Expert Systems With Applications 100 (2018) 224–233

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval

Amin Khatami[a,*], Morteza Babaie[b,c], H.R. Tizhoosh[c], Abbas Khosravi[a], Thanh Nguyen[a], Saeid Nahavandi[a]

[a] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
[b] Department of Mathematics and Computer Science, Amirkabir University of Technology, Iran
[c] KIMIA Lab at the University of Waterloo, Canada

ARTICLE INFO

Article history:
Received 29 July 2017
Revised 10 January 2018
Accepted 31 January 2018
Available online 6 February 2018

Keywords:
Content-based image retrieval
CBIR
Medical imaging
Deep learning
Radon

ABSTRACT

Closing the semantic gap in medical image analysis is critical. Access to large-scale datasets might help to narrow the gap. However, large and balanced datasets may not always be available. On the other side, retrieving similar images from an archive is a valuable task to facilitate better diagnosis. In this work, we concentrate on forming a search space, consisting of the most similar images for a given query, to be used for a similarity-based search technique in a retrieval system. We propose a two-step hierarchical shrinking search space when local binary patterns are used. Transfer learning via convolutional neural networks is utilized for the first stage of search space shrinking, followed by creating a selection pool using Radon transform for further reduction. The difference between two orthogonal Radon projections is considered in the selection pool to extract more information. The IRMA dataset, from ImageCLEF initiative, containing 14,400 X-ray images, is used to validate the proposed scheme. We report a total IRMA error of 168.05 (or 90.30% accuracy) which is the best result compared with existing methods in the literature for this dataset when real-time processing is considered.

1. Introduction

The main challenge in image search is to find the most relevant information among a set of images for a given query image. In medical domain, searching for similar cases in terms of the same anatomy and/or pathology can serve as “virtual peer review” for diagnostic purposes. Retrieving similar cases along with associated information and reports from repositories when the query is being treated by general practitioners and radiologists can establish a new level of comparative diagnostic which is presently absent and can immensely contribute to more accurate and robust diagnosis (Khatami et al., 2017d; 2017e; Zhang, Song, Cai, Liu, Liu, Pujol, Kikinis, Xia, Fulham, & Feng, 2016; Zhang, Huang, Li, & Metaxas, 2012).

Nowadays, the number of digital images generated by medical imaging devices is raising enormously. Managing and analyzing these large-scale repositories are becoming significantly complicated. If the size of the search space is too large making the image information not retrievable, the *semantic gap* will remain as an insurmountable challenge (Smeulders, Worring, Santini, Gupta, & Jain, 2000). The semantic gap is the amount of image information lost to a numerical representation of some features.

To make content-based image retrieval (CBIR) feasible, a robust searching scheme to find similar cases in large archives is needed. To overcome this challenge, hand-crafted features, deep features, dictionary approach, and other algorithms are used (Avni, Greenspan, Konen, Sharon, & Goldberger, 2011; Greenspan & Pinhas, 2007; Khatami, Khosravi, Lim, & Nahavandi, 2016; Khatami, Khosravi, Nguyen, Lim, & Nahavandi, 2017b; Kumar, Kim, Cai, Fulham, & Feng, 2013; Nanni, Brahnam, & Lumini, 2010; Rahman, Bhattacharya, & Desai, 2007; Xu, Lee, Antani, & Long, 2008).

There are two distinct strategies for image retrieval in a CBIR system. The first one is retrieving particular organs in particular modalities such as retrieving malignant lung nodules (Khatami et al., 2017c; Pan, Qiang, Yuan, & Wu, 2016) and liver lesions in CT Images (Napel et al., 2010), and chest structures from X-ray images (Shin et al., 2016). The second strategy concentrates on global similarity search in heterogeneous archives to identify and

* Corresponding author.
E-mail addresses: amin.khatami@deakin.edu.au, skhatami@deakin.edu.au (A. Khatami), mbabaie@uwaterloo.ca (M. Babaie), tizhoosh@uwaterloo.ca (H.R. Tizhoosh), abbas.khosravi@deakin.edu.au (A. Khosravi), thanh.nguyen@deakin.edu.au (T. Nguyen), saeid.nahavandi@deakin.edu.au (S. Nahavandi).

https://doi.org/10.1016/j.eswa.2018.01.056
0957-4174/

retrieve similar cases (Baâzaoui, Barhoumi, Ahmed, & Zagrouba, 2018; Greenspan & Pinhas, 2007; Seo, 2007). We follow the second strategy to propose a robust CBIR system which is based on two sequential procedures for shrinking the search space using deep networks and local similarity-based search techniques. However, utilizing deep networks for image retrieval in medicine is a challenging task due to two factors: (1) lack of large and "labelled" datasets for training, and (2) naturally existence of case "imbalance" in the medical domain.

Applying robust similarity search on properly shrunk search space, which is expected to make the semantic gap smaller, is a key to design accurate retrieval systems. However, utilizing high-dimensional feature spaces and the choice of reliable distance metric should be taken into account in order to produce meaningful results (Aggarwal, Hinneburg, & Keim, 2001; Alonso & Contreras, 2016; Hinneburg, Aggarwal, & Keim, 2000; Müller, Michoux, Bandon, & Geissbuhler, 2004).

Motivated by achievements of deep learning in computer vision and the applicability and practicability of Radon transform in medical domain, we utilize transfer learning via a convolutional neural network (CNN), as well as a Radon-based selection pool to sequentially shrink the search space. The smaller the search space with several right candidates, the better performance achieved by using the descriptors such as local binary patterns (LBP).

The contributions of this research are started with operating on a strongly imbalanced dataset and delivering the best accuracy and performance reported in literature so far. As well, to the best of our knowledge, this is the first work which creates a feature vector based on a classification-based shrunk search space for further shrinking. This contribution significantly improves the performance, as seen later. We also propose a two-step sequential shrinking search space, using a CNN and Radon transform, resulting in improved retrieval accuracy.

The rest of the paper is organized as following: Literature review is presented in Section 2. Descriptions of the proposed model are presented in Section 3. Section 4 explains the experimental results along with the analysis and discussions. A comprehensive performance comparison is also reported in Section 4, followed by concluding remarks in Section 5.

2. Related works

Several essential stages should be followed to obtain the similarity among features, resulted in a robust CBIR system: (1) *Content description:* the features of color, shape, texture, and so on should be extracted from images, (2) *Feature vectors:* an integrated feature vector should be properly assembled describing the information of the query and the images in datasets. Note that efficiency is important in this stage, and (3) *Similarity measure:* the metrics calculating the similarity among the feature vectors is paramount.

