Neurocomputing 357 (2019) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Breast cancer diagnosis through active learning in content-based image retrieval

Rafael S. Bressan[a], Pedro H. Bugatti[a], Priscila T.M. Saito[a,b,*]

[a] Department of Computing, Federal University of Technology - Parana, PR, Brazil
[b] Institute of Computing, University of Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 1 March 2019
Revised 22 April 2019
Accepted 15 May 2019
Available online 17 May 2019

Communicated by Yongdong Zhang

Keywords:
Machine learning
Computer vision
Image analysis
Active learning
Image retrieval
Breast cancer diagnosis

ABSTRACT

One of the cornerstones of content-based image retrieval (CBIR) for medical image diagnosis is to select the images that present higher similarity with a given query image. Different from previous literature efforts, the present work aims to seamlessly fuse a powerful machine learning strategy based on the active learning paradigm, in order to obtain greater efficacy regarding similarity queries in medical CBIR systems. To do so, we propose a new approach, named as Medical Active leaRning and Retrieval (MARRow) to aid the breast cancer diagnosis. It enables to deal with more feasible strategies, specifically for the medical context and its inherent constraints. We also proposed an active learning strategy to select a small set of more informative images, considering selection criteria based on not only similarity, but also on certain degrees of diversity and uncertainty. To validate our proposed approach, we performed experiments using public medical image datasets, different descriptors for each one and compared our approach against four widely applied and well-known literature approaches, such as: Traditional CBIR without relevance feedback strategies, Query Point Movement Strategy (QPM), Query Expansion (QEX) and SVM Active Learning (SVM-AL). From the experiments, we can observe that our approach presents a strong performance over state-of-the-art ones reaching a precision gain of up to 87.3%. MARRow also presented a well-suited and consistent increasing rate along the learning iterations. Moreover, our approach can significantly minimize the expert's involvement in the analysis and annotation process (reducing up to 88%). The results testify that MARRow improves the precision of the similarity queries. It is capable to explore at the maximum the experts' intentions, which are captured during the relevance feedback process, incrementally improving the learning model. Therefore, our approach can be suitable and applied in challenging processes, such as real and medical contexts, enhancing medical decision support systems (e.g. breast cancer diagnosis).

1. Introduction

Over the last decades, medical image databases have been growing due to technological advances in data acquisition and storage devices. Hence, the improvements of automatic retrieval and classification [1–10] approaches have become necessary to handle and organize such data. To perform these tasks we can use the content-based image retrieval (CBIR) process. It aims to retrieve images based on the similarity (or dissimilarity) between a given query image and an image dataset.

To compute these similarities, low-level features based on color, texture and/or shape are extracted from images [11]. Besides the set of features, the dissimilarity function (or distance function) and the expert (e.g. radiologist) interaction with the CBIR process are key aspects to obtain more precise results. Once each expert has his/her own perception and expertise, the relevance feedback (RF) paradigm can be applied to capture the expert's intention in a coarse-grained way [12]. It allows the expert to label the retrieved images as relevant or irrelevant regarding a given iteration. In other words, it leads to a query refinement. Then, when the CBIR process returns the similar images according to a query image, the pipeline can be fed with the relevance degree of each retrieved image. This information is aggregated with the image features and the distance function to perform a new query that is closer to the expert's intention. The RF process can be done until the expert is satisfied with the returned images.

Although there is a plenty of RF methods in the literature [13–16], to the best of our knowledge, the majority of them leave the definition of the degree of relevance and irrelevance to the ex-

* Corresponding author.
E-mail addresses: rafaelbressan@alunos.utfpr.edu.br (R.S. Bressan), pbugatti@utfpr.edu.br (P.H. Bugatti), psaito@utfpr.edu.br (P.T.M. Saito).

https://doi.org/10.1016/j.neucom.2019.05.041
0925-2312/

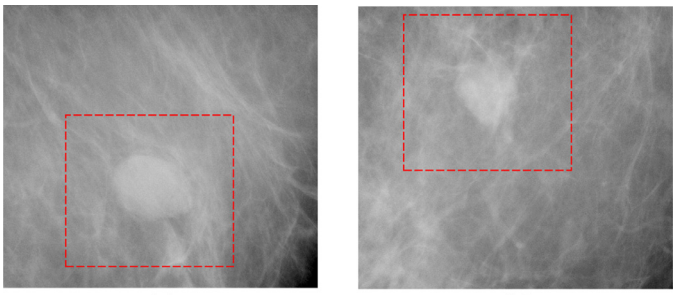


Fig. 1. Examples of informative (uncertain) samples from two different classes: (a) benign and (b) malignant lesions. It is possible to notice that both images (regions of interest from different classes) present a high similarity degree regarding their lesions (highlighted by dashed lines) and other tissues.

pert. Often these methods do not take into account or do not take full advantage of this information throughout the learning iterations. Besides that, once a query is refined at each iteration more and more relevant images are returned. However, all these images may not contribute to the learning process of a given image classifier. Therefore, to overcome these issues, active learning (AL) strategies can be embedded into the CBIR process.

AL [17] is a machine learning paradigm that selects the most informative samples for the learning process. It allows a small set of unlabeled learning samples to be selected and displayed, iteratively, for expert annotations. Then, the annotated set is used for training a classifier. Several active learning techniques [18,19] have been developed using different selection strategies to obtain the most informative samples. Although well-known and widely used in different domains, many of them are unfeasible, specifically for the medical context and its inherent constraints (e.g. related to dealing with large datasets, interactive response times, and minimal expert intervention in the learning process).

To take into account these characteristics, we propose in the next section an active learning strategy dedicated to the RF in the CBIR process, based on the uncertainty and diversity criteria. We focus our main attention on the medical context, specifically involving the diagnosis of breast cancer.

2. Methodology

We proposed an approach, named as Medical Active leaRning and Retrieval (MARRow), that aggregates active learning strategies for content-based breast image retrieval. In the first iteration, the expert performs the traditional annotation process (indicating relevant and irrelevant images, according to a given query image), as the classic RF loop. Next, from the second iteration forward, the selected and retrieved images, which will be trained in an active learning process, are those that will most contribute to the learning process of a given classifier. Unlike the literature works, in our approach the most informative images are those that present the best balance between not only the similarity with the query image, but also certain degrees of diversity and uncertainty. In other words, those images that are from different classes and difficult to differentiate, when we compare the query image semantics and the retrieved image ones (e.g. images at the boundaries of two different/overlapped classes).

For instance, in Fig. 1, we can see an example of two informative (uncertain) images located at the boundaries of two different classes, benign and malignant lesions, respectively, that will be presented to the expert (instead of only images closer to the query center). It is possible to notice that both images (regions of interest from different classes) present a high similarity degree regarding their lesions (highlighted by dashed lines) and other tissues. Through our approach, we can balance the learning process with the set of images that will most contribute to reach a faster and higher accuracy of the classifier. Then, consequently, it will improve the quality of the returned images in the CBIR process. It occurs because the classifier will be trained with the most informative (similar and uncertain) images.

Algorithm 1 and Fig. 2 present the main steps of our proposed approach. The dashed lines (Fig. 2) represent the cycle of the incremental learning process. In Step 1, given an image dataset $\mathcal{I}$, and a query image q, it is performed the selection of the best descriptor (best feature extractor and distance function pair). We analyzed several sets of feature extractors F_i and distance functions D_j (Algorithm 1, Line 1), due to their importance to the retrieval process. Afterwards, low-level features are extracted from $\mathcal{I}$, using the best feature extractor and generating the learning set $\mathcal{Z}_2$ (Algorithm 1, Line 2). We also extracted features from q, using the same extractor.

