FEEDBACK ALGORITHM

---

##### INPUT

The top 10 images similar to the input image

##### OUTPUT

Final results of similar images

##### 0.BEGIN

1. **REPEAT UNTIL** User satisfaction **OR** Keeping same results

2.   **OBTAIN** the feedback from user interaction on prior results (relevant or irrelevant to request);

3.   **LEARN** the system through a feedback algorithm;

4.   **REFINE** results **AND REMOVE** the irrelevant images ;

5. **END REPEAT**;

6. **DISPLAY** Final results;

7. **END**.

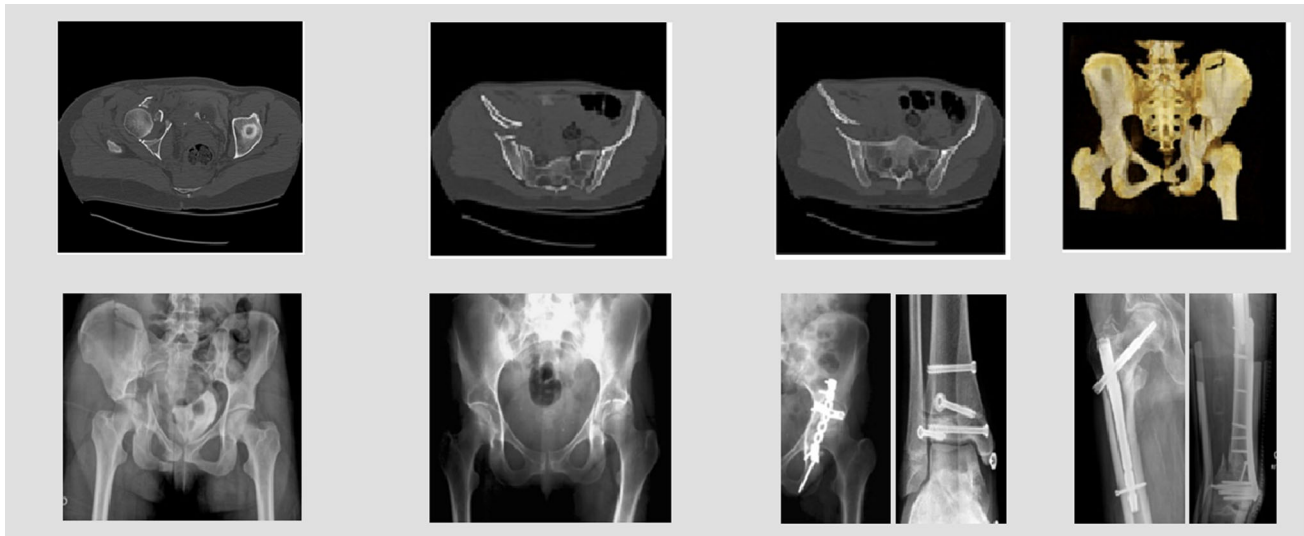
## 4.5 Experimental results

Experiments have been carried out to validate the efficiency of the proposed model to execute similarity queries, through a graphical user interface (GUI) belongs to our social network. We explain the results, comment on them, and compare our system with other existing systems.

### 4.5.1 Feature extraction tool

Features extraction can be achieved with the help of certain image manipulation tools such as ImageJ. This tool helps us to efficiently extract various features to be used in our application for some specific searches. ImageJ<sup>14</sup> is a public domain Java image processing and analysis program inspired by NIH Image for the Macintosh. It runs, either as an online applet (used for the online features extraction), integrated by a link in our social network, or as a downloadable application (used for the offline features extraction). It can display, edit, analyze, process, and save images. It can calculate pixel value statistics of user-defined selections. ImageJ was designed with an open

<sup>14</sup> <http://imagej.nih.gov/ij/>.



**Fig. 2** Examples of images in the database

architecture that provides extensibility via Java plug-in. User-written plugins make it possible to solve almost any image processing or analysis problem. The various features, which are described in Sect. 4.3, are extracted by using ImageJ, and then the feature's vector of each image is constructed and stored in the database.

#### 4.5.2 Data test and evaluation criteria

The relevance of our approach was evaluated on pre-selected medical images, which are supplied by internal physicians and residents during their training, diagnostic and images analysis in relation to patients' states and clinical cases, at Charles Nicolle Hospital in Tunis. This collection contains 500 medical images in the JPEG format. We perform experiments over images from this collection, classified in five categories in generally: Thorax, abdomen, lumbar spine, pelvis, skull, and others. Every database image is of size  $865 \times 705$  or  $787 \times 787$  pixels. For evaluation, we use all the images in the database. Figure 2 illustrates example of images in the database.

We evaluate the performance of this approach using the metric of precision, the recall, AP (average precision), and MAP (mean average precision) (Manning et al. 2008). The precision is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images. We denote to the precision by  $P$ . The equation of the precision is:

$$P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}. \quad (24)$$

The recall is defined as the ratio of the number of retrieved relevant images to the total number of relevant

images in the database. We denote to the recall by  $R$ . The equation of the recall is:

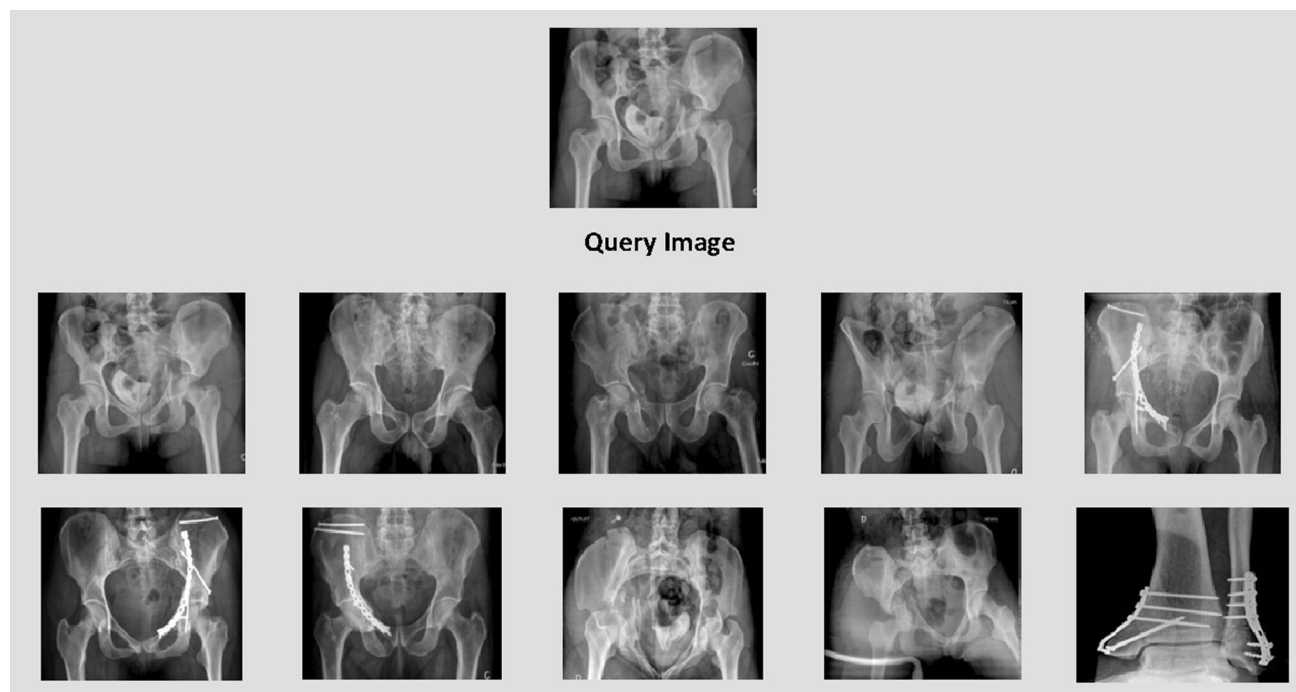
$$R = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in the DB}} \quad (25)$$

The MAP represents the quality of a system based on different levels of recall. If the precision score is 1.0, this means that every image retrieved by a search is relevant, but we do not know whether the search retrieves all the images relevant to the query. If the recall score is 1.0, this means that all relevant images are retrieved by the search, but we do not know whether the number of irrelevant images were also retrieved (Afifi and Ashour 2012).

#### 4.5.3 Evaluation of the image retrieval scheme

We perform some experiments to check the retrieval effectiveness of the proposed method without the relevance feedback process. The objective is to validate the efficiency of the retrieval model by the fusion between different features, which are described above, and compare its performance with other existing state-of-the-art methods in the literature. Some images were selected randomly from each category to test the system and retrieve 10 images similar to the query image. An example of a test is presented, as can be seen from Fig. 3, by selecting a random image from the pelvis class.

The operation of the scheme in a precision–recall curve is depicted as a traditional graph, which delivers a significant outcome after the database is known and was practiced by some past systems. The application of the precision–recall curve enables the evaluation of the proposed scheme. The option of some images at random from



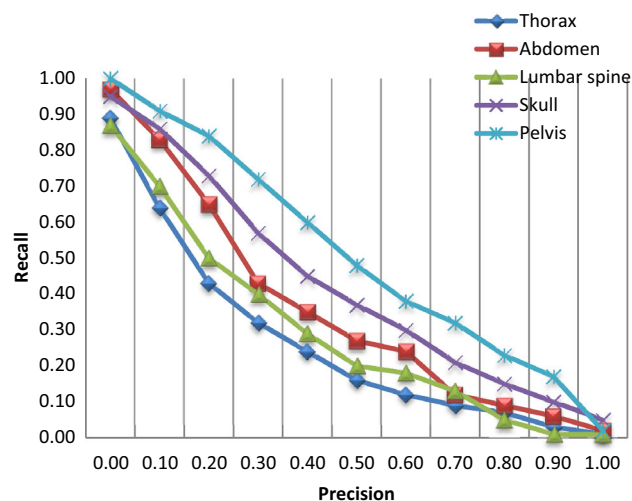
**Fig. 3** Pelvis query and the top 10 retrieved images

each class in the database was used as queries to compute the precision and recall data.

We note that our approach works well on each image in different classes. For every image, the precision of the retrieval solution is achieved by increasing the retrieved images amount. Figure 4 presents the precision–recall graph for query images among different classes.

Furthermore, we investigated the influence of the proposed set of features on our integrated CBIR system performance. One of the testing processes is divided into four phases. In the first phase, the proposed method will use the randomly selected images to retrieve similar images from the database, using the color feature only. In the second phase, the proposed method will retrieve similar images from the database, using the texture feature only. In the third phase, the proposed method will retrieve similar images from the database, using the shape feature only. In the fourth phase, the proposed method will retrieve images similar to the input image according to the color, texture and shape features, according to our methodology. After many tests, it is clear from Table 1 that proposed method works very well when we use both color, texture and shape feature to retrieve images similar to the input image. It shows the experimental results for each category, by calculating the average precision based on the computation of the precision for different tests.

Also, Fig. 5 shows a graph to visualize which method has more accurate results. This figure demonstrates also

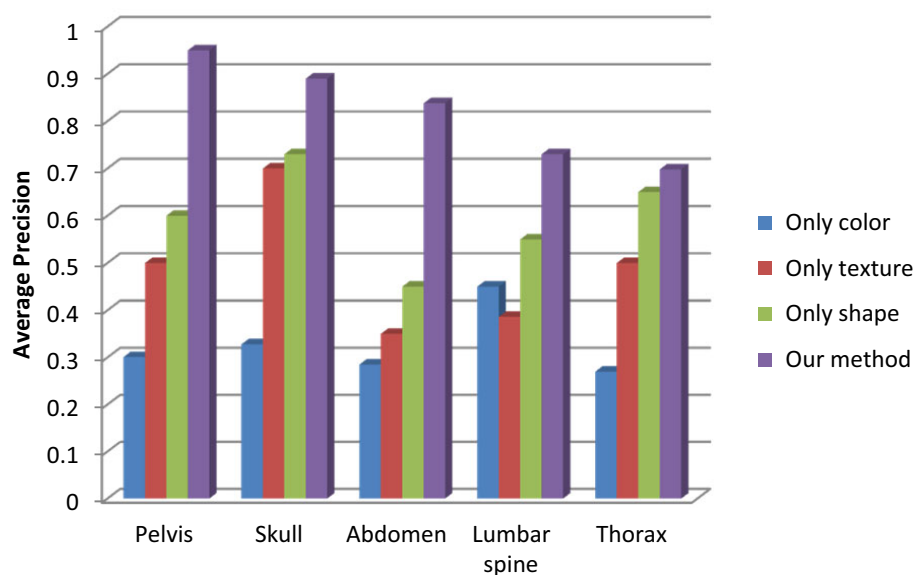


**Fig. 4** Precision–recall curves

**Table 1** Comparison of average precision obtained by different retrieval techniques

	Pelvis	Skull	Abdomen	Lumbar spine	Thorax
Only color	0.301	0.328	0.285	0.450	0.270
Only texture	0.500	0.700	0.350	0.386	0.500
Only shape	0.600	0.730	0.450	0.550	0.650
Our method	0.950	0.890	0.838	0.730	0.698

