

Transferred Semantic Scores for Scalable Retrieval of Histopathological Breast Cancer Images

Elaheh Mahraban Nejad¹ · Lilly Suriani Affendey¹ · Rohaya Binti Latip¹ · Iskandar Bin Ishak¹ · Rasoul Banaeeyan²

Received: 22 April 2018 / Revised: 17 June 2018 / Accepted: 7 July 2018 / Published online: 24 July 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract

Content-based medical image retrieval (CBMIR) is an active field of research and a complementary decision support tool for the diagnosis of breast cancer. Current CBMIR systems employ hand-engineered image descriptors which are not effective enough at retrieval phase. Besides this drawback, the so-called semantic gap in the CBMIR is not still addressed leaving the room for further improvements. To fill in the two mentioned existing gaps, we proposed a new retrieval method which exploited a deep pre-trained convolutional neural network model to extract class-specific and patient-specific tumorous descriptor to firstly train a binary breast cancer classifier and then a multi-patient classifier aiming for reducing dimensions of the raw deeply transferred features and obtaining semantic scores which significantly enhanced the performance in terms of mean average precision. We evaluated the method on scalable BreakHis dataset of histopathological breast cancer images. After conducting five sets of experiments, results demonstrated the superior effectiveness of the proposed semantic-driven retrieval methods by means of increased mean average precision and decreased dimensionality and retrieval time. In overall, an improvement of 29.03% was obtained by the proposed class-driven semantic retrieval method.

Keywords Breast cancer retrieval · Transferred semantic score · Deep convolutional neural network · Histopathological image

1 Introduction

The field of medical image analysis has received a significant attention over the recent years due to the ever-increasing number of patients diagnosed with different diseases, particularly cancers. Breast cancer is the second most prevalent cancer among women, and the number of patients diagnosed

with breast cancer is growing each year (IARC). The recent decrease in death rate due to breast cancer [19] is the result of early diagnosis and treatment, which highlights the excessive need for fostering an efficient, reliable and automated diagnostic system. The growth in the amount of annotated digital medical images has facilitated automating the diagnostic systems, and thus, developing more robust supervised techniques is now possible, expediting the diagnosis time as well as diminishing the pathologist interaction with the system. Owing to the applications of CBMIR (content-based medical image retrieval) systems [1, 8, 18], diagnosis can be performed with the aid of retrieval systems. In [6], a recent review of CBMIR systems is presented which highlights the recent advances are achieved by employing deep learning techniques introduced in [11].

CBMIR is an effective way of supplementing the diagnosis and treatment of breast cancer; their performance largely depends on how abstract the image descriptors are. Recently, deep Learning [11] methods—particularly convolutional neural networks—were employed in CBMIR task, [16, 24, 27], to obtain state-of-the-art results, but a few works

✉ Elaheh Mahraban Nejad
mahrabannejad@gmail.com

Lilly Suriani Affendey
lilly@upm.edu.my

Rohaya Binti Latip
rohayalt@upm.edu.my

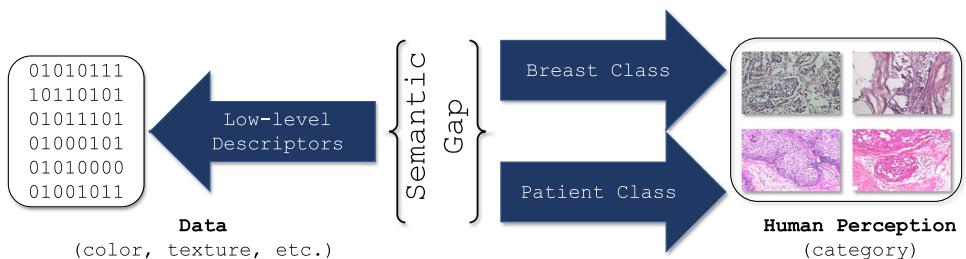
Iskandar Bin Ishak
iskandar_i@upm.edu.my

Rasoul Banaeeyan
banaeeyan@gmail.com

¹ Faculty of Computer Science and Information Technology,
University Putra Malaysia, Selangor, Malaysia

² Faculty of Engineering, Multimedia University, Cyberjaya,
Malaysia

Fig. 1 Visualization of the so-called semantic gap between low-level representation of the histopathological image data and the high-level perception of human in the form of classes/categories



have been conducted to enhance the retrieval performance on histopathological images [18].

Although the recent CBMIR systems have attained a good performance at retrieving the most relevant cases among medical images, they are still not able to fill in the so-called semantic gap between low-level descriptors (color, texture, etc.) and high-level perception of humans (classes) in the task of BC image retrieval. The presented challenge is visualized in Fig. 1.

In this paper, we, for the first time, conduct CBMIR on a new large-scale breast cancer (BC) histopathological image dataset introduced by [21] and gradually enhance the retrieval performance by performing a set of four experiments, out of which three methods utilize a deep pre-trained convolutional neural network (CNN) model to obtain abstract representation of BC images and consequently use these descriptors to train two ensemble semantic classifiers aiming to bridge the semantic gap on the BreakHis dataset. The main contribution of this work is twofold.

- We performed retrieval on the scalable BreakHis histopathological dataset of breast cancer images using both hand-engineered and deep transferred descriptors. We not only enhanced the retrieval performance by 8%, but also reduced the dimensionality by more than four times (from 4130 to 1000-D).
- We also proposed a semantic retrieval method which further improved the previous best performance by 29.03% and simultaneously reduced the dimensionality, sequentially, from 1000 to 82-D, and ultimately to 2-D with a superior enhancement in execution time from 250 to 95 μ s (more than 2.5 \times faster).

