A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal–Cell Carcinoma Cancer Detection

4 authors:

Angel Cruz-Roa
Universidad de los Llanos
48 PUBLICATIONS   **979** CITATIONS

SEE PROFILE

John Arevalo
National University of Colombia
22 PUBLICATIONS   **389** CITATIONS

SEE PROFILE

Anant Madabhushi
Case Western Reserve University
479 PUBLICATIONS   **8,256** CITATIONS

SEE PROFILE

Fabio A. González
National University of Colombia
227 PUBLICATIONS   **3,927** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Radiogenomics View project

Project Categorización de estructuras atmosféricas View project

A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection

Angel Cruz-Roa[a], John Arevalo[a], Anant Madabhushi[b], Fabio González[a]

[a]MindLab Research Group, Universidad Nacional de Colombia, Bogotá, Colombia
[b]Dept. of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

Abstract. This paper presents and evaluates a deep learning architecture for automated basal cell carcinoma cancer detection that integrates (1) image representation learning, (2) image classification and (3) result interpretability. A novel characteristic of this approach is that it extends the deep learning architecture to also include an interpretable layer that highlights the visual patterns that contribute to discriminate between cancerous and normal tissues patterns, working akin to a digital staining which spotlights image regions important for diagnostic decisions. Experimental evaluation was performed on set of 1,417 images from 308 regions of interest of skin histopathology slides, where the presence of absence of basal cell carcinoma needs to be determined. Different image representation strategies, including bag of features (BOF), canonical (discrete cosine transform (DCT) and Haar-based wavelet transform (Haar)) and proposed learned-from-data representations, were evaluated for comparison. Experimental results show that the representation learned from a large histology image data set has the best overall performance (89.4% in F-measure and 91.4% in balanced accuracy), which represents an improvement of around 7% over canonical representations and 3% over the best equivalent BOF representation.

1 Introduction

This paper presents a unified method for histopathology image representation learning, visual analysis interpretation, and automatic classification of skin histopathology images as either having basal cell carcinoma or not. The novel approach is inspired by ideas from image feature representation learning and deep learning [10] and yields a deep learning architecture that combines an autoencoder learning layer, a convolutional layer, and a softmax classifier for cancer detection and visual analysis interpretation.

Deep learning (DL) architectures are formed by the composition of multiple linear and non-linear transformations of the data, with the goal of yielding more abstract – and ultimately more useful – representations [10]. These methods have recently become popular since they have shown outstanding performance in different computer vision and pattern recognition tasks [2,8,10]. DL architectures are an evolution of multilayer neural networks (NN), involving different design and training strategies to make them competitive. These strategies include spatial invariance, hierarchical feature learning and scalability [10]. An interesting characteristic of this approach is that feature extraction is also considered as a part of the learning process, i.e., the first layers of DL

architectures are able to find an appropriate representation of input images in terms of low-level visual building blocks that can be learnt.

This work addresses the challenging problem of histopathology image analysis and in particular the detection of basal-cell carcinoma (BCC), the most common malignant skin cancer and may cause significant tissue damage, destruction and, in some cases, disfigurement. Unlike natural scene images, where typical automated analysis tasks are related to object detection and segmentation of connected regions that share a common visual appearance (e.g. color, shape or texture), histopathology images reveal a complex mixture of visual patterns. These patterns are related to high variability of biological structures associated with different morphology tissue architecture that typically tend to significantly differ in normal and diseased tissue. Another source of visual variability is the acquisition process itself, going from a 3D organ biopsy to a 2D sample (histopathological slide). This process involves different stages: sampling, cutting, fixing, embedding, staining and digitalization, each one contributing inherent artifacts [3]. Figure 1 shows histopathology image samples stained with hematoxylin-eosin (H&E) from cancerous and non-cancerous tissue samples. These images illustrate the high intra-class visual variability in BCC diagnosis, caused by the presence (or absence) of different morphological and architectural structures, healthy (eccrine glands, hair follicles, epithelium, collagen, sebaceous glands) or pathological (morpheiform, nodular and cystic change).
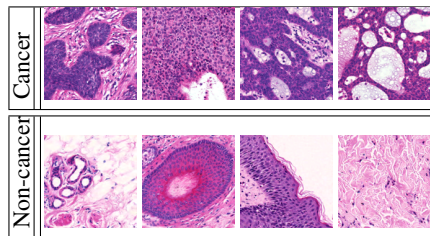

Fig. 1. Example of BCC histopathology images (both cancer and non-cancer) stained with H&E at 10X.

There is an extensive literature in automatic histopathology image analysis where different strategies for image representation have been tried: discrete cosine transform (DCT), wavelet coefficients, Gabor descriptors, and graph representations among others [7]. In all cases, the goal is to capture the visual features that better characterize the important biological structures related to the particular problem been addressed. This means that some representations may work better than others depending on the particular problem. On account of recent advances in computer vision [2,8], there is an encouraging evidence (mostly for natural scene images) that learned representations (induced from data) may have a better performance than canonical, predefined image feature representations. To the best of our knowledge, this is the first attempt to evaluate learned-from-data image features in BCC histopathology image analysis using a DL architecture. A related approach is a bag of features (BOF) representation, which attempts to learn a set of visual code words from training data and uses them to represent images. However, this representation strategy still needs the definition of a local feature descriptor in advance (e.g. raw-block, SIFT histogram, DCT coefficients). Previous research has suggested that the particular local feature descriptor choice has an important effect on BOF performance [4,5].