Considering Step 2, the learning set is partitioned in k clusters using a given clustering method. After the clustering process, it is generated the set of centroids $\mathcal{C}$ (Algorithm 1, Line 3). In addition to the images from $\mathcal{C}$, we selected the most similar images to q. To do so, we obtained the desired number of images from $\mathcal{L}_S$, which is ordered by an increasing order of distance from q (Algorithm 1, Lines 4 – 5). Then, those images are presented to the expert for annotation (as relevant/irrelevant). The annotated images constitute the initial training set $\mathcal{Z}_1$ (Algorithm 1, Line 6), which is used in the classifier training process, generating the first instance of the learning model $\mathcal{M}$ (Step 3, Line 7).

In Step 4, the current learning model actively participates in the process of selecting the most informative images to be used in its own training, and in order to improve the query result, returning more similar images. Therefore, the proposed selection criteria are based on uncertainty and similarity in relation to the query image q. So, the temporary training set $\mathcal{Z}'_1$ receives the most

Algorithm 1: Proposed Approach - MARRow.

input : query image q
output : final learning model $\mathcal{M}^{\Omega}$ and final list $\mathcal{L}_R^{\Omega}$ ordered by relevance (from the most similar to the least similar to q)
auxiliaries: image dataset I, sets of feature extractors F_i and distance functions D_j, learning set $\mathcal{Z}_2$, best feature extractor *BestFD.FeatureExtractor*, number of clusters k, set of centroids $\mathcal{C}$, ordered list $\mathcal{L}_S$ of the closest images to q, best distance function *BestFD.Distance*, number of desired samples *ns*, temporary training set $\mathcal{Z}'_1$, training set $\mathcal{Z}_1$, current learning model $\mathcal{M}$, number of selected samples *nu*.

1 *BestFD* ← findBestPairFeatureDistance(I, F_i, D_j);
2 $\mathcal{Z}_2$ ← featureExtraction(*BestFD.FeatureExtractor*);
3 $\mathcal{C}$ ← clusteringCentroids($\mathcal{Z}_2$, k);
4 $\mathcal{L}_S$ ← similaritySearch(q, $\mathcal{Z}_2$, *BestFD.Distance*);
5 $\mathcal{Z}'_1 \leftarrow \mathcal{C} \cup s_i \in \mathcal{L}_S$, $i = 1, 2, ..., ns$;
6 $\mathcal{Z}_1$ ← annotation($\mathcal{Z}'_1$);
7 $\mathcal{M}$ ← training($\mathcal{Z}_1$);
8 **repeat**
9 $\quad \mathcal{Z}'_1$ ← getALStrategy(*nu*, $\mathcal{M}$, $\mathcal{Z}_2 \backslash \mathcal{Z}_1$, q, *BestFD.Distance*);
10 $\quad \mathcal{Z}_1 \leftarrow \mathcal{Z}_1 \cup$ annotation($\mathcal{Z}'_1$);
11 $\quad \mathcal{M}$ ← reTraining($\mathcal{Z}_1$, $\mathcal{M}$);
12 **until** *satisfied*;
13 $\mathcal{M}^{\Omega} \leftarrow \mathcal{M}$;
14 $\mathcal{L}_R^{\Omega}$ ← sorting($\mathcal{Z}$, $\mathcal{M}_{\Omega}$), $\mathcal{L}_R^{\Omega} \subset I$;

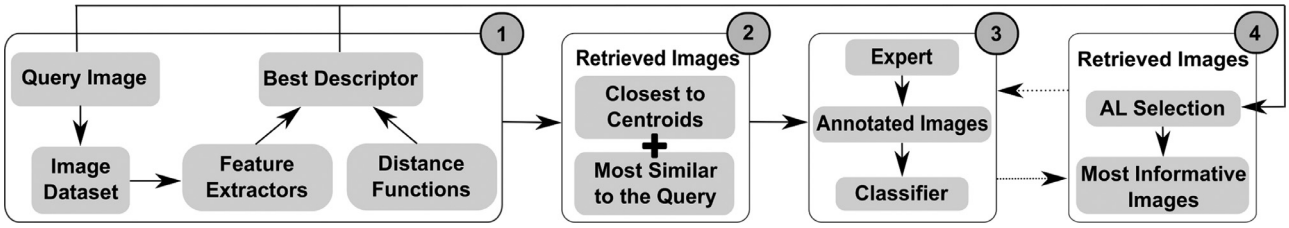


Fig. 2. Pipeline of the proposed approach.

informative images, obtained by our proposed active learning strategy (described by Algorithm 2).

Algorithm 2: Proposed Active Learning Strategy.

input : number of selected samples *nu*, current learning model $\mathcal{M}$, learning set $\mathcal{Z}_2 \backslash \mathcal{Z}_1$, query image q and best distance function *BestFD.Distance*
output : list of selected samples $\mathcal{L}_\mathcal{R}$
auxiliaries: learning set $\mathcal{Z}'_2$, ordered list of candidates $\mathcal{L}_\mathcal{C}$

1 **if** *size*$(\mathcal{Z}_2 \backslash \mathcal{Z}_1) < nu$ **then**
2 $\quad nu \leftarrow$ size$(\mathcal{Z}_2 \backslash \mathcal{Z}_1)$;
3 **end**
4 $\mathcal{Z}'_2 \leftarrow$ classifying$(\mathcal{Z}_2 \backslash \mathcal{Z}_1, \mathcal{M})$;
5 $\mathcal{L}_\mathcal{C} \leftarrow$ getSelection$(nu, \mathcal{Z}'_2, q, BestFD.Distance)$;
6 **while** $i \leq nu$ **do**
7 $\quad \mathcal{L}_\mathcal{R} \leftarrow \mathcal{L}_\mathcal{R} \cup s_i \in \mathcal{L}_\mathcal{C}$;
8 **end**

Then, the selected image set $\mathcal{Z}'_1$ is displayed to the expert. From the first learning iteration, this set of selected images are already previously labeled by the current instance of the model. So, the expert needs only to correct the labels of misclassified images (validating as relevant or irrelevant). The images confirmed and corrected properly by the expert are added to the previous training set $\mathcal{Z}_1$ (Algorithm 1, Line 10). It is important to emphasize that our strategy does not show samples that were already labeled ($\mathcal{Z}_2 \bigcap \mathcal{Z}_1 = \emptyset$ or in a simplified notation $\mathcal{Z}_2 \backslash \mathcal{Z}_1$). Afterwards, the training is performed again and a new instance of the model $\mathcal{M}$ is generated (Algorithm 1, Line 11).

Steps 3 and 4 (Algorithm 1, Lines 8 – 12) are repeated until the expert is satisfied with the results retrieved by the proposed learning process. Once satisfied, we can obtain a final learning model $\mathcal{M}^\Omega$ (which can be applied in an unlabeled dataset) and a final list ${\mathcal{L}_\mathcal{R}}^\Omega$ ordered by relevance (from the most similar to the least one) in relation to the query image q.

2.1. Active learning strategy

We also proposed a new active learning strategy (described by Algorithm 2) that selects a small set of more informative images. Our idea is to explore and use the knowledge of the classifier, which was obtained from the most informative samples, improving the image retrieval process.

Initially, Lines 1 – 3 from Algorithm 2 refer only to a control to verify if there are *nu* desired samples to be selected in the learning set $\mathcal{Z}_2 \backslash \mathcal{Z}_1$, since it is an iterative process, in which a set of images is selected at each iteration. So, the learning set is classified by the current learning model, generating the classified learning set $\mathcal{Z}'_2$ (Algorithm 2, Line 4). After the classification process, we obtain the list of the most informative (candidate) images $\mathcal{L}_\mathcal{C}$ (Algorithm 2, Line 5), according to the proposed selection criteria, which is deeply described by Algorithm 3. Then, we obtain the selected relevant set $\mathcal{L}_\mathcal{R}$, composed of the *nu* most informative (most uncertain and similar) samples from $\mathcal{L}_\mathcal{C}$ (Algorithm 2, Lines 6 – 8).