Texture descriptors are commonly used in medical image retrieval systems (Junior, Delgado, Gonçalves, & Nunes, 2009; Vipparthi & Nagar, 2014). As shown in Kashif, Deserno, Haak, and Jonas (2016); Sargent, Chen, Tsai, Wang, and Koppel (2009), it seems that keypoint-based descriptors such as SIFT, SURF, and ORB are not able to generate reliable feature points for some types of medical images. Dense sampling methods such as LBP are well-known in retrieval domain. Apparently, a thorough investigation is often required to obtain the best and efficient feature vectors representing images (Babaie et al., 2017a; Brahnam, Jain, Nanni, Lumini et al., 2014; Pietikäinen, Hadid, Zhao, & Ahonen, 2011). Global features are also widely used in medical image retrieval (Kumar et al., 2013). Radon transformation is mostly a global descriptor which extracts information of images from different directions. This transformation is widely utilized in medical domain due to easy implementation and efficient matching (Babaie, Tizhoosh, Zhu, & Shiri, 2017b; Clack & Defrise, 1994; Metz & Pan, 1995; Weisi et al., 2011). Moreover, Radon-based features may result in an efficient retrieval system by creating a short-length feature vector, which is in contrast to the aforementioned descriptors (Tizhoosh, 2015).

Image retrieval in medical application (IRMA) benchmark, (Lehmann et al., 2005; 2004b), explained in experimental results section, is an quite interesting medical dataset which consists of X-ray images of different body parts for patients of different age and gender (see Fig. 1). The images are assigned *IRMA codes* for benchmarking. As mentioned in Tommasi, Caputo, Welter, Güld, and Deserno (2009), Khatami et al. (2017a), Khatami, Babaie, Khosravi, Tizhoosh, and Nahavandi (2018) and Liu, Tizhoosh, and Kofman (2016), many studies have been performed on this benchmark. Radon-based annotations for medical image retrieval were proposed byTizhoosh (2015), resulting in the IRMA error of 470.57 (3) on the total test set of IRMA benchmark. In another study developed in Tizhoosh, Zhu, Lo, Chaudhari, and V. (2016a), a small number of equidistant projections of Radon was examined to generate a retrieval system on a set of IRMA dataset.

Based on a wealth research reports, a combination of Radon transformation with deep learning techniques on IRMA benchmark has proven to be a reliable approach to image search (Liu et al., 2016; Sze-To, Tizhoosh, & Wong, 2016; Tizhoosh, Mitcheltree, Zhu, & Dutta, 2016b).Liu et al. (2016) used CNN features (the information from the last fully connected layer) for a local search scheme, obtained by Radon barcodes to achieve the IRMA error of 224.13. Also, the IRMA error of 344.08 was obtained by Sze-To et al. (2016), using deep autoencoders and Radon projections. In another research, a deep auto-encoded Radon retrieval system was developed and achieved the IRMA error of 392.09 (Tizhoosh et al., 2016b).

As reported in Müller et al. (2009), an IRMA error of 178.93 was acquired by Idiap research team, utilising Support Vector Machines (SVMs) and the two descriptors of LBP, and modSIFT (Tommasi & Orabona, 2010). Avni, Goldberger, and Greenspan (2009) achieved an IRMA error of 169.5 by using a dictionary approach on IRMA dataset. More specifically, they developed a multi-resolution patch-based dictionary approach by using principal component analysis on the densely sampled patches, followed by an SVM classifier trained on the bag-of-words. Camlica, Tizhoosh, and Khalvati (2015) obtained an IRMA error of 146.55, which it is the lowest reported error so far. However, their saliency method is extremely sluggish such that they neglect the overhead for the saliency calculations and simply use the offline-generated maps during testing. For this reason, we do not compare our results with their method as this approach would be impractical for daily clinical practice.

3. Methodology

A sequential shrunk search space is introduced in this study for a uniform LBP descriptor (Ojala, Pietikainen, & Maenpaa, 2002). The shrunk space enables an efficient searching system. A distance metric is utilised to measure the similarity between two images. Properly shrinking the search space for LBP guarantees an accurate and robust retrieval system. The proposed retrieval system is summarized into three main parts, as depicted in Fig. 2. (1) First stage shrinking which is equipped by a transfer learning technique. A CNN model is utilised for this step. (2) Second stage shrinking which is obtained by defining a selection pool. Radon transformation is utilised to create the pool. (3) The last step is a local search procedure based on a similarity-based routine. The LBP descriptor is used to measure the similarity between images via the Manhattan metric.

A brief discussion of the two shrinking steps is explained, as follows.

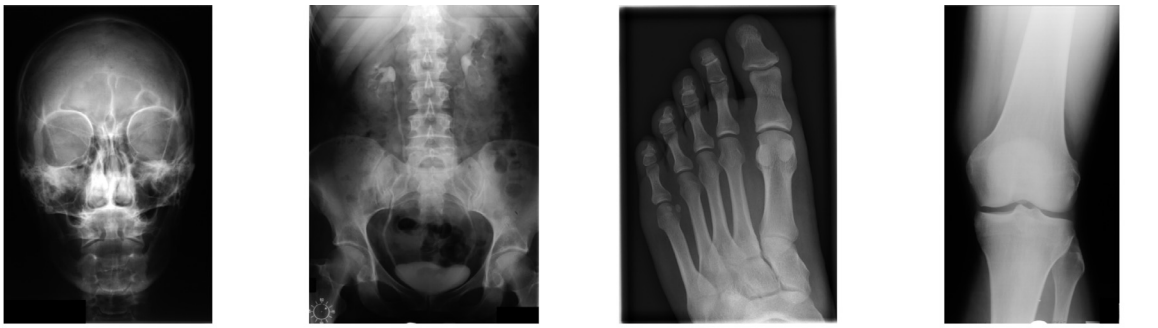

Fig. 1. Sample images from IRMA dataset with their IRMA codes TTTT-DDD-AAA-BBB.

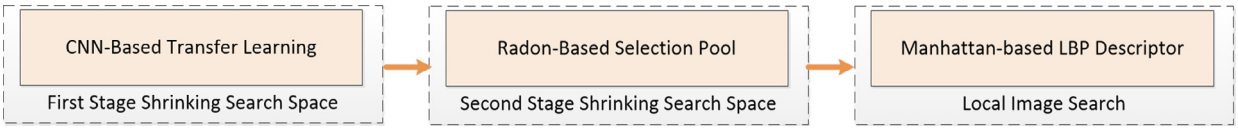

Fig. 2. The block diagram of the proposed CBIR scheme.