This paper is organized as follows. Section 2 reviews existing works and recent advances in medical image retrieval systems. Section 3 explains the research methods as well as the details of the conducted experiments. Section 4 demonstrates the experimental designs. Section 5 presents the results and provides a discussion on them. At the end, Sect. 6 summarizes the paper and suggests the further work.

2 Related work

Conventionally, low-level handcrafted features used to be employed in medical image retrieval systems; for instance,

the work in [30] proposed an scalable CBMIR system in which SIFT (scale-invariant feature transform) features [14] were extracted from breast cancer histopathological images and, then, they were used as input into a supervised kernel hashing technique to map 10,000 dimensional features to only ten binary bits for the retrieval purpose. This work achieved a classification accuracy of 88.1%, an acceptable retrieval time and also retrieval precision mean average precision (mAP) of 83.3. In [28], a retrieval system was proposed called PD-LST which was constructed based on BOVW (bag of visual words) and focused on dictionary pruning to identify discriminative characteristics among images. The evaluation results showed an improved retrieval performance as well as good efficiency at medical image retrieval phase. In [9], a retrieval system was proposed which extracted local features and also used BoW to quantize raw low-level descriptors; then, the quantized features were mapped into a two-dimensional inverted index file. The experiment was conducted on a digital database of screening mammography images, and enhancement on retrieval precision and diagnostic was observed. Later, in [29], a graph-based query framework was introduced to fuse the holistic and local features for a new retrieval technique which used unsupervised approach. The evaluation was performed on captured images of intraductal breast lesions, and its classification accuracy was reported as 91.67%; a second retrieval experiment on a second dataset was also performed which showed efficiency and scalability of the technique. In [27], a framework was proposed which was consisted of two components as (1) classification and (2) retrieval on ImageCLEF medical images. Regarding the classification task, SIFT and LBP descriptors were used in image-based and patch-based features; integration of SIFT and LBP in image-based resulted in an improved accuracy of 92%, and this fusion led to the state-of-the-art retrieval performance. In [24], the authors used pre-trained VGG-16 [20] to propose the framework for retrieval task. In this work, semantic descriptors—obtained by output of the last fully connected layer of a CNN model (softmax layer)—and a secondary vector of descriptors obtained from the layer before softmax were simultaneously utilized to conduct retrieval on three different datasets; as a result, enhanced computational time and mean average precision (mAP) were achieved. In [22], a scalable CBIR (content-based image

retrieval) system was proposed based on images of mammographic masses. The handcrafted SIFT features were extracted for a mammographic region of a queried image, and also contextual information was taken advantage to construct a vocabulary tree to refine the weight of tree nodes; the final results proved the ability of the system in obtaining a better retrieval precision, classification accuracy and scalability on a large-scale database of screening mammography. In [4], a CBMIR was proposed to support the decision of medical practitioners; features extraction was performed by utilizing an enhanced GLCM; as a result, retrieval precision was enhanced by employing ARM in terms of precision, recall and accuracy. In [17], a retrieval system was proposed aiming for content-based prostate MR image retrieval using a deeply learned hashing forest (DL-HF). The system was a generic system which could be generalized to other imaging modalities without requiring any image preprocessing or normalization and any dependence on user responses. In [16], a deep convolutional network was used for CBIR, which was trained for classification on multimodal body organ medical images; the learned features and classification output were employed in retrieval task which resulted in 99.77% accuracy in classification task and a mean average precision of 0.69 in retrieval. In [31], a method was presented for the task of medical sign recognition as a complementary decision support for the diagnosis of lung lesions; it was proposed to use deep learning in features extraction phase and apply supervised hashing methods to reduce the high-dimensional semantic features; also a weighted hamming distance was used at retrieval part. Finally, they achieved a mAP of 87.29% with a hash code length of 48 bits. In [25], a patch-based multimodal retrieval method was proposed which works based on a fusion of visual and text information. A deep convolutional neural network model was designed to obtain visual information, and an embedded pathology report text vector was used to train the model to reach the best precision as 0.54 in an iterative approach. In [5], a deep robust hashing framework was proposed, called RMIH, as an aid to decrease dimensions of features for retrieval purpose which works under a multiple-instance setting; finally, after the evaluation was conducted on mammography images, an enhanced retrieval performance was observed. Later in [10], it was shown that a pre-trained network was able to achieve better retrieval results than a task-specific trained CNN model; in other words, the authors showed that a low number of training images are not sufficient to train a CNN model from scratch, and instead relying on a pre-trained model would lead to better classification retrieval results. In [13], a retrieval system was proposed which used unsupervised hashing method to generate compact binary codes; based upon these codes, a retrieval precision of 84.00 and a classification of accuracy of 93% were achieved. In [26], a generic medical image retrieval system was proposed based on capturing visual