Algorithm 3: Proposed Selection Strategy.

input : number of selected samples *nu*,learning set $\mathcal{Z}'_2$,query image q and best distance function *BestFD.Distance*
output : list of selected candidate samples $\mathcal{L}_\mathcal{C}$.
auxiliaries: organized lists based on labels $\mathcal{L}_i$,nearest neighbor adjs,ordered lists of positive and negative samples $\mathcal{L}_\mathcal{P}$ and $\mathcal{L}_\mathcal{N}$, and list based on the center of mass of each label $\mathcal{L}com_i$.

1 $\mathcal{L}_\mathcal{C}, \mathcal{L}_i \leftarrow \emptyset$, i = {1, 0};
2 **for** *each* $s \in \mathcal{Z}'_2$ **do**
3 $\quad \mathcal{L}_i \leftarrow \mathcal{L}_i \cup \{s\}$, i = s.labelid;
4 $\quad$ adjs $\leftarrow$ get1NN(s);
5 $\quad$ **if** *s.labelid* $\neq$ *adjs.labelid* **then**
6 $\quad\quad \mathcal{L}_\mathcal{C} \leftarrow \mathcal{L}_\mathcal{C} \cup \{s, adjs\}$;
7 $\quad$ **end**
8 **end**
9 **if** $\mathcal{L}_\mathcal{C} \neq \emptyset$ **then**
10 $\quad \mathcal{L}_\mathcal{P} \leftarrow$ positiveSelection$(\mathcal{L}_\mathcal{C})$;
11 $\quad \mathcal{L}_\mathcal{N} \leftarrow$ negativeSelection$(\mathcal{L}_\mathcal{C})$;
12 $\quad \mathcal{L}_\mathcal{P} \leftarrow$ similaritySearch$(q, \mathcal{L}_\mathcal{P}, BestFD.Distance)$;
13 $\quad \mathcal{L}_\mathcal{N} \leftarrow$ similaritySearch$(q, \mathcal{L}_\mathcal{N}, BestFD.Distance)$;
14 $\quad \mathcal{L}_\mathcal{C} \leftarrow$ concatenation$(\mathcal{L}_\mathcal{P}, \mathcal{L}_\mathcal{N})$;
15 **end**
16 **if** *size*$(\mathcal{L}_\mathcal{C}) < nu$ **then**
17 $\quad \mathcal{L}com_i \leftarrow$ centerOfMass$(\mathcal{L}_i)$, i = {1, 0};
18 $\quad \mathcal{L}_\mathcal{P} \leftarrow$ decreasingOrder$(\mathcal{L}com_1, \mathcal{L}_1)$;
19 $\quad \mathcal{L}_\mathcal{N} \leftarrow$ decreasingOrder$(\mathcal{L}com_0, \mathcal{L}_0)$;
20 $\quad$ **repeat**
21 $\quad\quad \mathcal{L}_\mathcal{C} \leftarrow \mathcal{L}_\mathcal{C} \cup \{s, t\}$, $s \in \mathcal{L}_\mathcal{P}$ and $t \in \mathcal{L}_\mathcal{N}$;
22 $\quad$ **until** *size*$(\mathcal{L}_\mathcal{C}) = nc$;
23 **end**

For the proposed selection strategy (Algorithm 3), initially, an ordered list of candidates $\mathcal{L}_\mathcal{C}$ is created to receive the candidate samples to be displayed for the expert (Algorithm 3, Line 1). Learning lists $\mathcal{L}_i$, organized based on labels, are also created, where i denotes the ith class ($i = \{1, 0\}$, i.e. relevant and irrelevant classes, respectively). Then, each sample s from the learning set $\mathcal{Z}'_2$, which was previously labeled by the current learning model, is analyzed (Algorithm 3, Lines 2 – 8), in order to evaluate which ones are the most informative candidates. All samples are separated according to the class labels provided by the model. After that, each sample $s \in \mathcal{Z}'_2$ is stored in a list of labels $\mathcal{L}_i$, corresponding to their respective class label i ($i = s.labelid$) (Algorithm 3, Line 3).

So, we obtain the 1-nearest neighbor image (1-NN) to the image s, named as adjacent image *adjs* (Algorithm 3, Line 4). If the labels (provided by $\mathcal{M}$) from the images s and *adjs* are different (i.e. $s.labelid \neq adjs.labelid$, Line 5), both images are inserted into the list of candidates $\mathcal{L}_\mathcal{C}$ (Algorithm 3, Line 6). These images are

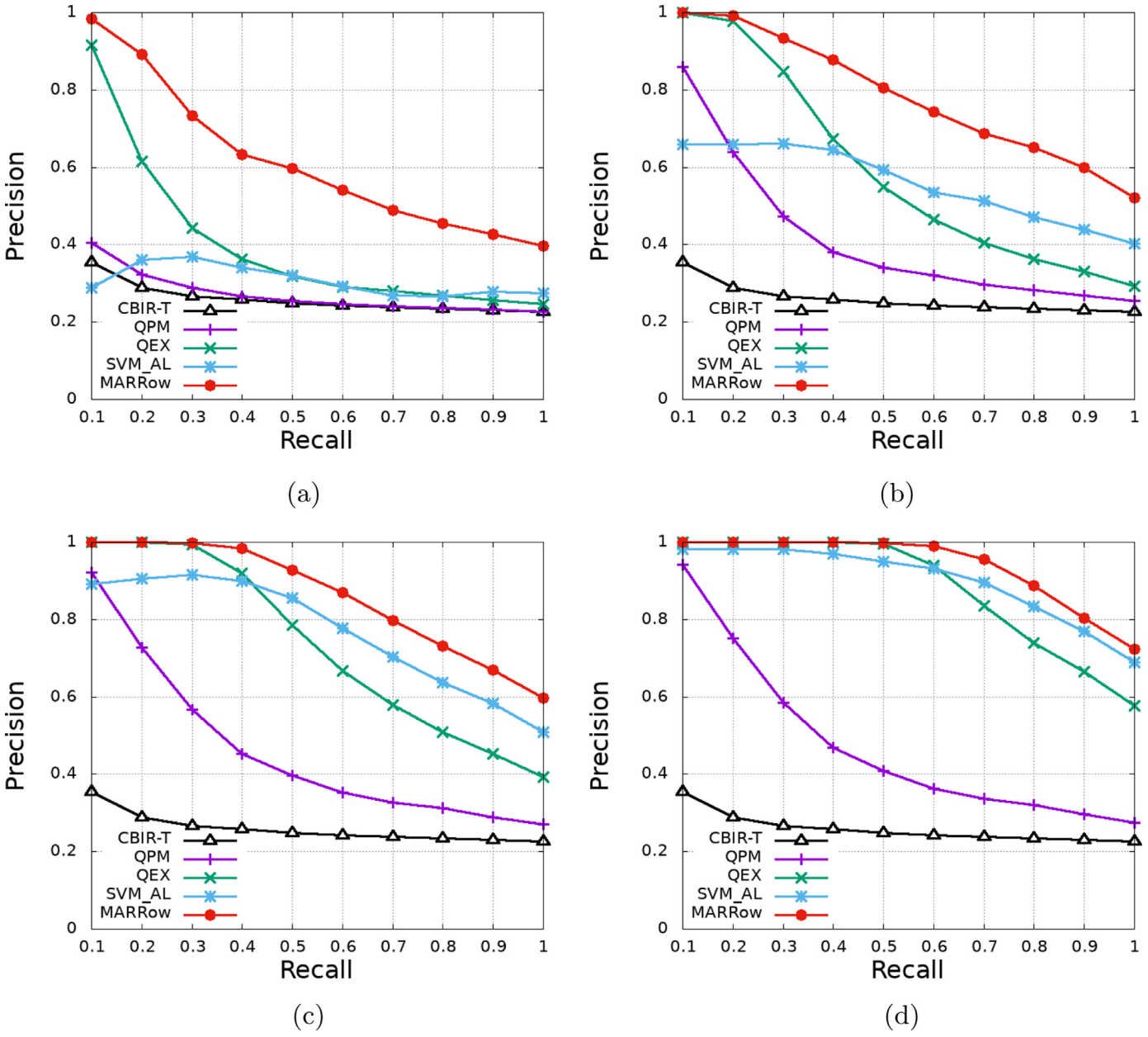


Fig. 3. P&R curves obtained by each approach over the I_1 dataset, considering the: (a) first, (b) third, (c) fifth and (d) eighth learning iterations.