- **First step shrinking**: first stage (Shrinking search space): transfer learning scheme is utilised for the first shrinking procedure using a CNN as a per-trained network. The intuition behind this is to improve generalization, as discussed in Erhan et al. (2010). Accordingly, the pre-training process acts as a regularization technique because the initial layers are properly initialised in a supervised manner with the pre-trained weights. Therefore, gradient dilution difficulty is not severe anymore.
  Two different datasets are required for the transfer learning process. Data augmentation, which is the most common technique to prevent overfitting in deep learning (Krizhevsky, Sutskever, & Hinton, 2012), is used to enlarge the original data set. It enlarges the data set artificially by utilising translation, deformation, and reflection operations. Note that the influence of different data augmentation methods has been already investigated (Salamon & Bello, 2017). Section 4 explains more about the distributions of the two augmented sets. A pre-training CNN is applied on the first augmentation set, followed by a fine-tuning procedure performing on the second enlarged set. The output of this stage is a categorised image space.
  CNN architecture: As investigated in Khatami et al. (2017a), different experiments were conducted based on the layer patterns introduced in AlexNet (Krizhevsky et al., 2012) and LeNet (LeCun, Bottou, Bengio, & Haffner, 1998). The Theano framework (Bastien et al., 2012) has been utilised for implementation. Note that the AlexNet has a very similar architecture to the LeNet, but it is deeper and larger. Moreover, the features of convolutional layers are stacked on top of each other. However, in the LeNet, each convolutional layer always immediately is followed by a pooling layer. Different kernel sizes were investigated for convolutional and pooling layers. As reported in Khatami et al. (2017a), the size of 3×3 was finally selected for this study. It was also found that LeNet performed better on the IRMA data set. It follows the layer pattern of LeNet, Conv-Pool, which has been implemented for MNIST (LeCun et al., 1998). Rectified linear unit (LeCun, Bengio, & Hinton, 2015), short ReLU, is used as activation function. It is calculated by (1) in which the output is zero if any input value is less than zero.

$$ReLU(x) = max(0, x) \tag{1}$$

where x is the input to a neuron.

- **Second step shrinking (Shrinking search space)**: A further shrinking of the search space is obtained from this step. The shrinking procedure utilises the Radon transformation to create a selection pool. The selection pool is important due to its impact to achieve an efficient and effective retrieval system, using LBP for the next step which is an image search procedure.

The following pseudo code explains the procedure, briefly:

1. First stage shrinking:
   - Step 1: Create two augmented sets based on the procedures mentioned in experimental results.
   - Step 2: Apply a pre-trained CNN architecture on the first augmented set, followed by a fine-tuning procedure on the second augmented set.
2. Second stage shrinking:
   - Step 1: Suppose a test query is fed to the CNN and the predicted category for the query is called L. Calculate eight different Radon projections of the test query as well as all images in the L category.
   - Step 2: Calculate the difference between each pair of orthogonal projections. This results in four Radon feature vectors.
   - Step 3: Use Manhattan metric to measure the similarity between the four Radon vectors of the test query and all the images from Category L. Then insert the top n similar images into the pool. Now, the selection pool includes $4 \times n$ images.
3. Similarity-based image search:
   - Step 1: Use the LBP descriptor, equipped with Manhattan metric to find the most similar image from the selection pool.

3.1. Stage 1: convolutional neural network and fine-tuning

Fig. 3 shows the first step of the proposed model. Two different augmented sets are created from the original IRMA set. (1) The first augmented dataset is obtained by three most popular augmentation strategies, namely, flipping, rotation, and scaling (Chatfield, Simonyan, Vedaldi, & Zisserman, 2014; Krizhevsky et al., 2012). This results in 85k images from 12,677 initial training images. (2) The second augmented training set is obtained by following the below procedure:

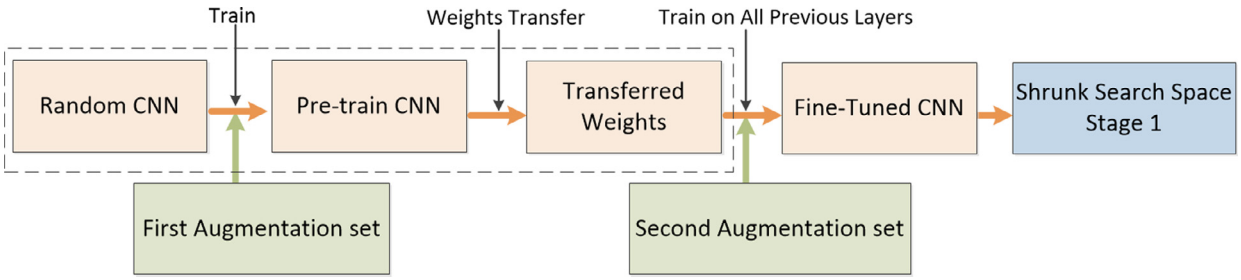

Fig. 3. The diagram for the first stage of the shrinking search space.

- Removing non-related parts: as conducted in Babaie et al. (2017b), the non-related parts such as burned-in annotations and other landmarks such as letters because of film digitization are discarded.
- Contrast adjustment: to highlight the differences between background and foreground, contrast adjustment is performed. It changes the contrast (brightness) of images by considering a specific threshold. Gamma correction with different gamma values is applied to varying the image contrast levels.
- Augmentation: the pre-processed data are augmented based on affine transformations of translation (shift) and shearing.

This results in 54k images from the original training set. The first augmented data set is used for the pre-training stage, while the second is utilised for further boosting during the fine-tuning phase. Considered by the similarity between both augmented data sets, we can have confidence that we would not overfit if fine-tuning is performed through the entire network (Goodfellow, Bengio, & Courville, 2016; Yosinski, Clune, Bengio, & Lipson, 2014). It is because the second augmented dataset is similar to the first one, however the distributions are different.

3.2. Stage 2: selection pool (shrink search space)

This stage creates a selection pool for further shrinking of the search space which is utilised for the next local search-based step. The role of this phase is important because it tackles the curses of dimensionality which are critical for search-based techniques. The input to this stage is the categorised images classified by the previous step. The selection pool uses similarity measures as below to select the n-top best images from the corresponding category, which have the most similarity with the given query image. This approach significantly reduces dimensions of the search space, resulting in a further shrinking. Note that, a robust and accurate retrieval system is promising if a properly shrunk search space is created for its searching phase.

To achieve this point, Radon transformation (Radon, 2005) is utilised. As mentioned in the literature, different feature descriptors might be used, however Radon transformation is selected due to its simple implementation and efficiency in matching. Radon is also a well-known technique in medicine, and a wealth of medical studies has been conducted based on this descriptor. Moreover, it results in an efficient retrieval system because it has less necessity to memory and storage. In other words, it creates a low-dimensional feature vector for further analysis, which is in contrast to the other descriptors such as LBP or HOG (Tizhoosh, 2015).

Radon transformation. Radon transformation (Radon, 2005), calculated by (2), is an integral conversion, computing the summation of information of an image from several angles, θ. Fig. 4 shows three Radon projections of a matrix, representing an image, from three different θ. The key point of Radon is the ability of this transformation to reconstruct the original image from different aspects.

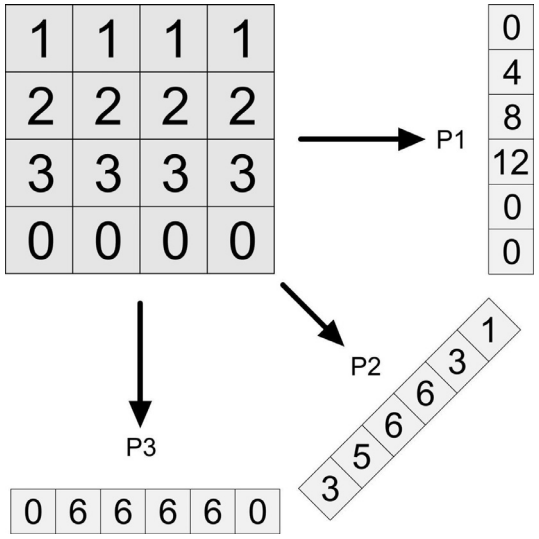

Fig. 4. An example of three projections of Radon on a matrix.