**Fig. 5** Average precision comparisons between color moments, texture features, shape features, and combination of color, texture and shape



**Table 2** Comparison of average precision obtained by different retrieval techniques

	Pelvis	Skull	Abdomen	Lumbar spine	Thorax
Color + texture	0.800	0.550	0.400	0.550	0.480
Color + shape	0.750	0.750	0.600	0.600	0.600
Shape + texture	0.680	0.780	0.650	0.650	0.700
Our method	0.950	0.890	0.838	0.730	0.698

that using the combination of color, texture and shape feature to represent the image and retrieve images similar to it, has more accuracy compared with only color feature or only texture feature or only shape feature. The results prove that the precision of the system based on all proposed integrated features is higher than for each feature separately.

Another testing process is divided into three phases. In the first phase, the proposed method will use the combination between color and texture features to retrieve similar images from the database. In the second phase, the proposed method will retrieve similar images from the database, using the combination between color and shape features. In the third phase, the proposed method will retrieve similar images from the database using the combination between texture and shape features. We compare these results with results performed by the combination of all features. It is clear, after many tests, from Table 2 that proposed method works also very well when we use both color, texture and shape feature to retrieve images similar to the input image. Figure 6 shows a graph to demonstrate which method has more accurate results.

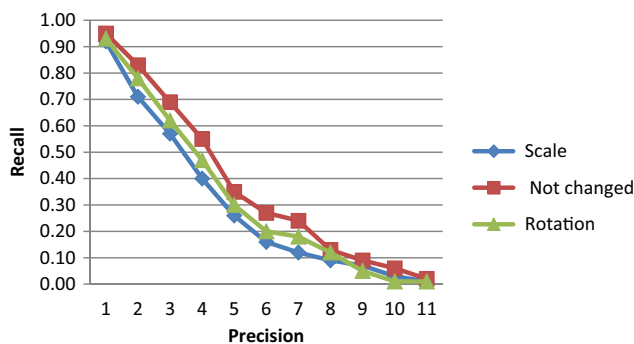
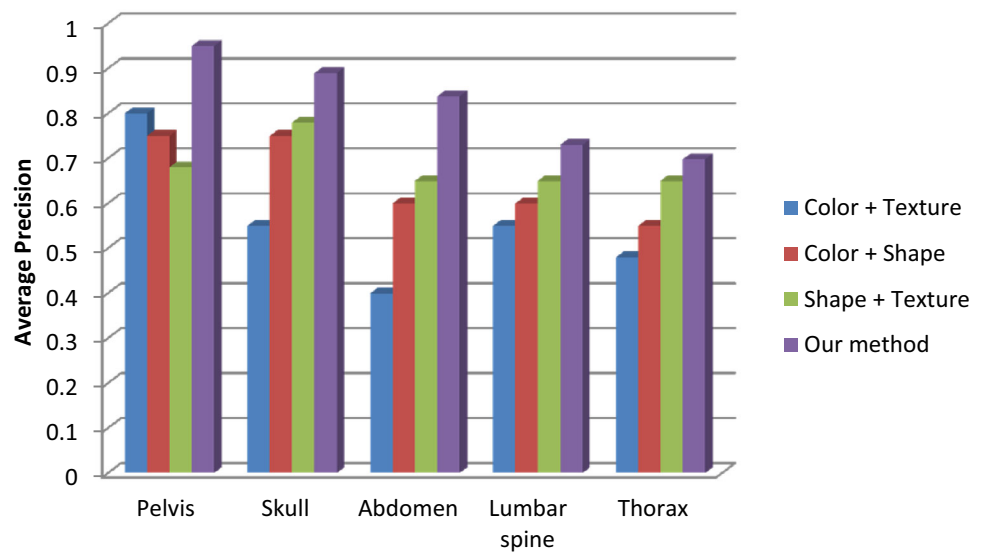
Moreover tests were performed on unchanged (original) images, as well as on modified images, from the database. In these experiments, all query images (changed and original)

were used to search for similar images stored in the databases. To investigate the retrieval of changed images, some query images were changed in terms of scale and rotation. The images were rotated by  $10^\circ$ ,  $45^\circ$ , and  $90^\circ$ . The scale change factors were 0.9 and 1.1. In these experiments, the precision decreased by approximately 6 %. The precision–recall curves for experiments on changed images are shown in Fig. 7.

#### 4.5.4 Comparative performance evaluation

We report experimental results that show the feasibility and utility of the proposed algorithm and compare its performance with four state-of-the-art methods. Ramamurthy et al. (2012), Wu and Tai (2009), Ramamurthy and Chandran (2012) and Bueno et al. (2010). To simulate the practical situation, the sequence of query images used in all the experiments is always, as for the other experiments, generated at random. We use the average precision (AP) and the mean average precision (MAP), as mentioned in Sect. 4.5.1, as our retrieval performance metrics for the comparison between methods. Table 3 shows that our proposed system performance is better than the other systems (Ramamurthy et al. (2012) and Ramamurthy and Chandran (2012)) for all classes. Figure 8 shows the comparison of the proposed system with other systems. It



**Fig. 6** Features comparison chart with the proposed method**Fig. 7** Precision versus recall curve for original and modified images

shows the average precision value of each system for each class via the value from Table 3 by a vertical bar.

According to results in Table 4, we see also that the image retrieval performance by the proposed method is competitive with the all other tested methods, in terms of mean average precision (MAP).

The comparison is carried out by an implementation of each schema and method (systems to compare with our schema) by using MATLAB 7.0 and records different results from tests, performed on our image dataset. We can be sure that there is no difference between founded results and published results. So, we can be sure, now, that our proposed system is superior compared to other existing systems.