considered the most informative (uncertain) candidate ones, since they are close to each other and have been labeled by the current instance of the model as being from different classes. These (diverse) samples can bring benefits rather than samples considered from the same class.

If the previous selection criterion is satisfied, the set of candidates $\mathcal{L}_\mathcal{C}$ is not empty (i.e. $\mathcal{L}_\mathcal{C} \neq \emptyset$, in Line 9). Then, samples from $\mathcal{L}_\mathcal{C}$ are separated in two auxiliary lists $\mathcal{L}_\mathcal{P}$ and $\mathcal{L}_\mathcal{N}$, containing the images classified as relevant (positive) and irrelevant (negative), respectively (Algorithm 3, Lines 10 – 11). Each sample from $\mathcal{L}_\mathcal{P}$ and $\mathcal{L}_\mathcal{N}$ are ordered by the distance to the query image q (Algorithm 3, Lines 12 – 13). So, L_C receives the concatenation of samples from $\mathcal{L}_\mathcal{P}$ and $\mathcal{L}_\mathcal{N}$ (Algorithm 3, Line 14).

If the desired number of images has not been obtained (i.e. size($\mathcal{L}_\mathcal{C} \leq nu$)), we considered another criterion for selecting images (Algorithm 3, Lines 16 – 23). The centers of mass com_i from each class i are located and stored in their corresponding list $\mathcal{L}com_i$ (Algorithm 3, Line 17). Afterwards, samples from $\mathcal{L}_i$, $i = \{1, 0\}$ (i.e. positive and negative samples) are organized in their respective lists ($\mathcal{L}_\mathcal{P}$ and $\mathcal{L}_\mathcal{N}$), in a descending order, according to the distances between them and the centers from $\mathcal{L}com_i$ (Algorithm 3, Lines 18 – 19).

Then, we can obtain a list of candidates $\mathcal{L}_\mathcal{C}$, selecting one image from each list $\mathcal{L}_\mathcal{P}$ and $\mathcal{L}_\mathcal{N}$, respectively, until the desired number of images is obtained (Algorithm 3, Lines 20 – 22). Finally, our selection strategy returns the set of the most informative (most uncertain and similar) candidates $\mathcal{L}_\mathcal{C}$.

3. Experiments

3.1. Datasets

The experiments were performed based on public image datasets from the MAMMOSET database [20], which is composed of regions of interest (ROIs) of mammograms from three datasets (VIENNA, MIAS and DDSM). The VIENNA dataset was created by

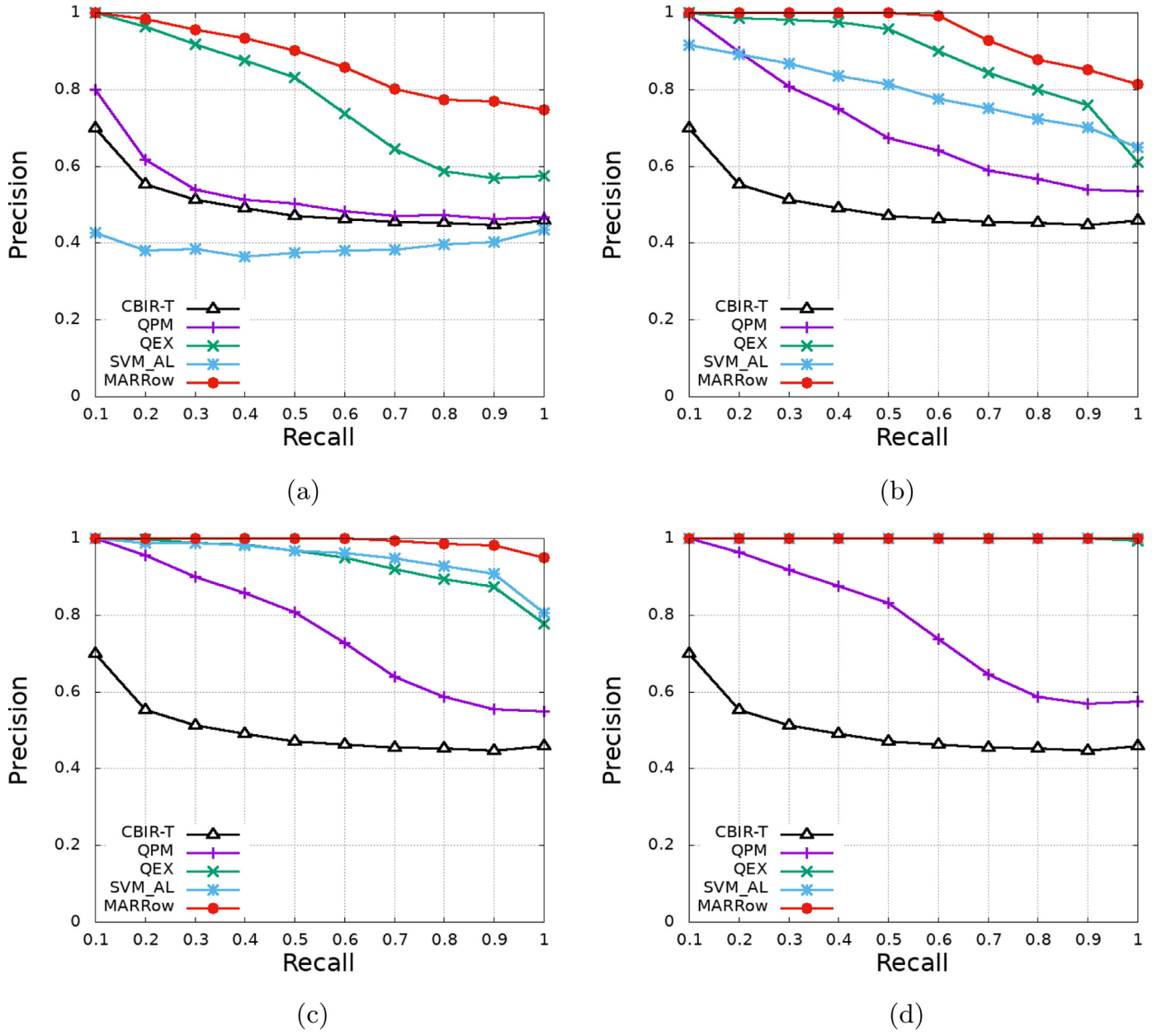


Fig. 4. P&R curves obtained by each approach over the I_2 dataset, considering the: (a) first, (b) third, (c) fifth and (d) eighth learning iterations.

the Department of Radiology at University of Vienna. It is composed of mammograms collected from the Breast Imaging Reporting and Data System (BI-RADS) Tutorium, which was carried out at the same university. The Mammographic Image Analysis Society (MINI-MIAS) repository is a reduced version of the MIAS dataset. The DDSM (Digital Dataset for Screening Mammography) repository is composed of medical breast images with 12 bits per pixel, organized into four categories based on the view of the breast image, which are (i) LCC: Left CranioCaudal, (ii) RCC: Right CranioCaudal, (iii) LMLO: Left MedioLateral Oblique, and (iv) RMLO: Right MedioLateral Oblique. For a more detailed description of the datasets see [20].