features of images which was done by descriptors such as correlogram, LGP, wavelet moments, etc. Then, at the retrieval time, several weighting schemes were applied based upon fractional brain storm optimization (FBSO), as well as fractional lion algorithm (FLA). In another recent work [18], a novel approach was proposed to encode the cell-level information of medical images for the construction of a binary code representation of that image in an unsupervised manner. The study in [3] presented a medical image retrieval technique based on image segmentation. A multi-panel medical image segmentation framework (MMISF) was proposed to divide a given image into different sub-images, and then by locating the largest inter-panel border, the accuracy and efficiency of the system were enhanced. In [12], a texture-based similarity point measurement technique was proposed to address the issue of low-tolerance measurements in the task of medical image retrieval. A clustered-based CNN model was proposed in [1] for the accurate classification and retrieval of endoscopic images. To this end, convolutional kernels of a pre-trained CNN model were analyzed and clustered into different groups based on their sensitivity to color and texture information; later the most informative descriptors were preserved using a spatial maximal activator pooling (SMAP) approach resulting in a compact and distinctive set of features. Later in a research by [8], local pattern descriptors (LPD) and gray-level co-occurrence matrix (GLCM) descriptors were extracted from MRI brain images to enhance the brain image retrieval system by presenting a texture-based feature vector as a fusion of these two. In [2], a low-dimensional binary codes construction technique was introduced in order to retrieve medical images more efficiently. The target compact codes were obtained by selecting among image features taken from a fully connected layer of a CNN model by employing fast Fourier transform (FFT). Lastly, the study in [23] presented a multi-texton representation technique for the purpose of medical image retrieval which uses filter bank responses in the local coordinate system of images as well as spatial pyramid matching technique; the superior performance of their method was later demonstrated through multiple experiments. Later, in [16] a CNN was designed in order to train a 24-class model for classification of radiology images of different body organs, for example liver, brain, lung. Then, the image features were extracted from the last fully connected layer (FC 3) of this trained CNN model to be employed for medical image retrieval purpose.

3 Methodology

The overview of retrieval framework is illustrated in Fig. 2 which consists of two components as off-line (in gray area) and online (in green area). During off-line phase, first, we

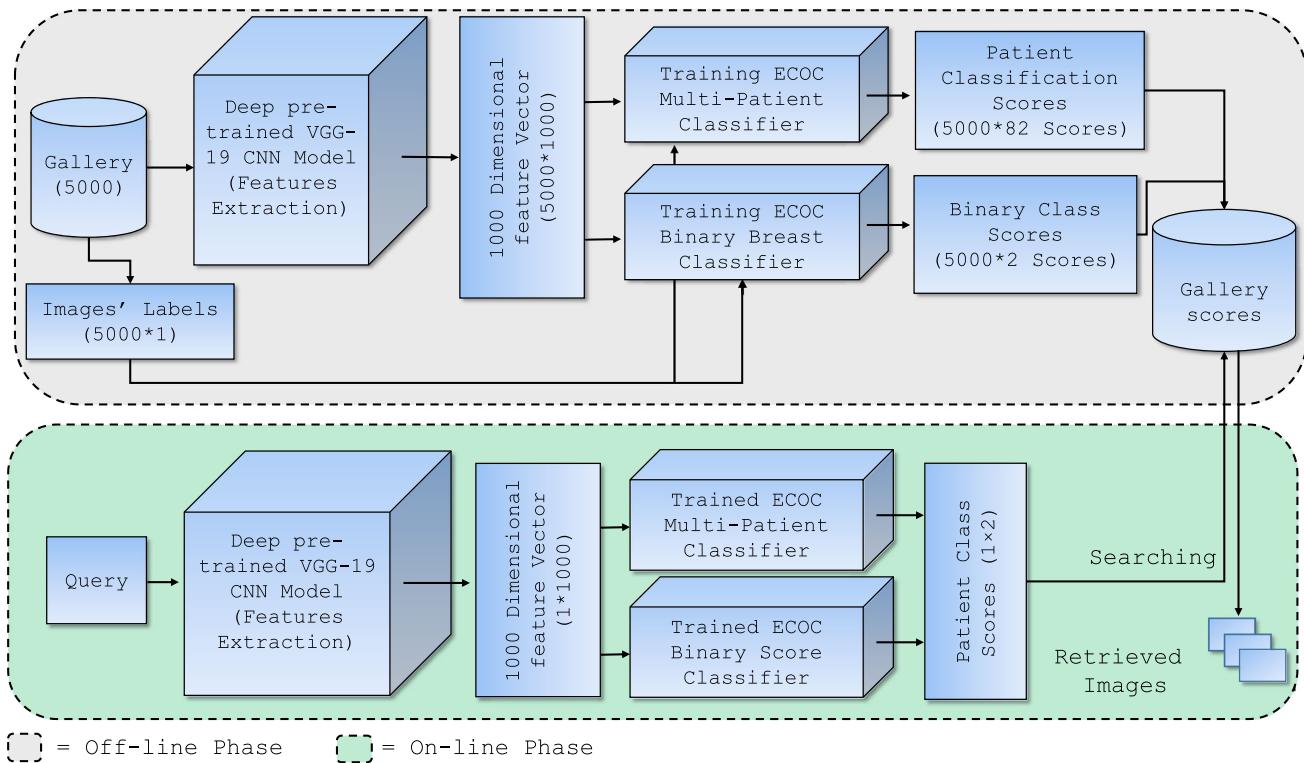


Fig. 2 Overview of the semantic diagram of the proposed retrieval system. Gray is showing off-line process and green showing online section

extract representative features from a deep pre-trained CNN model; second, we reduce the dimensionality of the raw descriptors by training two ensemble classifiers: (1) a binary classifier model to obtain a low-dimensional (2-D) classification score for breast cancer images and (2) a multi-patient classifier to obtain patient scores of a given query image (82-D). Through the online section, semantic-driven scores obtained from the off-line section are employed to perform a more precise and faster BC image retrieval.