#### 4.5.5 Evaluation of the image retrieval scheme with relevance feedback

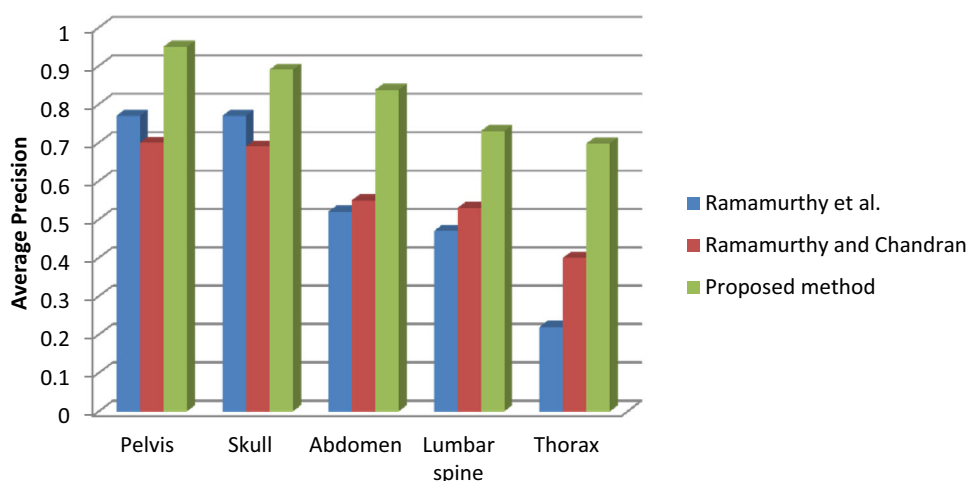
As mentioned above, the relevance feedback method works in an iterative fashion. At every iteration, a set of  $n$  images are shown. The user is then given the chance to classify them as relevant or non-relevant, and his/her selections are given as an input to the search algorithm. Finally, the algorithm computes a new set of  $k$  images which are used as the input for the next iteration. The feedback processes were performed 5 times, and number of returned images is 10. For completeness, a final experiment aims to evaluate the effectiveness and the performance of our method when it is applied in all iterations of the relevance feedback mechanism. In Fig. 9, we present an example of a scenario of judgment and labeling (relevant/irrelevant) performed by a user related to an input query image. Different results for this experiment are shown in Figs. 10 and 11.

Figures 10 and 11 shows the retrieval results, in terms of the AP, using the relevance feedback algorithm after the five feedback iterations. It is not difficult to see that, compared with the retrieval results without relevance feedback, the performance of our relevance feedback retrieval system is improved. However, as the number of the feedback iteration increases, the performance of our relevance feedback retrieval system becomes better and better, and it is more effective.

**Table 3** Comparison among the average precision of the proposed method, Ramamurthy et al. and Ramamurthy and Chandran

	Pelvis	Skull	Abdomen	Lumbar spine	Thorax
Ramamurthy et al.	0.770	0.770	0.520	0.470	0.220
Ramamurthy and Chandran	0.700	0.690	0.550	0.530	0.400
Proposed method	0.950	0.890	0.838	0.730	0.698

**Fig. 8** Average precision comparisons between the proposed system and existing systems



**Table 4** Average precision comparison between the proposed system and existing systems

	MAP
Ramamurthy et al.	0.55
Wu and Tai	0.51
Bueno et al.	0.75
Ramamurthy and Chandran	0.57
Our method	0.82

Figure 11 describes also detailed comparison of the average retrieval precision for different classes. From this figure, we observed that there is a rapid increase in retrieval performance with each feedback iteration. Retrieval performance is improved from 69.8 to 82 % from first iteration to the fifth iteration for the thorax class as an example. Results are also tabulated in Table 5. Users (physicians or patients) could use as many relevance feedback cycles as needed to improve the accuracy and the certainty of their diagnosis.

## 5 The evaluation of our methodology's impact in the decision-making process

After the navigation through our social network site and using the integrated CBMIR system, the specialists can judge the impact of our social network site in generally, and the integrated CBMIR system in particular.

### 5.1 Evaluation methodology's description

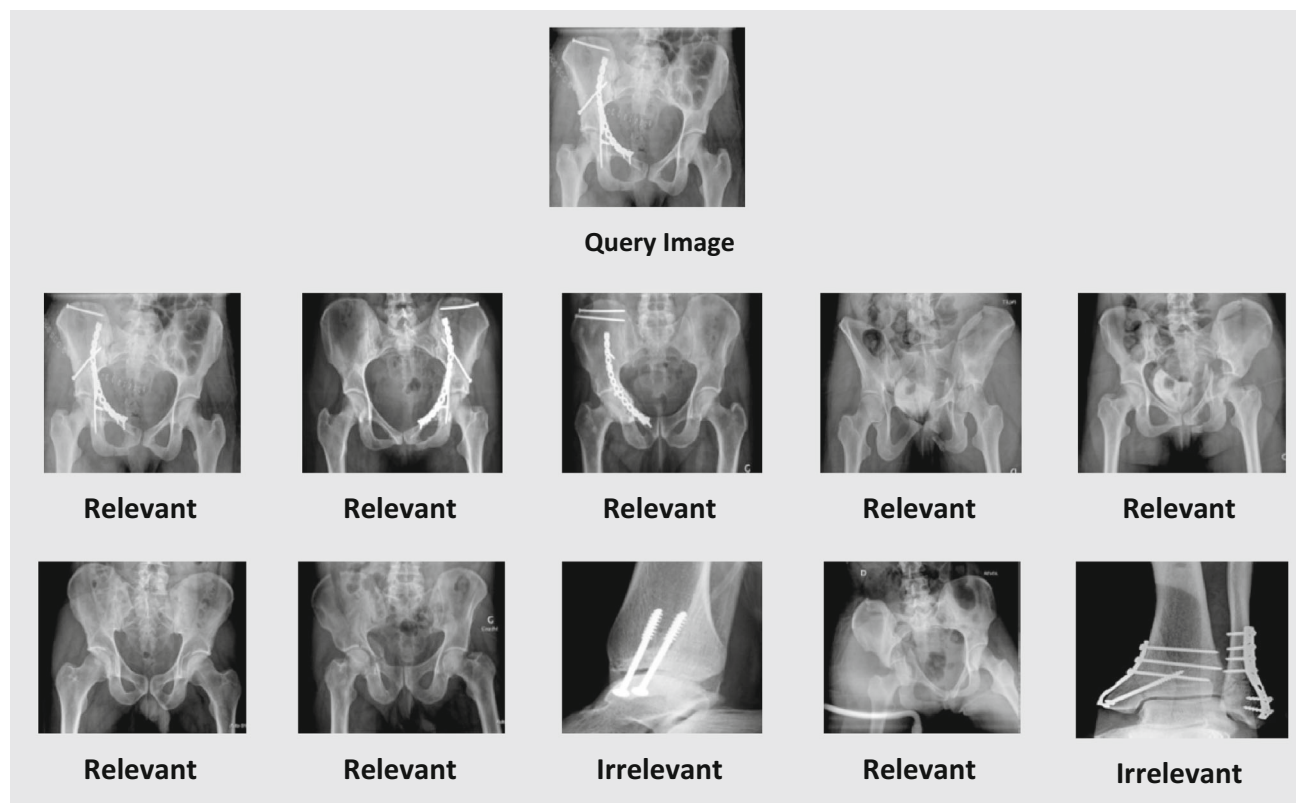
Specialists are asked to fill a questionnaire. Its aim was to identify relevant information on their actual needs. Analyses were performed to evaluate, specially, the satisfaction of the specialists when using the integrated CBMIR system