From these datasets, different subsets can be explored. Due to space constraints, we have selected some of them, in order to evaluate our approach in distinct complexity scenarios. Each subset considered in our experiments is named as I_{n_i}, where the higher the n_i, the higher (more challenging) the image dataset complexity. Tables 1–3 present the classes and the number of samples per class for each subset.

Table 1
Description of the dataset I_1 - VIENNA.

Classes	Images
birads3-calcification	33
birads4-calcification	113
birads5-calcification	32
birads3-mass	61
birads4-mass	72
birads5-mass	49
Normal	41

Table 2
Description of the datasets I_2 - MIAS, I_3 - DDSM, I_4 - MIAS-DDSM.

Classes	Images		
	MIAS	DDSM	MIAS-DDSM
Calcification-benign	14	615	629
Calcification-malign	12	547	559
Mass-benign	38	906	944
Mass-malign	20	824	844

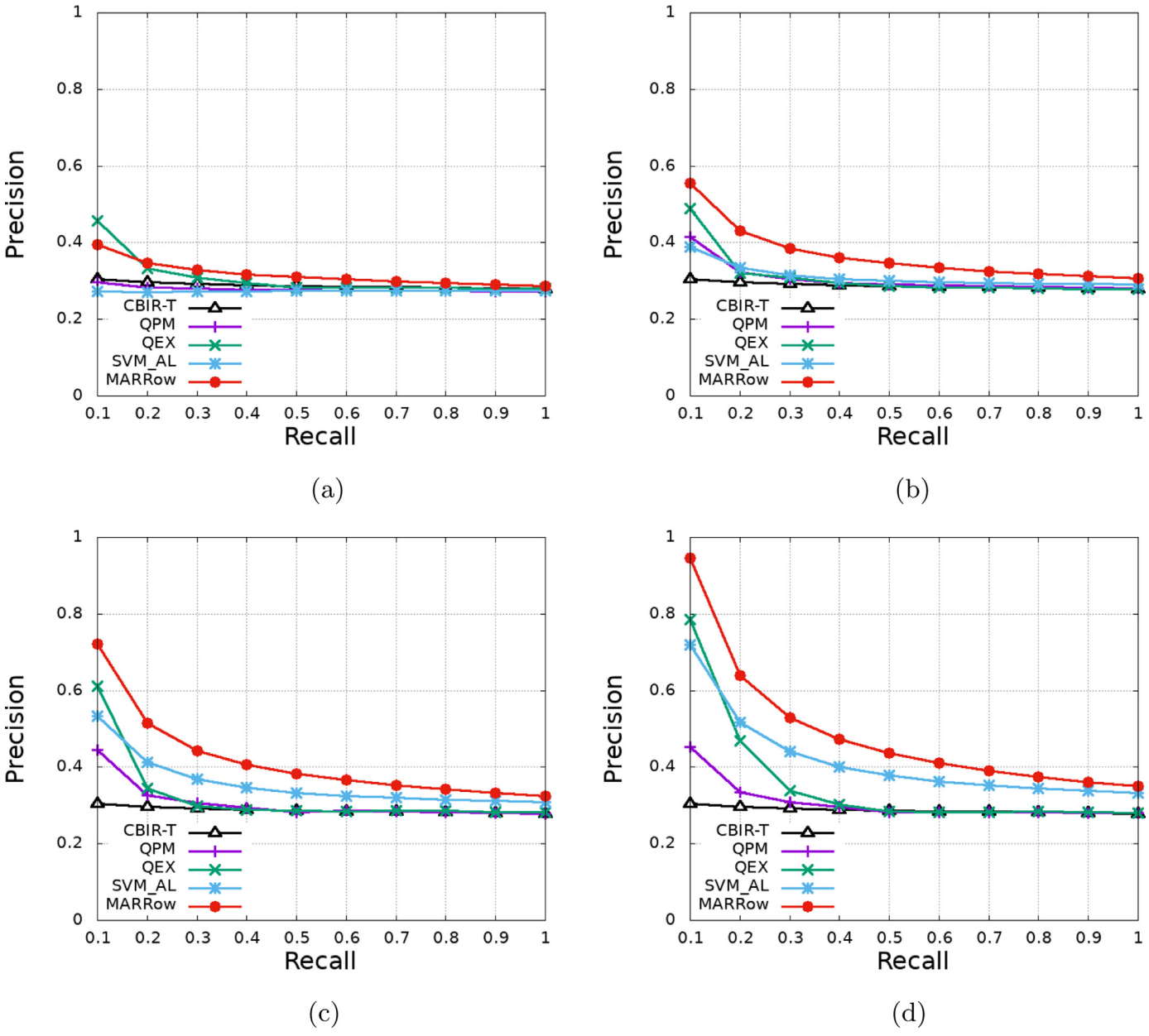


Fig. 5. P&R curves obtained by each approach over the I_3 dataset, considering the: (a) first, (b) third, (c) fifth and (d) eighth learning iterations.

Table 3
Description of the dataset I_5 - VIENNA-DDSM.

Classes	Images
birads1-calcification	1040
birads2-calcification	91
birads3-calcification	56
birads4-calcification	119
birads5-calcification	34
birads1-mass	1647
birads2-mass	58
birads3-mass	83
birads4-mass	75
birads5-mass	49
Normal	41

Table 4
Properties of each feature extractor applied to the datasets.

Feature extractor	Category	#Features
BIC Histogram [22]	Color	512
Edge Histogram [23]	Color	150
Normalized Histogram [24]	Color	256
Haralick [25]	Texture	24
Rotation Invariant LBP [26]	Texture	108
Texture Spectrum [27]	Texture	8
Daubechies [28]	Shape	16
Haar [28]	Shape	16
Zernike [29]	Shape	36

3.2. Scenarios

In order to corroborate the generalization and enable the replication of our approach, we used public medical image datasets described in Section 3.1. From these datasets, we extracted color, texture and shape-based features, detailed in Table 4, using different feature extractors. Each type of feature was compared with several distance functions, in order to obtain the best descriptor (a feature extractor joined with a distance function) to a given image dataset.