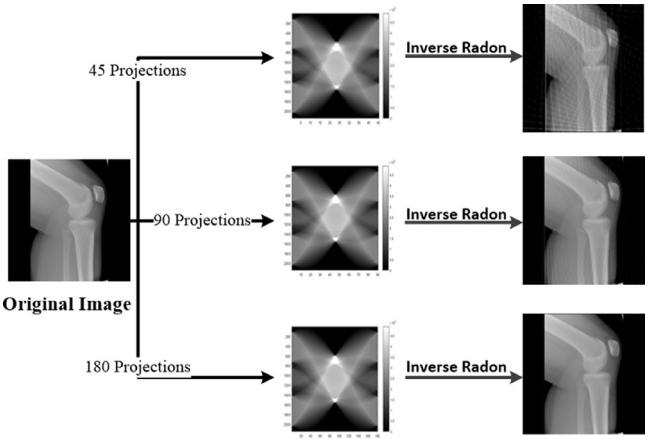

Fig. 5. The impact of utilising different numbers of Radon projections on an IRMA sample.

Note that the information of the image from each angle (θ) is depicted as a signal. As illustrated in Fig. 5, it is obvious that if the more angles are selected, the higher illuminated images are reconstructed. It is noted that the optimal number of angles is still a challenging problem.

$$R(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(\rho - xcos\theta - ysin\theta)dxdy, \quad (2)$$

where ρ refers to grey-level intensities of image f at position θ, and δ is a Dirac delta operator.

Radon transform with the differentiate two projections. Contrast, directionality, edges, and corner are some of the similarity properties describing an image. Extracting more relevant and proper features representing an image is a major challenge due to the rich content of the image. In other words, a huge dimensional feature vector in a CBIR system causes a difficulty for similarity measures to retrieve accurate images. To tackle the problem, Radon transform which is well-known in medicine and also obtains a low-dimensional feature vector is utilised. In order to construct an efficient and robust search-based scheme which uses LBP, we perform the Radon transformation to create a selection pool, obtaining by n-top best images from a corresponding category, which have the most similarity with a given query. The more projections are considered, the more information is extracted from the image. It is worth mentioning that using the more information results in more accurate and robust system; however computational costs are increased enormously. To solve the problem, several strategies, based on computational costs of searching scheme and number of projections, are defined in experimental results. The best scenario, proposed in this paper is to select *four* projections which are created based on the difference among *eight* projections, selected in the selection pool, as discussed below. It significantly reduces the search space dimensions as well as computational costs.

How to create the selection pool. The below three procedures are used to create the second stage of shrinking the search space, as illustrated in Fig. 6.

- *First procedure*: Suppose a query test image is given to the pre-trained CNN in the first stage. Radon transformation from *eight* different projections is performed on the test query. The similar strategy is performed on all the images in the corresponding category, predicted by the CNN. Note that the number of images here is significantly lower than that of the whole dataset.
- *Second procedure*: The difference between two orthogonal projections is calculated to keep more relevant features from the two orthogonal Radon projections. As discussed in the next section, it significantly reduces the retrieval error as compared with considering just one projection. It results in *four* pairs Radon feature vectors.
- *Third procedure*: Then, LBP descriptor with Manhattan metric is considered for measuring the similarity between the Radon feature vectors of the test query and those of the images from the corresponding category. The best n-top similar images are selected from the category and inserted into the selection pool as the retrieved candidates for the query. As discussed in Experimental Results, an investigation is conducted on the type of the metrics used for the measurement in the LBP. It is also followed by an analysis on the number of n, resulting in an accurate retrieval system. Therefore, the same process is performed on all four vectors in the previous step. Finally, we have a pool of images including $4 \times n$ training samples.

Note that regarding the second procedure, we do not concatenate these projections in the proposed pool selection because a fast search method performing on low-dimensional feature vectors, embedding in real-time retrieval systems, is investigated. Moreover, as experimented, concatenating the projections results in lower performance. Also, it is worth mentioning that an analysis on the optimal number of n, in the third procedure, is required to build an accurate retrieval system.

3.3. Stage 3: similarity-based image search

Image search procedure is discussed in this part. It is clear that properly reducing and shrinking a search space for a retrieval system considerably speed up the retrieval process as well as preserve retrieval accuracy. In our study, at the first stage, a shrunk search space, using a pre-training model was proposed. Then, the selection pool was created properly for further dimensionality reduction. Now, the LBP descriptor using Manhattan metric is utilised to find the most similar image from the selection pool, with respect to the query. It is a quick search procedure by the LBP because of performing on low-dimensional selection pool. The LBP creates a histogram of m bins per cell, set into $k \times k$ pixels, causing LBP histograms with length of $m \times k$. The low-dimensional selection pool obtains a small number of m, which results in a low length histogram for the LBP. Therefore, by considering $4 \times n$ training samples per query, we have at most $4 \times n$ for each LBP histograms, i.e. totally $4 \times n \times m \times k$. Note that $m \gg n$, k, and obtaining a considerable small m significantly reduces the search space dimension, and this tackles the curse of dimensionality problem. As explained in Experimental Results, the selection pool provides a significant small number of m which results in an efficient retrieval system.

4. Experimental results

4.1. IRMA dataset

The IRMA dataset includes a variety of x-ray images which have been randomly selected from radiology routine work at the Department of Diagnostic Radiology, Aachen University of Technology (RWTH), Aachen, Germany (Lehmann et al., 2005; 2004b)[1]. The radiology images represent various cases with respect to patients' age and gender, viewing positions, and pathologies. All radiology images have been rescaled to fit into a 512×512 bounding box maintaining the original aspect ratio via zero padding. All samples have a specific **IRMA code** defining 57 image categories. The training samples contain 12,677 radiographs with known categories. The test samples comprise of 1733 radiographs.

The complete IRMA code shows a string code of 13 characters, each in $\{0, \ldots, 9; a, \ldots, z\}$: TTTT-DDD-AAA-BBB. This string constitutes four mono-hierarchical axes: the technical code T (imaging modality), directional code D (body orientations), anatomical code A (the body region), and biological code B (the biological system examined). Fig. 7 depicts two sample images from the benchmark along with their IRMA code in the format TTTT-DDD-AAA-BBB.