3.1 Feature extraction

In this study, we employed the VGG-19 pre-trained deep CNN model as described in [20] which is trained on over 1M images of 1K different object classes to be used as a fixed feature extractor; therefore, tumors descriptors from breast cancer histopathological images are obtained from the softmax layer of this CNN model. Since acceptable input image size for VGG-19 is 224×224 , all original images are resized from 700×460 into 224×224 followed by a zero-mean image normalization. Figure 3 illustrates feature extraction process by employing VGG-19 in detail. Resized RGB images are fed to the deep pre-trained model as input, and then multiple series of convolutional filters are applied on them in each layer to generate the most abstract and distinctive representative information from the FC layer of the model. The

final output is a 1000-D transferred descriptors (probabilities) taken form the softmax layer.

3.2 Semantic patient/cancer scores

We take the advantage of ensemble Error Correction Output Code (ECOC) learning model as proposed in [7] to train the multi-patient and cancer classifiers, which reduces the dimensionality of transferred features from 1000-D into 82-/2-D patient-/class-driven semantic scores as illustrated in Fig. 4. ECOC starts by training several binary SVM (support vector machines) models which contains L learner as formulated in Eq. (1):

$$L = (P \times (P - 1))/2, \quad (1)$$

where P refers to the total number of patients ($P = 82$ in this study) in BreakHis dataset.

3.3 Retrieval

At retrieval time (online phase), for each image query, its semantic scores are obtained using the approach in the previous subsection, and then similarity distance between the query scores and that between gallery scores are measured in terms of Euclidean distance as given in Eq. 2.

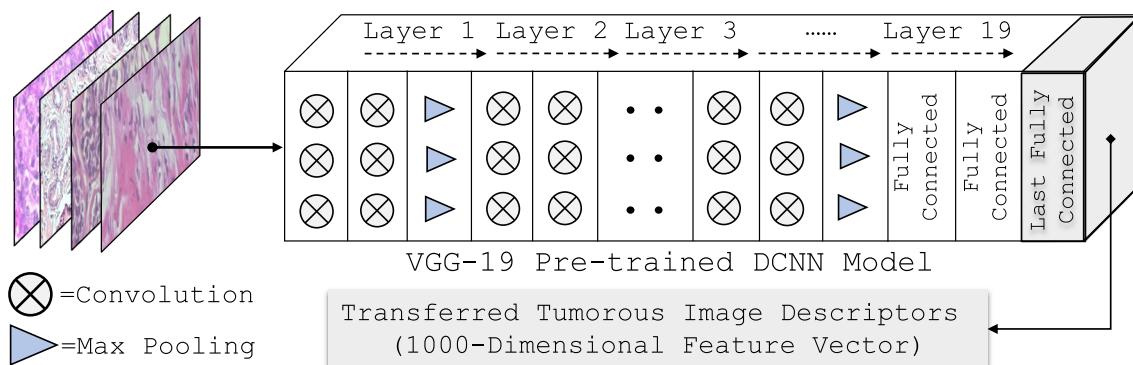
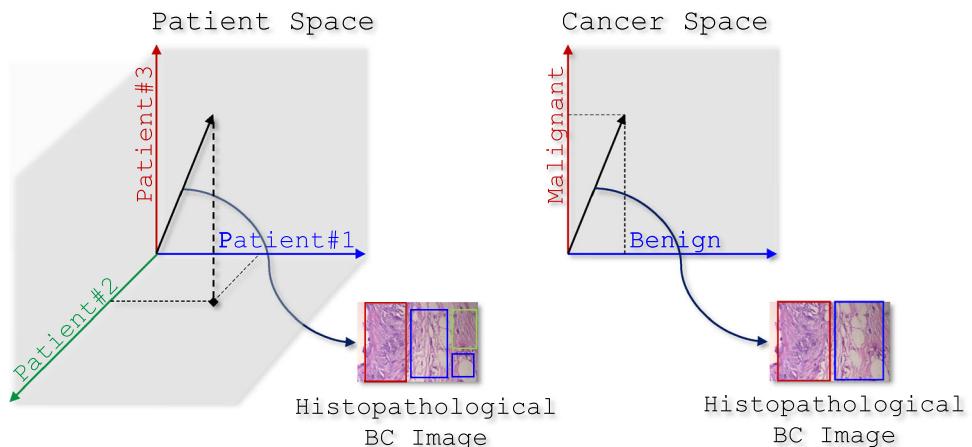


Fig. 3 Illustration of the VGG-19 pre-trained deep CNN model for extraction of transferred distinctive features

Fig. 4 Illustration of a sample BC image in two different semantic spaces: patient space and cancer space. Please notice that in this figure patient space is 3-D for the purpose of visualization, but the real patient vector employed in this research has 82 dimensions



$$d(q, g) = \sqrt{\sum_{i=1}^M (q - g_i)^2}, \quad (2)$$

where function $d(\cdot)$ calculates the distance between the query scores (q) and scores of the i th gallery image (g_i), and M is the total number of images in the gallery.