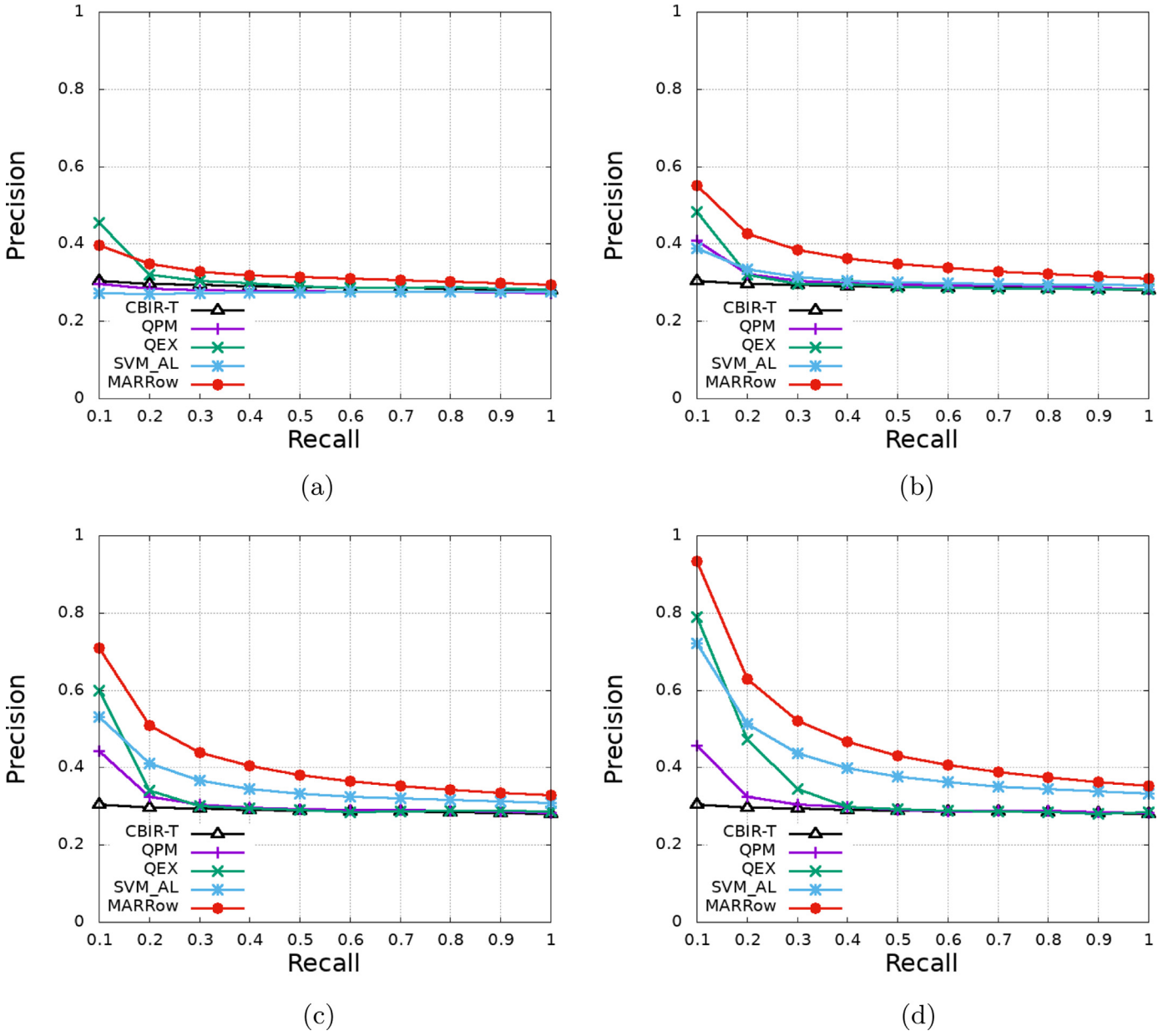


Fig. 6. P&R curves obtained by each approach over the I_4 dataset, considering the: (a) first, (b) third, (c) fifth and (d) eighth learning iterations.

To do so, seven different distance functions were considered: $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_\infty$, $\mathcal{X}_2$, Canberra, Jeffrey Divergence (JD), and dLog [21].

Learning strategies with RF in CBIR have been extensively studied to improve the semantic gap issue [11]. In order to show the efficacy of our approach, besides the Traditional CBIR (CBIR-T) without relevance feedback strategies, we presented comparisons with widely and well-known RF techniques, such as: Query Point Movement strategy (QPM) [16] and Query Expansion (QEX) [15]. QPM is based on the concept of moving the query center, throughout the iterations, towards more dense and relevant regions of the query space according to the expert intention. QEX promotes the dilation of the query aggregating to it new query centers.

Moreover, in order to improve the learning efficiency of the relevance feedback, active learning has been explored. Many active learning methods have been developed considering different selection criteria [18,19], and also applied in different classification tasks and domains. For instance, it is possible to choose samples near the decision boundary of a classifier [18,19]. The insight is to select the most diverse and uncertain samples which are close to the decision boundary of the classifier, demonstrating that they are the most difficult samples and, consequently, providing greater benefit to the model. In [18], it was proposed an active learning method with support vector machines (SVM-AL) for retrieval tasks. The method selects the samples which are closest to the classification boundary of the SVM classifier. There are also some latter research efforts [18,30,31]. However, they require the optimization of an objective function, resulting in high computational complexity. Then, in the present paper, we also presented comparisons with the well-known and pioneering SVM-AL proposed by Kremer et al. [18], which is closer to our proposed approach, since it fuses the active learning paradigm into the CBIR process.

Our approach can be instantiated considering any supervised classifier or clustering technique. However, the analysis of different classifiers and clustering techniques were not the main scope of the present work. Then, to generate the learning model, in our experiments, we used the *k*-Nearest Neighbor (*k*-NN) classifier. Regarding the clustering process, we considered the *k*-means technique.

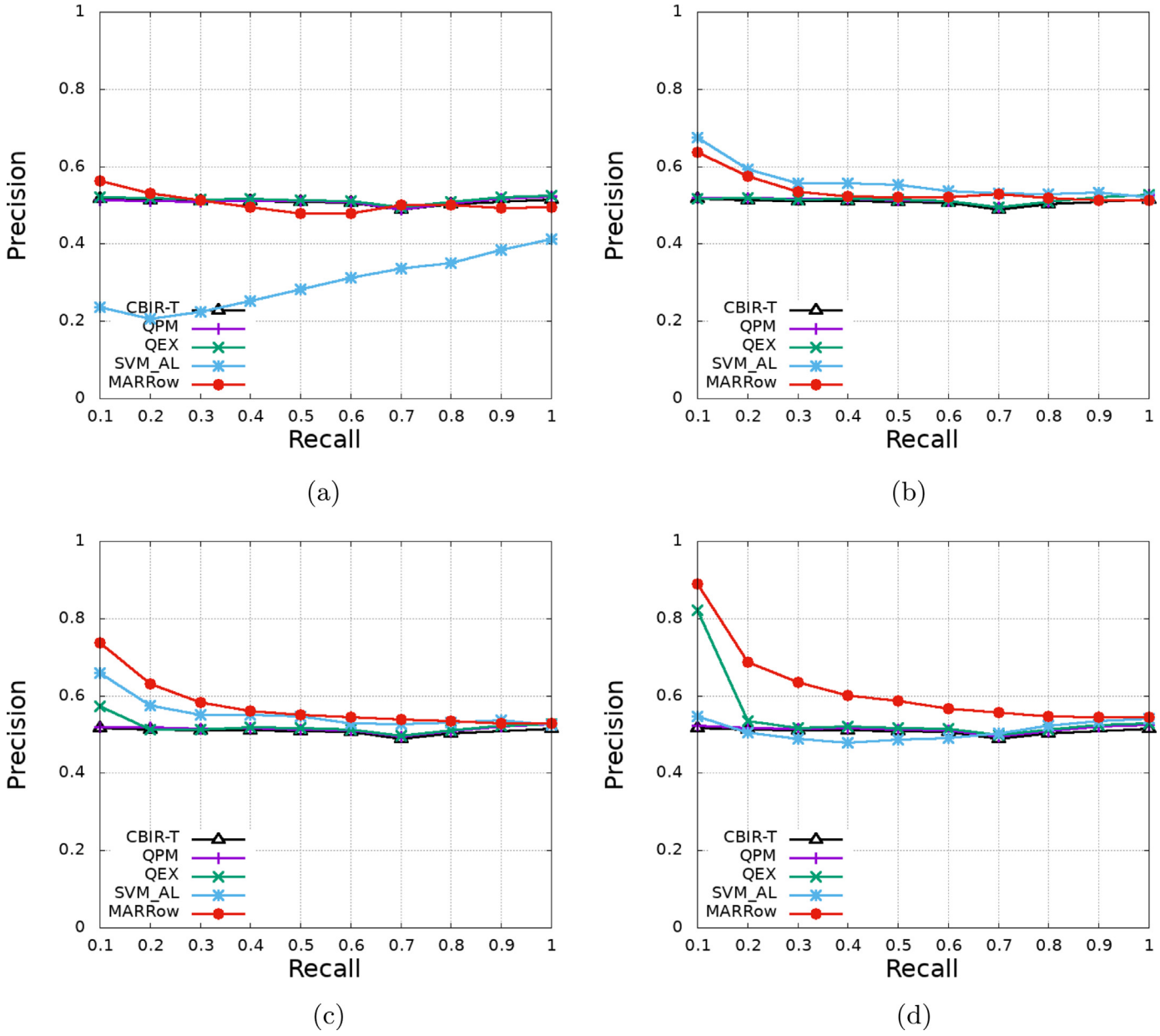


Fig. 7. P&R curves obtained by each approach over the I_5 dataset, considering the: (a) first, (b) third, (c) fifth and (d) eighth learning iterations.