The error evaluation procedure, proposed in ImageCLEF09, is used to calculate the retrieval error (Lehmann et al., 2004a; Lehmann, Schubert, Keysers, Kohnen, & Wein, 2003). The total IRMA error can be computed by

$$\sum_{i=1}^{I} \underbrace{\frac{1}{b_i}}_{(a)} \underbrace{\frac{1}{i}}_{(b)} \underbrace{\delta(l_i, \hat{l}_i)}_{(c)} \tag{3}$$

with

$$\delta(l_i, \hat{l}_i) = \begin{cases} 0 & \text{if } l_j = \hat{l}_j \ \forall j \le i \\ 0.5 & \text{if } l_j = * \ \exists j \le i \\ 1 & \text{if } l_j \neq l_j \ \exists j \le i \end{cases}$$

where b is the number of possible labels at position i, and δ is a decision measure carrying out 1 for incorrect label and 0 for correct one, when the image I_i is compared with the image $\hat{I}_i$.

4.2. Step 1: convolutional neural network and fine-tuning

The first stage of shrinking the search space is performed in this step. A deep learning technique is utilized for classifying the 12,677 IRMA images. All images are rescaled to zero-padded square

[1] URL to download the dataset: https://goo.gl/NX44yh.

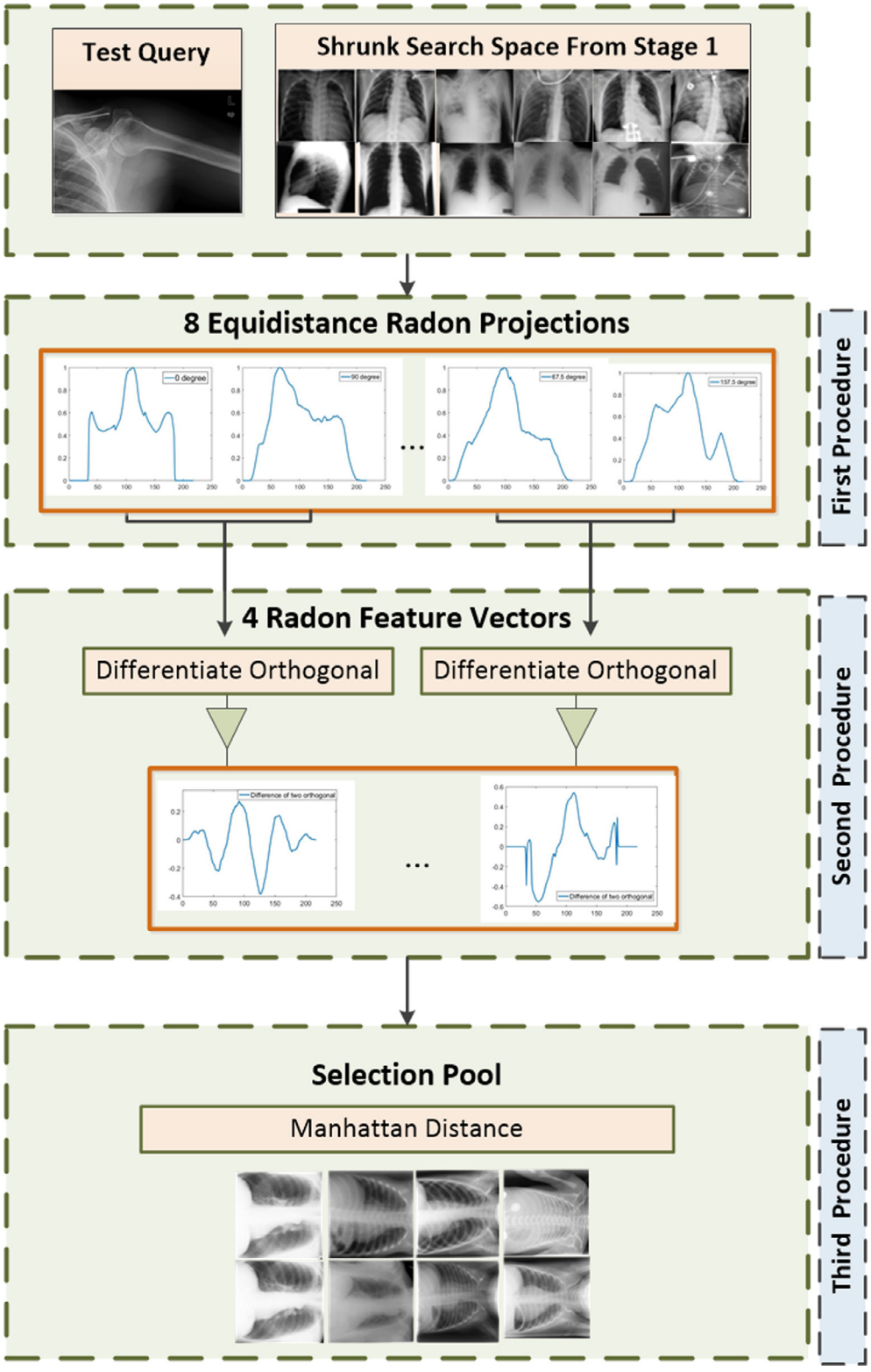

Fig. 6. The diagram for the second stage of the shrinking search space.

samples of size 56×56. Due to imbalance, augmentation techniques are required. The two augmented dataset including of 85k and 54k images are created, as mentioned in previous section.

Deep-based performances are investigated by two strategies of (1) using a CNN trained from scratch on a big dataset merging by the two augmented data sets, and (2) using a pre-trained network on the first augmented set, followed by a fine-tuning procedure on the weights of the pre-trained CNN which is optimized by back-propagation on the second augmented one. Superior performance is achieved by the latter. The fine-tuning is performed on all layers of the CNN. There are two reasons for this: (1) inspired by the fact that CNN features are more generic in early layers and

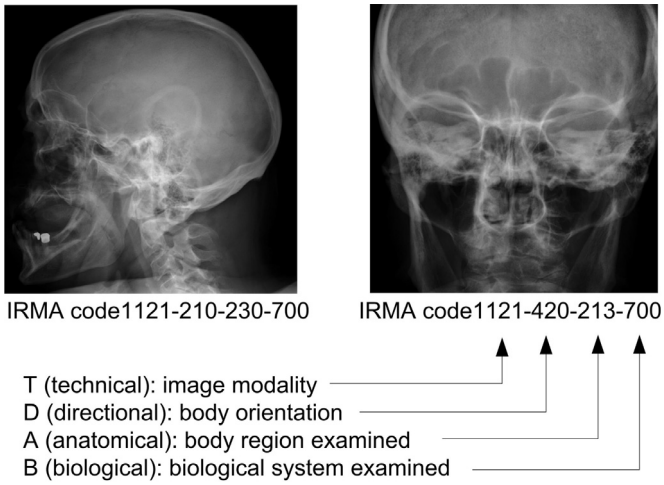

Fig. 7. A description of two 13-digit IRMA codes for two X-ray images.

Table 1
The deep CNN details.

Name	Filter size	Filter dimension	Stride
Conv1	3	16	1
ReLU1	1		1
Pooling1	2		2
Conv2	3	32	1
ReLU2	1		1
Pooling2	2		2
Conv3	3	64	1
ReLU3	1		1
Pooling3	2		2
Conv4	3	128	1
ReLU4	1		1
Pooling4	2		2
FC	1	625	1

more specific toward output layers due to the details of the classes contained in the original dataset; (2) considered by the similarity between both large-scaled augmented data sets, we can have more confidence that we would not overfit if fine-tuning is performed through the entire network (the second augmented dataset is similar to the first one, however the distributions are different) (Goodfellow et al., 2016; Yosinski et al., 2014).