4 Experiment

4.1 BreakHis dataset

In this study, we used BreakHis dataset which has been recently introduced in [21]. This dataset consists of 7,909 microscopic breast cancer histopathological images collected from 82 patients and categorized into two classes of benign and malignant in which each class holds four different subtypes (eight subtypes in total). The BC images are also taken under different magnification factors as $40\times$, $100\times$, $200\times$ and $400\times$. The distribution of BC images with respect to each subtype and magnification rate is detailed in Table 1. The query set consists of 160 images as detailed in the next

Table 1 Distribution of all histopathological BC images in BreakHis dataset [21] over four magnification factors and two class types

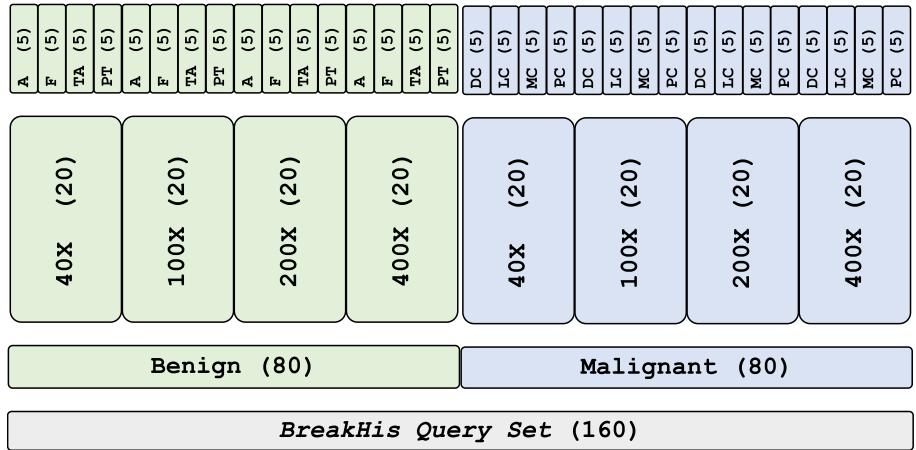
Magnification	Benign	Malignant	Total
$40\times$	625	1370	1995
$100\times$	644	1437	2081
$200\times$	623	1390	2013
$400\times$	588	1232	1820
Total	2480	5429	7909
No. of patients	24	58	82

subsection, the training set consists of 2749 and the remaining 5000 were used as gallery.

4.2 Query selection

The overall procedure for query selection is demonstrated in Fig. 5. To evaluate the proposed semantic BC image retrieval systems, 160 query images are randomly selected from dataset in a way that represents equal number of benign and malignant cases (each cancer class contains 80 images which are highlighted by green and blue in Fig. 5). Analogously, each cancer class represents all the four different

Fig. 5 Demonstration of image selection for construction of the query set from the BreakHis dataset [21]. The numbers in blue area represent distribution of malignant images in the query set with respect to each magnification rate and subtype, and the numbers in green indicate statistics of benign images in the query set (color figure online)



magnification factors while holding a fixed number of 20 images per magnification rate, and each set of images—belonging to different magnification rate—represents all the four variations in cancer subtypes by including five images per subtype. Overall, of 7909 images, 160 are taken out as the query set and the remaining 7749 are kept as gallery set.

4.3 Performance metric

Mean average precision (mAP) and precision at rank @ k are used as performance metrics. Average precision (AP) is formulated in Eq. (3).

$$AP = \frac{1}{N} \sum_{k=1}^N \frac{TP_k}{TP_k + FP_k} \quad (3)$$

where N represents the total number of retrieved images, TP is the true positive and FP is the false positive and k denotes different retrieval ranks.

For the evaluation of the proposed retrieval systems, five different experiments are conducted using the same experimental procedure on the same dataset. The first method evaluates the system using 4130-D hand-engineered LBP [15] descriptors, the second uses 1000-D deep transferred descriptors, the third utilizes patient-driven semantic scores with dimensionality of 82, the fourth uses the state-of-the-art medical image retrieval technique proposed in [16] and finally the fifth method is based on class-driven semantic scores of dimension 2.

5 Results and discussion

At this section, we present the experimental result as reported in Table 2 followed by a discussion on results with the aid of the mAP-recall curves presented in Fig. 8. Several sample

retrieval results of each method proposed in this paper are also visualized in Fig. 6.

5.1 Experimental results

According to Table 2, best retrieval performances in terms of mAP, P at 10, P at 20, P at 30, dimensionality size and retrieval time were achieved by method 4 with a significant difference in mAP and descriptor length as compared to first three methods, with exception of the forth method ([16]) which has the same descriptor size of two as the first method. Through performing four retrieval experiments, mAP was incrementally enhanced by 8.00%, 12.96%, 1.64 and ultimately reached the best highest improvement of 29.03%, respectively, from methods 1 to 2, methods 2 to 3, methods 3 to 4, and methods 4 to 5. Unlike method 4, the performance results reported by method 1 were the lowest regarding all the performance metrics used for the evaluation. Comparing method 2 and method 3, it was observed that although there was a difference of 0.05 in the mAP, their precision at ranks 10, 20 and 30 were almost the same with the marginal values of 0.01, 0.02 and 0.03, respectively. However, regarding the length of the descriptors, method 3 has a largely reduced size of 82 which is approximately 12 times more compact than the size of descriptor employed by method 2.