To evaluate our proposed approach, we generated Precision and Recall (P&R) graphs [32]. As a rule of thumb, the closer the P&R curve to the top of the graphic, the better is the technique. To build the P&R graphs, we performed several similarity queries based on the *k*-nearest neighbor operator and randomly choosing the query images from the image datasets. The number of images retrieved by each similarity search was defined as 30 (based on daily medical practice routine). When a given image class contains less samples than 30, the value of *k* is set according to the number of samples of such class. To summarize the results, we employed the mean average precision (MAP), as defined in [32].

Table 5
Best distance function for each feature extractor and dataset $\mathcal{I}$. The highlighted data indicate the best descriptor.

	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}_3$	$\mathcal{I}_4$	$\mathcal{I}_5$
BIC	$\mathcal{L}_\infty$	$\mathcal{L}_2$	Canberra	Canberra	Canberra
Edge Histogram	JD	$\mathcal{L}_2$	Canberra	Canberra	JD
Norm. Histogram	$\mathcal{L}_\infty$	$\mathcal{L}_2$	Canberra	Canberra	Canberra
Haralick	JD	Canberra	$\mathcal{L}_2$	Canberra	Canberra
LBP	**JD**	$\boldsymbol{\mathcal{L}_2}$	$\mathcal{L}_\infty$	$\mathcal{L}_\infty$	$\mathcal{L}_\infty$
Texture Spectrum	Canberra	Canberra	$\mathcal{L}_2$	$\mathcal{L}_2$	$\mathcal{L}_\infty$
Daubechies	JD	$\mathcal{L}_2$	$\mathcal{L}_\infty$	$\mathcal{L}_\infty$	$\boldsymbol{\mathcal{L}_\infty}$
Haar	$\mathcal{L}_1$	Canberra	Canberra	dLog	Canberra
Zernike	$\mathcal{L}_\infty$	$\mathcal{L}_1$	$\boldsymbol{\mathcal{X}_2}$	$\boldsymbol{\mathcal{X}_2}$	$\mathcal{L}_1$

3.3. Results and discussion

Initially, we performed an analysis of the best descriptors for the mammographic image datasets. Table 5 presents the best distance function for each feature extractor, and the best descriptor highlighted for each image dataset. The best descriptors were LBP-JD, LBP-$\mathcal{L}_2$, Zernike-$\mathcal{X}_2$, Zernike-$\mathcal{X}_2$ and Daubechies-$\mathcal{L}_\infty$ for the datasets $I_1 - I_5$, respectively. Then, these descriptors were considered in the experiments to compare our proposed approach against the state-of-the-art ones.

Table 6
Overall MAP throughout the (1st–8th) iterations. Bold data correspond to the best results.

Dataset	QPM	QEX	SVM-AL	MARRow
I_1 - VIENNA	65.7	92.7	76.4	**97.0**
I_2 - MIAS	89.4	97.6	86.5	**99.7**
I_3 - DDSM	34.6	42.4	41.4	**53.3**
I_4 - MIAS-DDSM	34.4	42.1	41.3	**52.7**
I_5 - VIENNA-DDSM	51.7	54.8	53.3	**63.1**

Figs. 3–7 show the results comparing our proposed approach (MARRow) against CBIR-T, QPM, QEX and SVM-AL approaches over (a) first, (b) third, (c) fifth and (d) eighth iterations, using the datasets I_1–I_5, respectively. The precision obtained by the traditional CBIR (CBIR-T) approach was illustrated as a baseline in all graphs. MARRow presented higher precisions than the state-of-the-art approaches, considering all iterations and datasets. We can clearly note that MARRow also presented a well-suited and consistent increasing rate along the iterations.

For instance, in Fig. 3(a), MARRow presented a precision up to 1.8 and 2.4 times better when compared with QPM and SVM-AL, respectively. Besides, it reached precision gains of 87.3% at a recall level of 50% against QEX. Considering the first recall levels, our approach also presented good results against QEX, for instance, 7.6%, 45%, and 65.3%, at recalls from 10% to 30%, respectively.

The same behavior can be observed with the other datasets (Figs. 4–7), which represent different scenarios of complexity. As the complexity increases, they present considerably difficult classes (see Section 3.1), due to the intrinsic inter-class similarity, which leads to a harder separation between relevant and irrelevant images, and a fine-grained annotation process. Despite these issues, MARRow presented the best results in comparison with the other approaches. Analyzing the first iterations (Figs. 5–7), all approaches almost ties. However, at further iterations, MARRow presented a better and more consistent precision growth. This is a key ingredient of MARRow. While other approaches reach a saturation point, MARRow is capable to mitigate this problem throughout the iterations. It is possible to note that more naive approaches (e.g. QPM) presented a stronger saturation plateau (i.e. saddle point).

Summarizing the results, Table 6 presents the overall MAP obtained by each approach, throughout the (1st to 8th) learning iterations. According to our extensive experimental evaluation, MARRow presented the best precisions for all datasets. Through our approach it was also possible to minimize the computational time of the learning process, once it reduces the expert's involvement in the analysis and annotation process (reducing up to 88%). This reduction occurs because the expert does not need to annotate (correct) the labels of all samples, as required by the literature works. Our approach enables to obtain a more robust classifier (i.e. it has fewer misclassifications, as can be seen from the presented results, e.g. see Table 6), as more informative samples are selected for its learning.

4. Conclusion

In this paper, we proposed the MARRow (**M**edical **A**ctive lea**R**ning and **R**etrieval) approach, which aggregates AL and RF methods in the medical image domain. We also proposed a new AL strategy that was capable to be seamless integrated into the CBIR core process in order to mitigate several drawbacks, regarding the efficacy and efficiency of such domain. This is because the proposed AL strategy selects a small set of more informative images, considering selection criteria based on not only similarity, but also on certain degrees of diversity and uncertainty. These selected images can bring benefits rather than those from the same class usually considered by literature works.

From the experiments, it is straightforward to notice that our approach not only improves in a great extent the precision of the medical similarity queries, but also boosts the efficiency of the process. Our approach overcomes the other state-of-the-art approaches, reaching precision gains of up to 87.3%. The results testify that MARRow is feasible to be applied in challenging processes, such as medical image analysis. As future works, we intend to propose other AL strategies, in order to improve the selection of the most informative samples and then the quality of the retrieved images.

Conflict of interest

None.

Acknowledgments

This work has been supported by National Council for Scientific and Technological Development - CNPq (grants #431668/2016-7, #422811/2016-5); Coordination for the Improvement of Higher Education Personnel - CAPES; Fundação Araucária; SETI; and UTFPR.