Table 1 presents in more detail the effective CNN structure used for the transfer learning scenario. The processed and augmented data are fed into the pre-trained CNN, which is followed by fine-tuning on the second augmented set. After fine-tuning, test queries are fed into the CNN to obtain the predicted categories. A shrunk search space per query is the output of this stage. It shows which category the query belongs to. However, in practical applications, the selected category contains many samples. A further shrinking procedure is necessary. Therefore, a selection pool is proposed to generate an efficient and effective retrieval system with respect to the space of all predicted categories, called *shrunk space*.

Table 3
Using different distance metrics for LBP histogram matching.

Metrics utilised in LBP	IRMA Error
Our proposed model with LBP using Manhattan metric	**168.05**
Our proposed model with LBP using Euclidean metric	172.25
Our proposed model with LBP using Cosine metric	177.41

4.3. Step 2: selection pool (shrinking the search space)

Different numbers of Radon projections are applied to the test query as well as all images of the *shrunk search space*. In previous studies (Babaie et al., 2017b; Tizhoosh, 2015), eight equidistant Radon projections $X \in \{0^\circ, 22.5^\circ, 45^\circ, \ldots, 157^\circ\}$ were conducted. As illustrated in Fig. 6, each projection depicts the information of an image as a signal of size 56, which is equal to the number of dimensions of the image. The main drawback of this approach is loss of the information once a single projection is considered for retrieval. Therefore, to improve the retrieval performance, the difference of two orthogonal Radon projections, denoted by D_i, $i \in \{1, 2, 3, 4\}$, is considered. Each D_i is a difference between each of the orthogonal pairs of X, such as $\{0^\circ, 90^\circ\}$. This results in more expressive feature vectors with size of 4×56, rather than 8×56. A similar procedure is performed on the test query and all images belonging to the corresponding shrunk search space per query. Then, the LBP search is applied to create the selection pool.

4.4. Step 3: similarity-based image search

Suppose a test query is fed into the system. The output is a selection pool with a size of $100(4 \times 25)$ feature vectors, obtained by applying the 4 orthogonal Radon transformations on the 25 images. The LBP uses the Manhattan metric to find the best feature vector of the test query within the selection pool. Specifically, the LBP descriptor has a radius of 1 and 8 neighbours per pixel. It creates a histogram of 59 bins per cell, set into 12×12 pixels. This results in LBP histograms with of length 8496. Accordingly, for each query, at most 100 LBP histograms with 8496 bins are created in the selection pool.

4.5. Discussion

The two-step hierarchical shrinking phase, derived by the CNN (step 1) and the selection pool (step 2), creates a shrunk search space which facilitates the operation of an efficient and robust retrieval system for an average sized and imbalanced data set. To configure a more accurate retrieval model, different scenarios are scrutinised, as follows. Table 2 compares the performance of the scenarios.

- Scenario 1 (CNN-based TL + LBP): The first stage shrinking, based on transfer learning, is followed by applying LBP with the Manhattan metric to retrieve the images. This strategy produces an IRMA error of 198.20, which is considerably higher than that of the proposed model. Moreover, compared with the proposed model, the computational cost is high.

Table 2
A comparison among different scenarios.

Scenarios	Time in min:s (for 1733 test images)	IRMA error
Scenario 5 (Our proposed model)	**16:21**	**168.05**
Scenario 1 (CNN-based TL + LBP)	29:36	198.20
Scenario 2 (SP without differentiation + LBP)	89:08	311.83
Scenario 3 (SP with differentiation + LBP)	61:43	280.11
Scenario 4 (CNN-based TL + Scenario 2 + LBP)	19:26	186.41

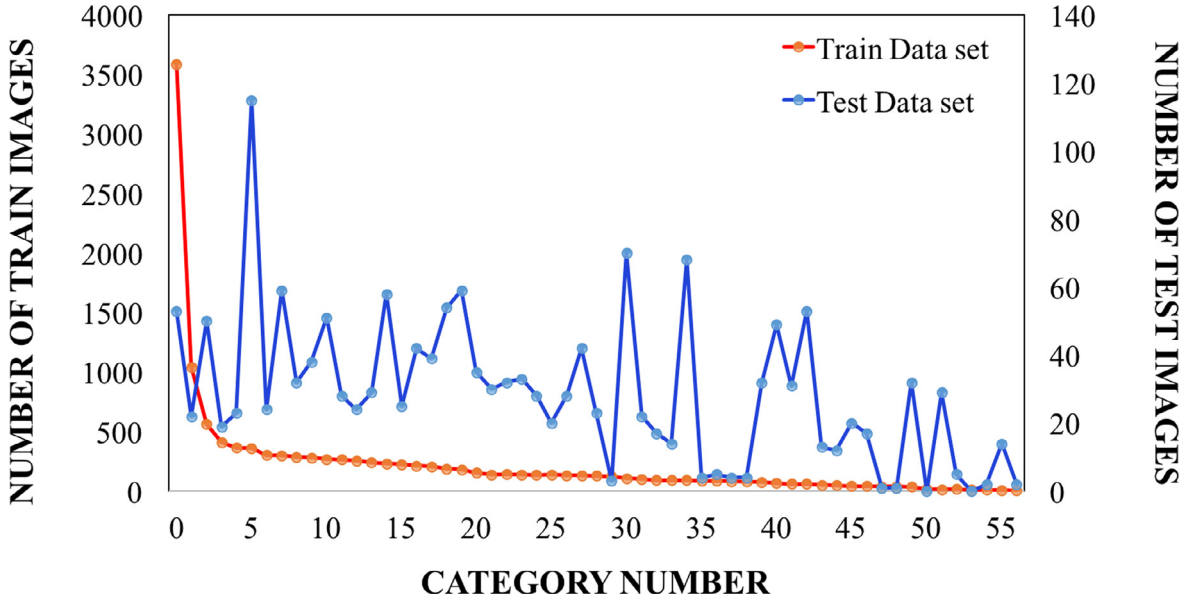

Fig. 8. Imbalance in both training an testing portions of IRMA images.

Table 4
Comparisons of our results to the numbers reported in literature for IRMA images. The methods with * are reported in Tommasi et al. (2009) and the methods with ** are reported in Liu et al. (2016). Accuracies are also provided according to accuracy $= 1 - \frac{\text{error}}{1733}$.