5.2 Discussion on results

The most important observation in Table 2 is the superior retrieval results of the 2-D class-driven semantic retrieval over the 82-D patient-driven semantic retrieval and the method in [16] (forth method) since—by assumption—the higher descriptor size is expected to lead to improved retrieval which is not the case in this observation though all three types of descriptors are semantic driven. This implies the fact that the semantic gap (as visualized in Fig. 1) is best addressed by relying on class categories than the patient categories, while

Table 2 Demonstration of mAP results, dimensionality sizes (Dim) and retrieval time (R.T) of five conducted experiments on BreakHis dataset [21]

Method	mAP	P@10	P@20	P@30	Dim	R.T (μ s)
Method 1. LBP	0.50	0.62	0.59	0.57	4130	410
Method 2. P.CNN	0.54	0.80	0.77	0.74	1000	250
Method 3. Sem.P	0.61	0.81	0.79	0.77	82	110
Method 4. T.CNN	0.62	0.71	0.69	0.67	2	95
Method 5. Sem.C	0.80	0.83	0.82	0.81	2	95

Retrieval performances are also represented at three different ranks as P@10, P@20 and P@30. The best results are presented in bold. 1—LBP local binary pattern descriptors, 2—P.CNN pre-trained CNN descriptors, 3—Sem.P semantic patient scores, 4—T.CNN trained CNN descriptors same as in [16] and 5—Sem.C semantic class scores

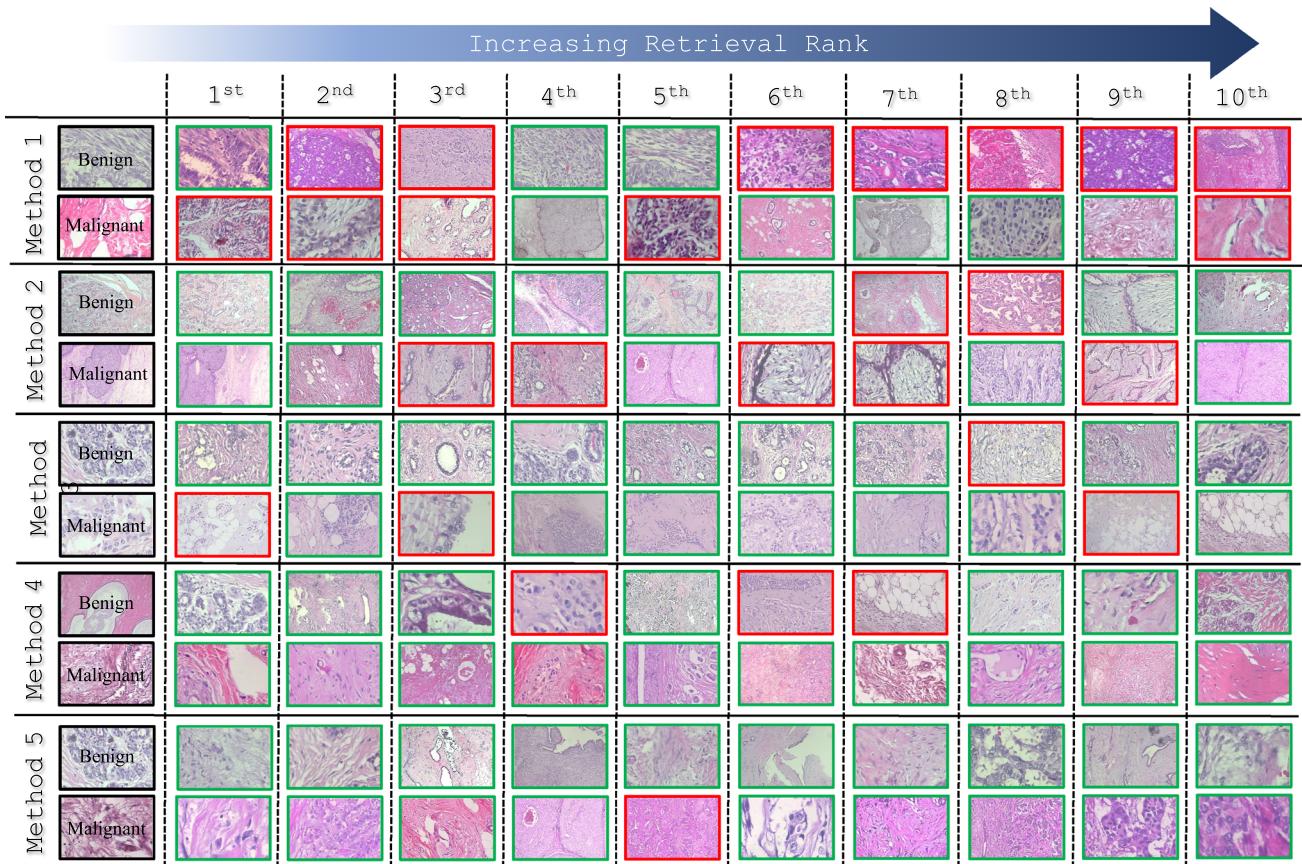


Fig. 6 Visualization of some randomly selected query images. For each method, one benign and one malignant are selected as query; the retrieval results are, then, represented in front of each query while TP (true positive) is highlighted by green rectangles and FP (false positive) by red rectangles

patient-driven semantic retrieval still performs better than the deep transferred descriptors.

As seen in Table 2, the enhancement (decrease) of retrieval time is the direct result of reduction in descriptor size, but its relationship is not linearly related to the size as illustrated in Fig. 7. Another interesting fact is the enhanced performances obtained at P@10, P@20 and P@30 with respect to methods 2 and 3 compared to method 4 while method 4 achieves better overall precision (mAP) in contrast to methods 2 and 3. The improved performance at lower recalls for methods 2 and 3 can specify the potential application of these methods

as a complementary decision support since a good medical image retrieval should have a good precision in early recall rates. This is further supported by the retrieval curves as in Fig. 8. The same figure also shows the early convergence of the retrieval curves for the first four methods which highlights the fact that at higher recall rates their performance becomes almost identical—particularly methods 1 and 2 which began converging earlier. Using the same analogy, the superior performance of the method 5 can be concluded because of its late convergence toward low precision even at higher recall rates.