in clinical practice. We consider that the judgment of specialists as a reference for patient's judgement. Basically, the questionnaire, as a semi-structured one, was composed of five personal questions, and 18 regarding the system assessment. The latter were elaborated based on usability and functional suitability features, which allow the evaluation how easy it was to use the medical social network and the CBMIR system (user interface esthetics, learn ability, operability). Functional suitability features evidenced the set of functions that respond to the internal and external demands of the system (subdivided into functional adequacy and functional correctness), as listed in ISO/IEC 25010 (2011). Moreover, a set of features was prepared with the aim of complementing ISO/IEC 25010 features according to this study approach (general views and opinions, new recommendations and requirements for system improvement and/or to solve limitations). In order to perform the analysis of our methodology, we conducted experiments with the support of 10 specialists, including radiologists and resident physicians from Charles Nicolle Hospital in Tunis. Twenty (20) images were used as query centers in such process. Each specialist analyzed five images on average, and each image was analyzed by two different specialists on average.

### 5.2 Results and interpretation

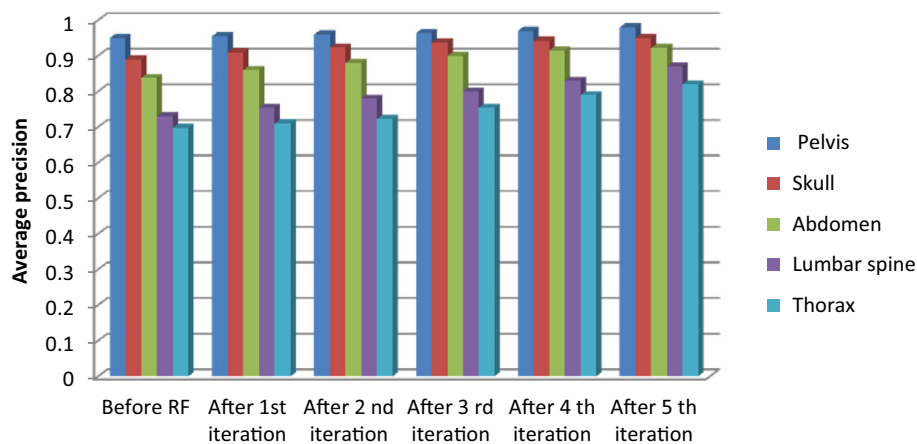
Finally, the specialists answered the questionnaire described above. The questionnaire showed that:

- 100 % found it easy to navigate through the medical social network site;
- 100 % had some previous knowledge of these sites;
- 100 % found that it is important to use a social network site dedicated to health care community (physicians and patients);



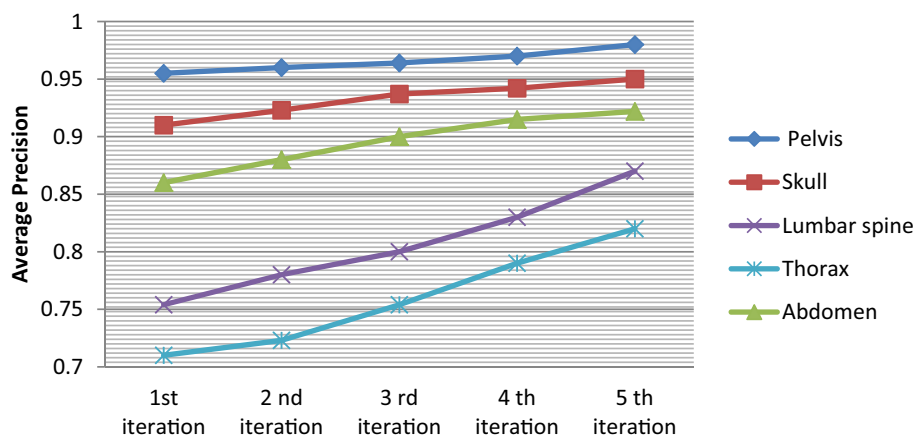
**Fig. 9** An example of labeling performed by a user

**Fig. 10** Average precision comparison before and after relevance feedback



- 90 % believes in the need of our site in the clinical routine;
- 90 % found that it is important to integrate a CBMIR system through the medical social network site;
- 80 % found it easy to operate the CBMIR system;
- 40 % had some previous knowledge of this type of system;
- 90 % believes that the CBMIR system helps in the diagnostic;
- 90 % are satisfied about the interaction with the system to identify the relevant among the irrelevant images (about the RF method);
- 90 % indicates that our system can help to train medical students and radiologists at Medical School;
- 90 % believes in the need of our system in the clinical routine.

Based on these findings, we can say that specialists have high expectations when using the medical social network

**Fig. 11** Relationship between average precision and number of feedback iterations**Table 5** Average precision value on each feedback iteration

	Before RF	After 1st iteration	After 2nd iteration	After 3rd iteration	After 4th iteration	After 5th iteration
Pelvis	0.950	0.955	0.960	0.964	0.970	0.980
Skull	<b>0.890</b>	<b>0.910</b>	<b>0.923</b>	<b>0.937</b>	<b>0.942</b>	<b>0.950</b>
Abdomen	<b>0.838</b>	<b>0.860</b>	<b>0.880</b>	<b>0.900</b>	<b>0.915</b>	<b>0.922</b>
Lumbar spine	<b>0.730</b>	<b>0.754</b>	<b>0.780</b>	<b>0.800</b>	<b>0.830</b>	<b>0.870</b>
Thorax	0.698	0.710	0.723	0.754	0.790	0.820

site. We can say, also, that specialists have high expectations when using the integrated CBMIR system in clinical practice as a tool that makes relevant information and helps them in the decision making when they use the social network site. They need this system in the diagnostic aid, in order to establish the diagnosis exactly from existing comments and posts, so our image retrieval scheme can be directly used in the diagnosis process. They need it also for medical teaching by comparing similar cases and their particularities or comparing similar cases with different diagnosis and for medical research by finding, for example, certain types of images to be included in a study. With the results obtained, interesting facts can be analyzed regarding the application of the CBMIR system in clinical practice and the importance of calibrating the system correctly.