	IRMA error	Accuracy
Proposed method	**168.05**	**90.30%**
TAUbiomed *	169.5	90.22%
diap*	178.93	89.68%
CNNC+RBC**	224.13	87.00%
FEITIJS*	242.46	86.00%
SuperPixel**	249.34	85.61%
VPA*	261.16	84.93%
SP-R**	311.8	82.01%
MedGIFT *	317.53	81.68%
VPA*	320.61	81.15%
SP-RBC**	356.57	79.42%
IRMA*	359.29	79.27%
MedGIFT*	420.91	75.57%

- Scenario 2 (Selection pool (SP) without differentiation + LBP): A selection pool with 8 equidistant Radon projections, as conducted in Babaie et al. (2017b), is used here. The SP procedure is followed by a local search using the LBP. The weak performance of 311.88 IRMA error shows that not using the CNN-based TL with the differentiate projections significantly affects the performance.
- Scenario 3 (SP with differentiation + LBP): 4 orthogonal Radon transformations are created from the previous scenario. The IRMA error of 280.11 shows an improvement, as compared with scenario 2, however the performance is considerably lower than that of the proposed model.
- Scenario 4 (CNN-based TL + Scenario 2 + LBP): A combination of scenarios 1 and 2 is investigated. The model is efficient, however, it creates the SP with 8 projections. The IRMA error of 186.41 shows the improvement, as compared with those from previous scenarios but its performance is still lower than that of the proposed model.
- Scenario 5 **Our proposed model (CNN-based TL + Scenario 3 + LBP)**: A combination of scenarios 1 and 3 is investigated. This is the most efficient model among all scenarios. The two sequential shrinking stages provide a search space which leads to accurate retrieval outcomes. This results in an IRMA error of 168.05, which is the best performance reported in the literature so far when efficiency is considered as well for real-time processing.

As explained in Table 3, different distance measures like Euclidean, Cosine, and Manhattan are investigated for LBP histogram matching. The Manhattan metric is selected after investigation.

Table 4 compares the proposed technique with several models reported in literature (Liu et al., 2016; Tommasi et al., 2009). Note that to the best of the author's knowledge, the proposed model outperforms all other methods reported in the literature based on the IRMA benchmark data set. Moreover, with respect to the computational cost, the proposed technique is efficient, and can be implemented in a real-time medical application. As experienced, the total cost to retrieve 1733 test images is of 16 min, that is 0.55 s for each query, which is considerably lower than the cost of 87 min before the shrinking.

IRMA dataset is quite challenging. The dataset poses several challenges to deep solutions. First of all, it is not large enough to provide discriminatory information for separating 57 classes. As well, the dataset is extremely imbalanced (see Fig. 8). Many artefacts and burnt-in annotations due to digitization cause additional difficulties.

The experiments show that when transfer learning is performed on two augmented data sets with different distributions, instead of a large one, it increases the performance. Considering a two-step hierarchical shrinking scheme significantly reduces the computational cost. Proposing a hierarchical shrunk space, derived from a classification phase and a properly created selection pool, can reduce the effect of imbalanced data distribution. In fact, the proposed method decreases the amount of image information lost by proposing the sequential shrinking search space properly. Accordingly, it improves the semantic gap which exists between the information extracted from machinery and the high-level features interpreted by humans. Considering a well-designed hierarchical shrunk search space can outperform the dictionary method, which is a very effective method to extract proper features in a retrieval domain.

5. Conclusions

A robust CBIR system often requires an accurate feature extraction technique and an efficient search model. In some real-world applications like medicine, it is a challenging task to properly achieve these goals concurrently because of some problems like imbalance and lack of big labelled data. Properly shrinking the search space is important to propose a similarity-based scheme for retrieval purposes. Hence, a deep structural method with the employment of transfer learning is proposed to obtain a robust feature extraction stage, resulting in an accurate classification system. This shrunk data categories along with the further shrinking steps, derived by a projection-based selection pool, result in a two-step hierarchical shrinking phase which enables a robust feature representation system for CBIR tasks. This contribution beats several state-of-the-art methods, especially dictionary approach on a strongly imbalanced IRMA dataset.

References

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *International conference on database theory* (pp. 420–434). Springer.

Alonso, I., & Contreras, D. (2016). Evaluation of semantic similarity metrics applied to the automatic retrieval of medical documents: An umls approach. *Expert Systems with Applications, 44*, 386–399.

Avni, U., Goldberger, J., & Greenspan, H. (2009). Addressing the imageclef 2009 challenge using a patch-based visual words representation.. *Clef (working notes)*.

Avni, U., Greenspan, H., Konen, E., Sharon, M., & Goldberger, J. (2011). X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Transactions on Medical Imaging, 30*(3), 733–746.

Baâzaoui, A., Barhoumi, W., Ahmed, A., & Zagrouba, E. (2018). Modeling clinician medical-knowledge in terms of med-level features for semantic content-based mammogram retrieval. *Expert Systems with Applications, 94*, 11–20.

Babaie, M., Kalra, S., Sriram, A., Mitcheltree, C., Zhu, S., Khatami, A., et al. (2017a). Classification and retrieval of digital pathology scans: A new dataset. *Cvmi workshop@ cvpr*.

Babaie, M., Kalra, S., Sriram, A., Mitcheltree, C., Zhu, S., Khatami, A., et al. (2017b). Classification and Retrieval of Digital Pathology Scans: A New Dataset. In *Computer Vision for Microscopy Image Analysis (CVMI). Workshop held in Conjunction with the Computer Vision and Pattern Recognition (CVPR) Conference*.

Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I., Bergeron, A., et al. (2012). Theano: New features and speed improvements. In *Proceedings of deep learning workshop, NIPS 2012,* arxiv preprint arxiv:1211.5590.

Brahnam, S., Jain, L. C., Nanni, L., Lumini, A., et al. (2014). *Local binary patterns: New variants and applications*. Springer.

Camlica, Z., Tizhoosh, H., & Khalvati, F. (2015). Medical image classification via svm using lbp features from saliency-based folded data. *Machine learning and applications (ICMLA), the 14th international conference on*.

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *Published in proceedings of BMVC 2014*.

Clack, R., & Defrise, M. (1994). Cone-beam reconstruction by the use of radon transform intermediate functions. *JOSA A, 11*(2), 580–585.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., & Bengio, S. (2010). Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research, 11*(Feb), 625–660.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org.

Greenspan, H., & Pinhas, A. T. (2007). Medical image categorization and retrieval for pacs using the gmm-kl framework. *IEEE Transactions on Information Technology in Biomedicine, 11*(2), 190–202.

Hinneburg, A., Aggarwal, C. C., & Keim, D. A. (2000). What is the nearest neighbor in high dimensional spaces? In *26th international conference on very large databases* (pp. 506–515).

Junior, O. L., Delgado, D., Gonçalves, V., & Nunes, U. (2009). Trainable classifier-fusion schemes: An application to pedestrian detection. *Intelligent transportation systems*: 2.

Kashif, M., Deserno, T. M., Haak, D., & Jonas, S. (2016). Feature description with sift, surf, brief, brisk, or freak? A general question answered for bone age assessment. *Computers in Biology and MNeural Information Processing Systemsedicine, 68*, 67–75.

Khatami, A., Babaie, M., Khosravi, A., Tizhoosh, H., & Nahavandi, S. (2018). Parallel deep solutions for image retrieval from imbalanced medical imaging archives. *Applied Soft Computing, 63*, 197–205.