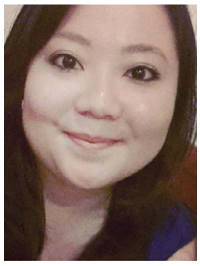
References

[1] C. Yan, L. Li, C. Zhang, B. Liu, Y. Zhang, Q. Dai, Cross-modality bridging and knowledge transferring for image understanding, IEEE Trans. Multimed. PP (2019) 1–10.
[2] C. Yan, Y. Tu, X. Wang, Y. Zhang, X. Hao, Y. Zhang, Q. Dai, Stat: Spatial-temporal attention mechanism for video captioning, IEEE Trans. Multimed. (2019).
[3] C. Yan, H. Xie, J. Chen, Z. Zha, X. Hao, Y. Zhang, Q. Dai, A fast uyghur text detector for complex background images, IEEE Trans. Multimed. 20 (12) (2018) 3389–3398.
[4] T. Turki, Z. Wei, Boosting support vector machines for cancer discrimination tasks, Comput. Biol. Med. 101 (2018) 236–249.
[5] I. Fondón, A. Sarmiento, A.I. Garca, M. Silvestre, C. Eloy, A. Polnia, P. Aguiar, Automatic classification of tissue malignancy for breast carcinoma diagnosis, Comput. Biol. Med. 96 (2018) 41–51.
[6] S. Liu, J. Zeng, H. Gong, H. Yang, J. Zhai, Y. Cao, J. Liu, Y. Luo, Y. Li, L. Maguire, X. Ding, Quantitative analysis of breast cancer diagnosis using a probabilistic modelling approach, Comput. Biol. Med. 92 (2018) 168–175.
[7] J. Chen, U. Scholz, R. Zhou, M. Lange, Lailaps-qsm: a restful api and java library for semantic query suggestions, PLOS Comput. Biol. 14 (3) (2018) 1–10.
[8] A. Kihm, L. Kaestner, C. Wagner, S. Quint, Classification of red blood cell shapes in flow using outlier tolerant machine learning, PLOS Comput. Biol. 14 (6) (2018) 1–15.
[9] G.H.B. Miranda, J.C. Felipe, Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization, Comput. Biol. Med. 64 (2015) 334–346.
[10] J.M. Tenório, A.D. Hummel, F.M. Cohrs, V.L. Sdepanian, I.T. Pisa, H. de Fátima Marin, Artificial intelligence techniques applied to the development of a decision support system for diagnosing celiac disease, Int. J. Med. Inf. 80 (11) (2011) 793–802.
[11] P. Malode, S.V. Gumaste, A review paper on content based image retrieval, Intl. Res. J. Eng. Technol. 20 (2015) 883–885.
[12] D.C.N.W. Uluwitige, S. Geva, G. Zuccon, V. Chandran, T. Chappell, Effective user relevance feedback for image retrieval with image signatures, in: Australasian Document Computing Symposium, ACM, 2016, pp. 49–56.
[13] L. Feng, S. Liu, Y. Xiao, Q. Hong, B. Wu, A novel CBIR system with WLLTSA and ULRGA, Neurocomputing 147 (2015) 509–522.
[14] X.-Y. Wang, Y.-W. Li, H.-Y. Yang, J.-W. Chen, An image retrieval scheme with relevance feedback using feature reconstruction and svm reclassification, Neurocomputing 127 (2014) 214–230.
[15] C. Carpineto, G. Romano, A survey of automatic query expansion in information retrieval, ACM Comput. Surv. 44 (1) (2012) 1:1–1:50.
[16] D. Liu, K.A. Hua, K. Vu, N. Yu, Fast query point movement techniques for large cbir systems, IEEE Trans. Knowl. Data Eng. 21 (5) (2009) 729–743.
[17] B. Settles, Active Learning Literature Survey, Technical Report, Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
[18] J. Kremer, K. Steenstrup Pedersen, C. Igel, Active Learning With Support Vector Machines, 4, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2014, pp. 313–326.
[19] M. Wang, X.-S. Hua, Active learning in multimedia annotation and retrieval: A survey, ACM Trans. Intell. Syst. Technol. 2 (2) (2011) 10:1–10:21.
[20] P.H. Oliveira, L.C. Scabora, M.T. Cazzolato, M.V.N. Bedo, A.J.M. Traina, C. Traina-Jr., MAMMOSET: An Enhanced Dataset of Mammograms, in: Satellite Events of the Brazilian Symposium on Databases, SBC, 2017, pp. 256–266.
[21] H. Samet, Foundations of Multidimensional and Metric Data Structures, Elsevier, Amsterdam, 2006.
[22] R.O. Stehling, M.A. Nascimento, A.X. Falcão, A compact and efficient image retrieval approach based on border/interior pixel classification, in: Intl. Conf. on Information and Knowledge Management, 2002, pp. 102–109.

[23] C.S. Won, D.K. Park, S.-J. Park, Efficient use of mpeg-7 edge histogram descriptor, Electron. Telecommun. Res.Instit. J. 24 (1) (2002) 23–30.
[24] M.S. Nixon, A.S. Aguado, Feature Extraction & Image Processing for Computer Vision, Academic Press, 2012.
[25] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. SMC-3 (6) (1973) 610–621.
[26] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recognit. 43 (3) (2010) 706–719.
[27] D.-C. He, L. Wang, Texture unit, texture spectrum, and texture analysis, IEEE Trans. Geosci. Remote Sensing 28 (4) (1990) 509–512.
[28] J.Z. Wang, G. Wiederhold, O. Firschein, S.X. Wei, Content-based image indexing and searching using daubechies' wavelets, Int. J. Digit. Libr. 1 (4) (1998) 311–328.
[29] A. Khotanzad, Y.H. Hong, Invariant image recognition by zernike moments, IEEE Trans. Pattern Anal. Mach.Intell. 12 (5) (1990) 489–497.
[30] S.C.H. Hoi, M.R. Lyu, A semi-supervised active learning framework for image retrieval, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 2005, pp. 302–309.
[31] L. Wang, K.L. Chan, Z. Zhang, Bootstrapping svm active learning by incorporating unlabelled images for image retrieval, in: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 1, 2003. I–I
[32] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology Behind Search, 2nd, Addison-Wesley Publishing Company, USA, 2011.

Rafael S. Bressan received his B.Sc. degree in Information Systems (2005) from the State University of the North of Parana (UENP), Brazil. In 2018, he received his M.Sc. in Computer Science from the Federal University of Technology - Parana, Brazil. Currently, he has been a Professor at the Department of Computing, University of Northern Parana (UNOPAR), Brazil. His research interests include image analysis, machine learning and pattern recognition.

Pedro H. Bugatti received a B.Sc. degree in computer science (2005) from the Euripides Soares da Rocha University of Marília, Brazil. In 2008, he received his M.Sc. in Computer Science from the University of São Paulo (ICMC-USP), Brazil. In 2012, he received his Ph.D. in Computer Science from the University of São Paulo (ICMC-USP), Brazil. Currently, he has been a Professor at the Department of Computing, Federal University of Technology - Parana (UTFPR), Brazil. His research interests include image analysis, machine learning, pattern recognition and content-based image retrieval.



Priscila T.M. Saito received a B.Sc. degree in computer science (2007) from the Euripides Soares da Rocha University of Marília, Brazil. In 2010, she received her M.Sc. in Computer Science from the University of São Paulo (ICMC-USP), Brazil. In 2014, she received her Ph.D. in Computer Science from the University of Campinas (IC-UNICAMP), Brazil. Currently, she has been a Professor at the Department of Computing, Federal University of Technology - Parana (UTFPR), Brazil. She is also a member of the PLOS ONE Editorial Board, acting as an academic editor. Her research interests include image analysis, machine learning and pattern recognition.