## 6 Conclusion and future work

Social image indexing and retrieval in the large databases of the social networks advanced the challenges to form a new problem that needs special handling. Content-based image retrieval presents an interesting research topic to manipulate medical images. Further, medical image indexing and retrieval is a challenging task on its own, and thus different solutions have been proposed, trying to address different angles of the problem. For this reason, the integration of CBIR systems through social networks

becomes important to handle the huge volume of uploaded images by different users. So, we proposed in this paper a medical social network destined to patients and physicians, in order to minimize medical errors by fostering collaboration between them. We, also, proposed the integration of a CBMIR system through our medical social network site. We presented an effective image retrieval system based on the combination and the fusion between color (gray level), texture and shape features. Therefore, color, shape and texture features are the most important features that were considered when developing a CBIR method. For the color content extraction, a well-known and powerful technique like the gray-level histogram and gray-level moment can be used, the co-occurrence matrix for texture content extraction was used, and statistical region descriptors for shape content extraction was used. Experimentations were performed by a medical image database from Charles Nicolle Hospital, in Tunis. The comparability of our proposed scheme with the other existing CBIR systems demonstrates the usage of the same database for system evaluation. Furthermore, it also proves that our system is better off than the other systems while the outcomes are satisfactory. We implemented, in our work, a relevance feedback approach to improve the CMBIR method quality, dealing with the “semantic gap problem.” By using the relevance feedback technique, we maximized the accuracy of the integrated CBMIR method, guided by the user’s interaction. The experiments showed that the proposed method is



effective in improving the query precision, contributing to bridge the semantic gap and achieving improvement in the query results of up to 20 %. We presented, also, a study about the evaluation of our social network site in generally, and the integrated CBMIR system in particular in order to analyze the impact of our methodology in a real clinical environment. Different results obtained by the questionnaire show us that a CBMIR method is useful in real environments to assist specialists, during the decision-making process, and the integration of these systems through medical social network site is very successful to improve diagnostic. Finally, different analysis presented, also, interesting results about the acceptance and viability of using our medical social network site.

Future work will focus on presenting an analysis approach for existing comments. The objective is to extract keywords from comments must be used to give an overview, a summary, of what exist on comments and to annotate images in order to facilitate the search task. The indexation is used to improve the image search from textual queries.

## References

- Afifi AJ, Ashour WM (2012) Content-based image retrieval using invariant color and texture features. In: International conference on digital image computing techniques and applications (DICTA). IEEE, Fremantle, WA, pp 1–6
- Agma J, Traina M, André G, Balan R, Bortolotti LM, Traina C Jr (2007) Content-based image retrieval using approximate shape of objects. In: The 17th IEEE symposium on computer-based medical systems (CBMS'07), pp 91–96
- Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C et al (2003) Automated storage and retrieval of thin section CT images to assist diagnosis: system description and preliminary assessment. *Radiology* 228(1):265–270
- Akgül C, Rubin D, Napel S, Beaulieu C, Greenspan H, Acar B (2011) Content-based image retrieval in radiology: current status and future directions. *J Digit Imaging* 24:208–222
- Almansoori W, Zarour O, Jarada TN, Karampales P, Rokne J, Alhaji R (2011) Applications of social network construction and analysis in the medical referral process. In: Proceedings of the 2011 IEEE ninth international conference on dependable, autonomic and secure computing (DASC '11)
- Antani S, Long LR, Thoma GR, Lee DJ (2003) Evaluation of shape indexing methods for content-based retrieval of X-ray images. In: Yeung MM, Lienhart RW, Li CS (eds) Proceedings of SPIE 5021, pp 405–416
- Antani S, Lee D, Long LR, Thoma GR (2004) Evaluation of shape similarity measurement methods for spine X-ray images. *J Vis Commun Image Represent* 15(3):285–302
- Arevalillo-Herráez M, Ferri FJ, Moreno-Picot S (2013) A hybrid multi-objective optimization algorithm for content based image retrieval. *Appl Soft Comput* 13:4358–4369
- Ashish O, Manpreet S (2012) Content based image retrieval system for medical databases (CBIR-MD)—lucratively tested on endoscopy, dental and skull images. *IJCSI Int J Comput Sci Issues* 9(1):300–306
- Bach JR, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R et al (1996) Virage image search engine: an open framework for image management. In: Sethi IK, Jain RC (eds) Proceedings of SPIE, vol 2670, no 1, pp 76–87
- Bhattacharjee N, Parekh R (2011) Skin texture analysis for medical diagnosis. In: The international conference on communication, computing and security. New York, pp 301–306
- Bueno R, Ribeiro MX, Traina AJM (2010) Improving medical image retrieval through multi-descriptor similarity functions and association rules. In: IEEE 23rd international symposium on computer-based medical systems (CBMS), pp 309–314
- Bugatti PH, Ponciano-Silva M, Agma J, Traina M, Traina C Jr, Marques P (2009) Content-based retrieval of medical images: from context to perception. In: 22nd IEEE international symposium on computer-based medical systems (CBMS), pp 1–8
- Bugatti PH, Ribeiro MX, Traina JM, Traina C Jr (2011) Feature selection guided by perception in medical CBIR systems. In: First IEEE international conference on healthcare informatics, imaging and systems biology, pp 323–330
- Chechik G, Sharma V, Shalit U, Bengio S (2010) Large scale online learning of image similarity through ranking. *J Mach Learn Res* 11:1109–1135
- Cox JJ, Miller ML, Minka TP, Papathomas TV, Yianilos PN (2000) The Bayesian image retrieval system, PicHunter: theory, implementation and psychophysical experiments. *IEEE Tran Image Process* 9(1):20–37
- Daniel RG, Liza SR, Jennifer LK (2013) Dangers and opportunities for social media in medicine. *Clin Obstet Gynecol*. doi:10.1097/GRF.0b013e318297dc38
- Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. *ACM Comput Surv* 40(2):5:1–5:60
- De Oliveira JEE, Machado AMC, Chavez GC, Lopes APB, Deserno TM, De Araujo AA (2010) Mammosys: a content-based image retrieval system using breast density patterns. *Comput Methods Programs Biomed* 99(3):289–297
- Doganay S (2014) Healthcare social networks: new choices for doctors, patients. <http://www.informationweek.com/healthcare/patient-tools/healthcare-social-networks-new-choices-for-doctors-patients/d/d-id/1234884>
- Doulamis N, Doulamis A (2006) Evaluation of relevance feedback schemes in content-based in retrieval systems. *Signal Process Image Commun* 21(4):334–357
- Fazal M, Baharum B (2013) Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain. *J King Saud Univ Comput Inf Sci* 25:207–218
- Feldman DL (2012) Medical social media networks: communicating across the virtual highway. *Q J Health Care Pract Risk Manag Infocus* 18(1):2–5
- Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B et al (1995) Query by image and video content: the QBIC system. *Computer* 28(9):23–32
- Franklin V, Greene S (2007) Sweet talk: a text messaging support system. *J Diabetes Nurs* 11(1):22–26
- Goh K-S, Chang EY, Li B (2005) Using one-class and two-class SVMs for multiclass image annotation. *IEEE Trans Knowl Data Eng* 17(10):1333–1346
- Gong J, Sun S (2011) Individual doctor recommendation model on medical social network. In: Proceedings of the 7th international conference on advanced data mining and applications (ADMA'11)
- Greenspan H (2007) Medical image categorization and retrieval. For PACS using the GMM-KL framework. *IEEE Trans Inf Technol BioMed* 11:190–202
- Grenier C (2003) The role of intermediate subject to understand the structuring of an organizational network of actors and