Khatami, A., Babaie, M., Khosravi, A., Tizhoosh, H., Salaken, S. M., & Nahavandi, S. (2017a). A deep-structural medical image classification for a radon-based image retrieval. In *Electrical and computer engineering (ccece), 2017 IEEE 30th Canadian conference on* (pp. 1–4). IEEE.

Khatami, A., Khosravi, A., Lim, C. P., & Nahavandi, S. (2016). A wavelet deep belief network-based classifier for medical images. In *International conference on neural information processing* (pp. 467–474). Springer.

Khatami, A., Khosravi, A., Nguyen, T., Lim, C. P., & Nahavandi, S. (2017b). Medical image analysis using wavelet transform and deep belief networks. *Expert Systems with Applications*.

Khatami, A., Mirghasemi, S., Khosravi, A., Lim, C. P., Asadi, H., & Nahavandi, S. (2017c). A swarm optimization-based kmedoids clustering technique for extracting melanoma cancer features. In *International conference on neural information processing* (pp. 307–316). Springer.

Khatami, A., Tai, Y., Khosravi, A., Wei, L., Dalvand, M. M., Peng, J., & Nahavandi, S. (2017d). A haptics feedback based-lstm predictive model for pericardiocentesis therapy using public intraoperative data. In *International conference on neural information processing* (pp. 810–818). Springer.

Khatami, A., Tai, Y., Khosravi, A., Wei, L., Dalvand, M. M., Zou, M., & Nahavandi, S. (2017e). A deep learning-based model for tactile understanding on haptic data percutaneous needle treatment. In *International conference on neural information processing* (pp. 317–325). Springer.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

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. *Journal of Digital Imaging, 26*(6), 1025–1039.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.

Lehmann, T. M., Güld, M. O., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., et al. (2005). Automatic categorization of medical images for content-based retrieval and data mining. *Computerized Medical Imaging and Graphics, 29*(2), 143–155.

Lehmann, T. M., Güld, M. O., Thies, C., Plodowski, B., Keysers, D., Ott, B., & Schubert, H. (2004a). Irma-content-based image retrieval in medical applications. *Medinfo*, 842–848.

Lehmann, T. M., Schubert, H., Keysers, D., Kohnen, M., & Wein, B. B. (2003). The irma code for unique classification of medical images. In *Medical imaging 2003* (pp. 440–451). International Society for Optics and Photonics.

Lehmann, T. M., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., et al. (2004b). Content-based image retrieval in medical applications. *Methods of Information in Medicine, 43*(4), 354–361.

Liu, X., Tizhoosh, H. R., & Kofman, J. (2016). Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In *Neural Networks (IJCNN), 2016 International Joint Conference on.* (pp. 2872–2878). IEEE.

Metz, C. E., & Pan, X. (1995). A unified analysis of exact methods of inverting the 2-d exponential radon transform, with implications for noise control in spect. *IEEE Transactions on Medical Imaging, 14*(4), 643–658.

Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., et al. (2009). Overview of the clef 2009 medical image retrieval track. In *Workshop of the cross-language evaluation forum for european languages* (pp. 72–84). Springer.

Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications clinical benefits and future directions. *International Journal of Medical Informatics, 73*(1), 1–23.

Nanni, L., Brahnam, S., & Lumini, A. (2010). A local approach based on a local binary patterns variant texture descriptor for classifying pain states. *Expert Systems with Applications, 37*(12), 7888–7894.

Napel, S. A., Beaulieu, C. F., 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 1. *Radiology, 256*(1), 243–252.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 971–987.

Pan, L., Qiang, Y., Yuan, J., & Wu, L. (2016). Rapid retrieval of lung nodule ct images based on hashing and pruning methods. *BioMed Research International, 2016*.

Pietikäinen, M., Hadid, A., Zhao, G., & Ahonen, T. (2011). *Local binary patterns for still images*. Springer.

Radon, J. (2005). 1.1 Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Classic Papers in Modern Diagnostic Radiology, 5*.

Rahman, M. M., Bhattacharya, P., & Desai, B. C. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Transactions on Information Technology in Biomedicine, 11*(1), 58–69.

Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters, 24*(3), 279–283.

Sargent, D., Chen, C.-I., Tsai, C.-M., Wang, Y.-F., & Koppel, D. (2009). Feature detector and descriptor for medical images. *Spie medical imaging*. International Society for Optics and Photonics. 72592Z–72592Z

Seo, K.-K. (2007). An application of one-class support vector machines in content-based image retrieval. *Expert Systems with Applications, 33*(2), 491–498.

Shin, H.-C., Roberts, K., Lu, L., Demner-Fushman, D., Yao, J., & Summers, R. M. (2016). Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2497–2506).

Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349–1380.

Sze-To, A., Tizhoosh, H. R., & Wong, A. K. (2016). Binary codes for tagging x-ray images via deep de-noising autoencoders. In *Neural networks (IJCNN), 2016 international joint conference on* (pp. 2864–2871). IEEE.

Tizhoosh, H. (2015). Barcode annotations for medical image retrieval: A preliminary investigation. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 818–822). IEEE.

Tizhoosh, H., Zhu, S., Lo, H., Chaudhari, & Mehdi, V. T. (2016a). Minmax radon barcodes for medical image retrieval. *International Symposium on Visual Computing*, 617–627.

Tizhoosh, H. R., Mitcheltree, C., Zhu, S., & Dutta, S. (2016b). Barcodes for medical image retrieval using autoencoded radon transform. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (pp. 3150–3155). IEEE.

Tommasi, T., Caputo, B., Welter, P., Güld, M. O., & Deserno, T. M. (2009). Overview of the clef 2009 medical image annotation track. In *Workshop of the cross-language evaluation forum for European languages* (pp. 85–93). Springer.

Tommasi, T., & Orabona, F. (2010). Idiap on medical image classification. In *Imageclef* (pp. 453–465). Springer.

Vipparthi, S. K., & Nagar, S. (2014). Expert image retrieval system using directional local motif xor patterns. *Expert Systems with Applications, 41*(17), 8016–8026.

Weisi, L., Tao, D., Kacprzyk, J., Li, Z., Izquierdo, E., & Wang, H. (2011). *Multimedia analysis, processing and communications*: 346. Springer Science & Business Media.

Xu, X., Lee, D.-J., Antani, S., & Long, L. R. (2008). A spine X-ray image retrieval system using partial shape matching. *IEEE Transactions on Information Technology in Biomedicine, 12*(1), 100–108.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320–3328).

Zhang, F., Song, Y., Cai, W., Liu, S., Liu, S., Pujol, S., et al. (2016). Pairwise latent semantic association for similarity computation in medical imaging. *IEEE transactions on Biomedical Engineering, 63*(5), 1058–1069.

Zhang, S., Huang, J., Li, H., & Metaxas, D. N. (2012). Automatic image annotation and retrieval using group sparsity. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42*(3), 838–849.