Fig. 7 Illustration of effect of descriptor size on the retrieval time in our experiments

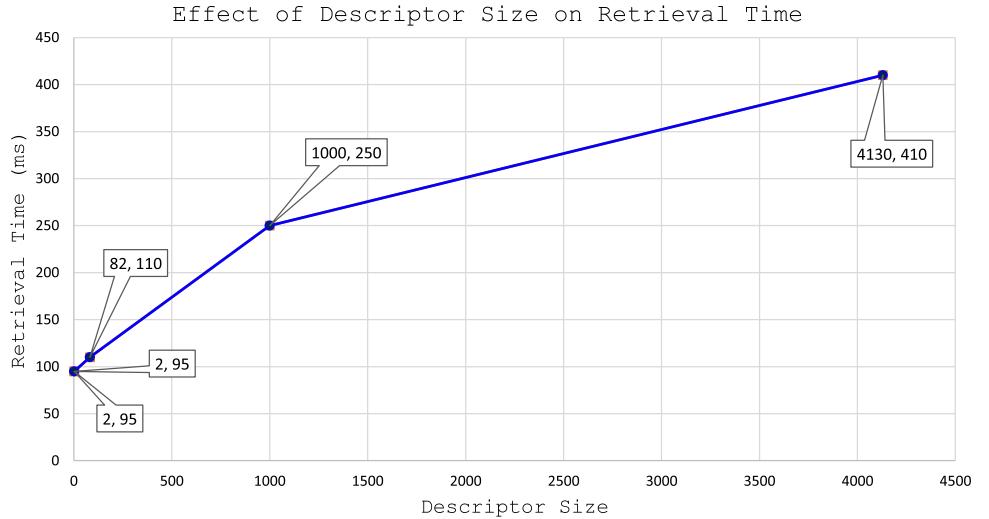
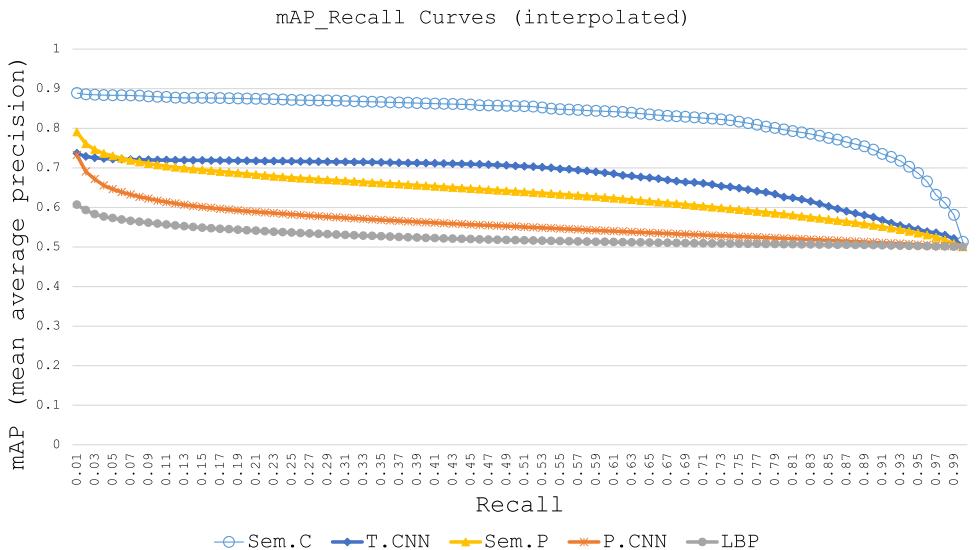


Fig. 8 Interpolated mAP–recall curves corresponding to the four experiments conducted in this paper. Method 1—LBP: local binary pattern descriptors, Method 2—P.CNN: pre-trained CNN descriptors, Method 3—Sem.P: semantic patient scores, Method 4—T.CNN: trained CNN descriptors same as in [16] and Method 5—Sem.C: semantic class scores



According to interpolated mAP–recall curves in Fig. 8, it can be observed that the representative curves corresponding to all methods (with the exception of method two) do not cross each other which indicates the absolute incremental enhancement on the retrieval results at all possible retrieval ranks (ranging from 0 to 1). In other words, the four methods are either always acting superior or inferior—at all possible ranks—as compared to each other indicating the consistency of the used descriptors under all possible magnification factors as well as the cancer subtypes. Please notice the query set is intentionally selected in a way to encompass all possible variations in the dataset as illustrated in Fig. 5. In the same figure, it is observed that at 100% recall, the mAP of all five methods meets at the same point on the graph (recall=0.5) which can be explained by equal number of malignant and benign images in the query set as demonstrated in Fig. 5.

6 Conclusion

In this paper, we developed five different BC (breast cancer) image retrieval systems based on different descriptors such as LBP (hand-engineered), deep transferred descriptors (from a pre-trained CNN model), patient-driven semantic scores (from multi-patient ensemble learner) and ultimately class-driven semantic scores (from a binary breast cancer learner). Aiming to address the semantic gap (as visualized in Fig. 1) in the breast cancer image retrieval, we performed, for the first time, multiple experiments on large-scale BreakHis dataset of histopathological images and demonstrated the superior performance of the semantic-driven approaches as against to non-semantic approaches.