- technology—case of a care network. In: Proceedings of the 9th conference of the association information and management, Grenoble
- Güld MO, Thies C, Fischer B, Lehmann TM (2007) A generic concept for the implementation of medical image retrieval systems. *Int Med Inform* 76(2–3):252–259
- Harishchandra H, Mushigeri S, Niranjana UC (2014) Medical image retrieval—performance comparison using texture features. *Int J Eng Res Dev* 9(9):30–34
- Hoi SCHH, Jin R, Zhu JK, Lyu MR (2009) Semi-supervised SVM batch mode active learning and its applications to image retrieval. *ACM Trans Inf Syst* 27(3):1–29
- Hsu W, Antani S, Long LR, Neve L, Thoma GR (2009) SPIRS: a web based image retrieval system for large biomedical databases. *Int J Med Inform* 78(1):13–24
- Hu W, Xie N, Li L, Zeng X (2011) A survey on visual content-based video indexing and retrieval. *IEEE Trans Syst Man Cybern C Appl Rev* 41(6):797–819
- Iakovidis D, Pelekis N, Kotsifakos E, Kopanakis I, Karanikas H, Theodoridis Y (2009) A pattern similarity scheme for medical image retrieval. *IEEE Trans Inf Technol Biomed* 13(4):442–509
- ISO, IEC 25010 (2011) Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuARE)—System and software quality models. International Standards Organization, Geneva
- John C, Kazunori O (2012) A comparative study of similarity measures for content-based medical image retrieval. <http://ceur-ws.org/Vol-1178/CLEF2012wn-ImageCLEF-CollinsEt2012.pdf>
- Keyers D, Dahmen J, Ney H, Wein BB, Lehmann TM (2003) Statistica framework for model-based image retrieval in medical applications. *J Electron Imaging* 12(1):59–68
- Komali A et al (2012) 3D color feature extraction in content-based image retrieval. *Int J Soft Comput Eng (IJSCE)* 2(3):560–563
- Kumar A, Kim J, Cai W, Fulham M, Feng D (2013) Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. *J Digit Imaging*. doi:10.1007/s10278-013-9619-2
- Lee DJ, Antani S, Long LR (2003) Similarity measurement using polygon curve representation and Fourier descriptors for shape-based vertebral image retrieval. In: Sonka M, Fitzpatrick JM (eds) Proceedings of SPIE, vol 5032, pp 1283–1291
- Lee DJ, Antani S, Chang Y, Gledhill K, Long LR, Christensen P (2009) CBIR of spine X-ray images on intervertebral disc space and shape profiles using feature ranking and voting consensus. *Data Knowl Eng* 68(12):1359–1369
- Lew MS, Sebe N, Djeraba C, Jain R (2006) Content-based multimedia information retrieval: state of the art and challenges. *ACM Trans Multimed Comput Commun Appl* 2(1):1–19
- Li J (2014) Data protection in healthcare social networks. *J IEEE Softw* 31(1):46–53
- Liu Y, Zhang D, Lu G, Ma W-Y (2007) A survey of content based image retrieval with high-level semantics. *Pattern Recognit Lett* 40(1):262–282
- MacArthur SD, Brodley CE, Shyu CR (2000) Relevance feedback decision trees in content-based image retrieval. In: Proceedings of the IEEE work-shop content-based access of image and video libraries, pp 68–72
- Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, New York
- Messaoudi A, Bouslimi R, Akaichi J (2013) Indexing medical images based on collaborative experts reports. *Int J Comput Appl* 70(5):1–9
- Mezaris V, Kompatsiaris I, Strintzis MG (2005) An ontology approach to object based image retrieval. In: Proceedings of the international conference on image processing, pp 511–514
- Müller H, Rosset A, Garcia A, Vallée JP, Geissbühler A (2005) Benefits of content-based visual data access in radiology. *Radiographics* 25(3):849–858
- Nandagopalan S, Adiga BS, Deepak N (2008) A universal model for content-based image retrieval. *World Acad Sci Eng Technol* 46:644–647
- Napel SA, Beaulieu CF, Rodriguez C, Cui J, Xu J, Gupta A et al (2010) Automated retrieval of CT images of liver lesions on the basis of image similarity: method and preliminary results. *Radiology* 256(1):243–252
- Patil PB, Kokare MB (2011) Relevance feedback in content based image retrieval: a review. *J Appl Comput Sci Math* 10:41–47
- Pentland A, Picard RW, Sclaroff S (1996) Photobook: content-based manipulation of image databases. *Int J Comput Vis* 18:233–254
- Qian X, Tagare HD, Fulbright RK, Long R, Antani S (2010) Optimal embedding for shape indexing in medical image databases. *Med Image Anal* 14(3):243–254
- Rajakumar K, Muttan S (2013) MRI image retrieval using Wavelet with Mahalanobis distance measurement. *J Electr Eng Technol* 8(5):1188–1193
- Ramamurthy B, Chandran KR (2012) Content based medical image retrieval with texture content using gray level co-occurrence matrix and k-means clustering algorithms. *J Comput Sci* 8(7):1070–1076
- Ramamurthy B, Chandran KR, Meenakshi VR, Shilpa V (2012) CBMIR: content based medical image retrieval system using texture and intensity for dental images. In: International conference eco-friendly computing and communication systems, ICECCS, vol 305, pp 125–134
- Rocchio JJ (1971) Relevance feedback in information retrieval: SMART retrieval system, 1st edn. Prentice Hall, Upper Saddle River, pp 323–341
- Rui Y, Huang TS, Chang SF (1999) Image retrieval: current techniques, promising directions, and open issues. *J Vis Commun Image Represent* 10(1):39–62
- Selvarani G, Annadurai S (2007) Medical image retrieval by combining low level features and DICOM features. In: International IEEE conference on computational intelligence and multimedia applications, pp 587–589
- Seng WC, Mirisae SH (2009) A content-based retrieval system for blood cells images. In: International IEEE conference on future computer and communication, pp 412–415
- Shanmugapriya N, Nallusamy R (2014) A new content based image retrieval system using GMM and relevance feedback. *J Comput Sci* 10(2):330–340
- Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS (1999) ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. *Comput Vision Image Underst* 75(1–2):111–132
- Sidong L, Lei J, Weidong C, Lingfeng W, Eberl S, Fulham MJ, Dagan F (2010) Localized multiscale texture based retrieval of neurological image. In: IEEE 23rd international symposium on computer-based medical systems (CBMS), pp 243–248
- Singh P, Singh S, Kaur G (2009) Efficient techniques used in CBMIR for medical image retrievals. *Proc World Acad Sci Eng Technol* 38:434–437
- Singh J, Kaleka JS, Sharma R (2012) Different approaches of CBIR techniques. *Int J Comput Distributed Syst* 1(2):76–78
- Smeulders A, Worring M, Santini S, Gupta A, Jain R (2000) Content based image retrieval at the end of the early years. *IEEE Trans Pattern Anal Mach Intell* 22(12):1349–1380
- Song Y (2012) Image Analysis for Automatic Phenotyping Measurements. Biomathematics and Statistics Scotland (BioSS). Wageningen
- Stojmenovic M, Nayak A (2008) Measuring the related properties of linearity and elongation of point sets. In: 13th Iberoamerican