In future work, we intend to design a re-ranking scheme to further enhance the retrieval performance by refining the top retrieved BC images. We will also be experimenting with deeper pre-trained CNN models to extract even more cancer-

specific transferred descriptors assuming that more layers provide more abstract representation of BC images.

Acknowledgements The authors would like to thank Fabio Alexandre Spanhol from Federal University of Paraná, Brazil, for providing BreakHis pathological dataset of breast cancer images.

References

1. Ahmad J, Muhammad K, Lee MY, Baik SW (2017) Endoscopic image classification and retrieval using clustered convolutional features. *J Med Syst* 41(12):196
2. Ahmad J, Muhammad K, Baik SW (2018) Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features. *J Med Syst* 42(2):24
3. Ali M, Dong L, Akhtar R (2017) Multi-panel medical image segmentation framework for image retrieval system. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-017-5453-8>
4. Boomilingam T, Subramaniam M (2017) An efficient retrieval using edge glcm and association rule mining guided ipso based artificial neural network. *Multimed Tools Appl* 76(20):21729–21747
5. Conjeti S, Paschali M, Katouzian A, Navab N (2017) Learning robust hash codes for multiple instance image retrieval. arXiv preprint [arXiv:1703.05724](https://arxiv.org/abs/1703.05724)
6. Das P, Neelima A (2017) An overview of approaches for content-based medical image retrieval. *Int J Multimed Inf Retr* 6(4):271–280
7. Escalera S, Pujol O, Radeva P (2009) Separability of ternary codes for sparse designs of error-correcting output codes. *Pattern Recognit Lett* 30(3):285–297
8. Jenitta A, Ravindran RS (2017) Image retrieval based on local mesh vector co-occurrence pattern for medical diagnosis from mri brain images. *J Med Syst* 41(10):157
9. Jiang M, Zhang S, Fang R, Metaxas DN (2015) Leveraging coupled multi-index for scalable retrieval of mammographic masses. In: 12th International Symposium on Biomedical Imaging (ISBI), IEEE , pp 276–280
10. Kieffer B, Babaie M, Kalra S, Tizhoosh H (2017) Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. arXiv preprint [arXiv:1710.05726](https://arxiv.org/abs/1710.05726)
11. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436
12. Li W, Pan H, Li P, Xie X, Zhang Z (2017) A medical image retrieval method based on texture block coding tree. *Signal Process: Image Commun* 59:131–139
13. Liu J, Zhang S, Liu W, Deng C, Zheng Y, Metaxas DN (2017) Scalable mammogram retrieval using composite anchor graph hashing with iterative quantization. *IEEE Trans Circuits Syst Video Technol* 27(11):2450–2460
14. Lowe DG (1999) Object recognition from local scale-invariant features. In: The proceedings of the 7th IEEE international conference on Computer vision, 1999. IEEE 2:1150–1157
15. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Machine Intell* 24(7):971–987
16. Qayyum A, Anwar SM, Awais M, Majid M (2017) Medical image retrieval using deep convolutional neural network. *Neurocomputing* 266:8–20
17. Shah A, Conjeti S, Navab N, Katouzian A (2016) Deeply learnt hashing forests for content based image retrieval in prostate mr images. In: Medical imaging 2016: image processing, international society for optics and photonics, vol 9784, p 978414
18. Shi X, Xing F, Xu K, Xie Y, Su H, Yang L (2017) Supervised graph hashing for histopathology image retrieval and classification. *Med Image Anal* 42:117–128
19. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. *CA: Cancer J Clin* 63(1):11–30
20. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
21. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) A dataset for breast cancer histopathological image classification. *IEEE Trans Biomed Eng* 63(7):1455–1462
22. Srikar E, Kumar BK (2016) CBIR and sift based computer-aided diagnosis of mammographic masses
23. Tang Q, Yang J, Xia X (2018) Medical image retrieval using multi-texton assignment. *J Digital Imag* 31(1):107–116
24. Varga D, Szirányi T (2016) Fast content-based image retrieval using convolutional neural network and hash function. In: IEEE international conference on systems, man, and cybernetics (SMC), 2016, IEEE, pp 002636–002640
25. Wu G, Coupé P, Zhan Y, Munsell B, Rueckert D (2015) Patch-based techniques in medical imaging. Springer, Berlin
26. Yadav P (2018) Cluster based-image descriptors and fractional hybrid optimization for medical image retrieval. *Cluster Comput*. <https://doi.org/10.1007/s10586-017-1625-6>
27. Zare MR, Müller H (2016) A medical x-ray image classification and retrieval system. *PACIS* 13
28. Zhang F, Song Y, Cai W, Hauptmann AG, Liu S, Pujol S, Kikinis R, Fulham MJ, Feng DD, Chen M (2016) Dictionary pruning with visual word significance for medical image retrieval. *Neurocomputing* 177:75–88
29. Zhang X, Dou H, Ju T, Zhang S (2015a) Fusing heterogeneous features for the image-guided diagnosis of intraductal breast lesions. In: IEEE 12th international symposium on biomedical imaging (ISBI), 2015, IEEE, pp 1288–1291
30. Zhang X, Liu W, Dundar M, Badve S, Zhang S (2015b) Towards large-scale histopathological image analysis: hashing-based image retrieval. *IEEE Trans Med Imaging* 34(2):496–506
31. Zhao JJ, Pan L, Zhao PF, Tang XX (2017) Medical sign recognition of lung nodules based on image retrieval with semantic features and supervised hashing. *J Comput Sci Technol* 32(3):457–469