- congress on pattern recognition, CIARP 2008. Havana, Cuba, pp 102–111
- Stojmenovic M, Jevremovic A, Nayak A (2013) Fast iris detection via shape based circularity. In: 8th IEEE conference on industrial electronics and applications (ICIEA), pp 747–752
- Su Z, Zhang H, Li S, Ma S (2003) Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. *IEEE Trans Image Process* 12(8):924–936
- Swarnambiga A, Vasuki SM (2013) Distance measures for medical image retrieval. *Int J Imaging Syst Technol* 23:9–21
- Tieu K, Viola P (2003) Boosting image retrieval. In: Proceedings of the IEEE conference on computer vision pattern recognition, pp 228–235
- Wang M, Hua XS (2011) Active learning in multimedia annotation and retrieval: a survey. *ACM Trans Intell Syst Technol* 2(2):10–31
- Wang M, Li H, Tao D, Lu K, Wu X (2012) Multimodal graph-based reranking for web image search. *IEEE Trans Image Process* 21(11):4649–4661
- Wang X-Y, Li Y-W, Yang H-Y, Chen J-W (2014) An image retrieval scheme with relevance feedback using feature reconstruction and SVM reclassification. *Neurocomputing* 127:214–230
- Wu C, Tai X (2009) Application of gray level variation statistic in gastroscopic image retrieval. In: Eighth IEEE/ACIS international conference on computer and information science, pp 342–346
- Xiang-Yang W, Bei-Bei Z, Hong-Ying Y (2012) Content-based image retrieval by integrating color and texture features. *Multimed Tools Appl* 68(3):545–569
- Xie Y, Chen Z, Cheng Y, Zhang K, Agrawal A, Liao WK, Choudhary A (2013) Detecting and tracking disease outbreaks by mining social media data. In: Proceedings of the twenty-third international joint conference on artificial intelligence (IJCAI'13)
- Xin J, Jin JS (2004) Relevance feedback for content-based image retrieval using Bayesian network. In: Proceedings of the pan-sydney area workshop on visual information processing (VIP '05). Australian Computer Society, Inc Darlinghurst Australia, pp 91–94
- Xu X, Lee DJ, Antani S, Long L (2008) A spine X-ray image retrieval system using partial shape matching. *IEEE Trans Inf Technol Biomed* 12(1):100–108
- Yogapriya J, Ila V (2013) An integrated framework based on texture features, cuckoo search and relevance vector machine for medical image retrieval system. *Am J Appl Sci* 10(11):1398–1412
- Yue J, Li Z, Liu L, Fu Z (2010) Content-based image retrieval using color and texture fused features. *Math Comput Model* 54:1121–1127
- Zeyad SY, Dzulkifli M, Tanzila S, Mohammed HA, Amjad R, Al-R Mznah, Al-D Abdullah (2014) Content-based image retrieval using PSO and k-means clustering algorithm. *Arab J Geosci* 8(8):6211–6224. doi:[10.1007/s12517-014-1584-7](https://doi.org/10.1007/s12517-014-1584-7)
- Zhang D, Lu G (2004) Review of shape representation and description techniques. *Pattern Recognit* 37(1):1–19
- Zhang J, Ye L (2010) Series feature aggregation for content-based image retrieval. *Comput Electr Eng* 36(4):691–701
- Zhang G, Ma ZM, He Y, Zhao T (2008) Texture characteristic extraction for dominant directions in content-based medical image retrieval. In: IEEE international conference on biomedical engineering and informatics (BMEI), pp 253–257
- Zhang D, Islam MdM, Lu G (2012) A review on automatic image annotation techniques. *Pattern Recognit* 45:346–362
- Zhi W, Wenwu Z, Peng C, Lifeng S, Shiqiang Y (2013) Social media recommendation. *Soc Media Retr Comput Commun Netw*. doi:[10.1007/978-1-4471-4555-43](https://doi.org/10.1007/978-1-4471-4555-43)