This paper presents a convolutional auto-encoder DL architecture for histopathology image classification as a tool to support BCC diagnosis. The DL architecture is enhanced by an interpretation layer that highlights the image regions that most contribute in the discrimination of healthy tissue from cancer. The main novel contributions of this work are:

- A BCC histopathological image analysis method, which integrates, in a unified DL model, the following functionalities: image feature representation learning, image classification, and prediction interpretability.
- An evaluation of learned-from-data image representations in BCC histopathology image analysis, which shows that this approach could produce improved classification performance while enhancing model interpretability.
- A novel strategy to exploit the information in the intermediate representation layers of the DL architecture to produce visual interpretable predictions. In that sense this method is analogous to a digital stain which attempts to identify image regions that are most relevant for making diagnostic decisions.

While there has been some previous related work in the use of DL architectures for automatic segmentation and classification in histopathology images for breast cancer cell detection, unlike our approach, the methods in [6,11] do not focus on image representation learning. Hence these methods do not analyze the learned features and do not explore their potential for visual prediction interpretation. Prediction interpretability is not an important issue when analyzing natural images, so it has not been typically studied in classical computer vision literature. However, in the context of systems for performing predictions and decision support there is a need to explain and identify those visual patterns which are relevant for prediction. While some approaches have been for visual prediction interpretation [1,3,4,5], these approaches have not used a DL architecture and in all of them the interpretation ability is provided by an additional stage, subsequent to image classification. By contrast, in our approach, visual interpretability is tightly integrated with the classification process.

2 Overview of the DL Representation and Classification Method

The new method for histopathology image representation learning, BCC cancer classification and interpretability of the results of the predictor, is based on a multilayer neural network (NN) architecture depicted in Figure 2. The different stages or modules of the framework, corresponding to different layers of the NN are described as follows:

Step 1. Unsupervised feature learning via autoencoders: Training images are divided into small patches (8×8 pixels), which are used to train an autoencoder NN [10] with k hidden neurons. This produces a set of weight vectors $W = \{w_1, \ldots, w_k\}$, which can be interpreted as image features learned by the autoencoder NN. The autoencoder looks for an output as similar as possible to the input [10]. This autoencoder learns features by minimizing an overall cost function with a sparsity constraint to learn compressed representations of the images defined by $J_{sparse}(W) = J(W) + \beta \sum_{j=1}^{k} KL(\rho || \hat{\rho}_j)$, where $J(W)$ is the typical cost function used to train a neural network, β controls the
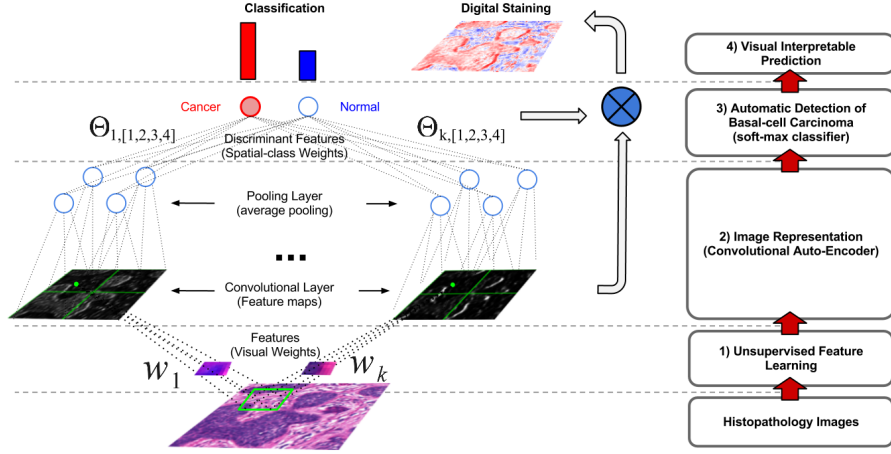
Fig. 2. Convolutional auto-encoder neural network architecture for histopathology image representation learning, automatic cancer detection and visually interpretable prediction results analogous to a *digital stain* identifying image regions that are most relevant for diagnostic decisions.

weight of sparsity penalty term. $KL(\rho||\hat{\rho}_j)$ corresponds to Kullback–Leibler divergence between ρ, desired sparsity parameter, and $\hat{\rho}_j$, average activation of hidden unit j (averaged over the training set).

Step 2. Image representation via convolution and pooling: Any feature w_k can act as a filter, by applying a convolution of the filter with each image to build a feature map. The set of feature maps form the convolutional layer. Thus, a particular input image is represented by a set of k features maps, each showing how well a given pattern w_i spatially matches the image. This process effectively increases the size of the internal representation ($\approx k \times$ the size of the original representation) of the image. The next layer acts in the opposite direction by summarizing complete regions of each feature map. This is accomplished by neurons that calculate the average (pool function) of a set of contiguous pixels (pool dimension). The combination of convolution and pooling provide both translation invariance feature detection and a compact image representation for the classifier layer.

Step 3. Automatic detection of BCC via softmax classifier: A softmax classifier, which is a generalization of a logistic regression classifier [2], takes as input the condensed feature maps of the pooling layer. The classifier is trained by minimizing the following cost function: $J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)}\log h_\Theta(x^{(i)}) + (1+y^{(i)})\log(1-h_\Theta(x^{(i)}))\right]$, where $\left\{(x^{(1)},y^{(1)}),\ldots,(x^{(m)},y^{(m)})\right\}$ is the corresponding training set of m images, where the i-th training image is composed of $y^{(i)}$ class membership and $x^{(i)}$ image representation obtained from the output of the pooling layer, and Θ is a weight vector dimension $k \times n$ (where n is the pool dimension). The output neuron has a sigmoid activation function, which produces a value between 0 and 1 that can be interpreted as the probability of the input image being cancerous.

Step 4. Visual interpretable prediction via weighted feature maps: The softmax classifier weights (Θ) indicate how important a particular feature is in discriminating between cancer and non-cancer images. A weight Θ_j (associated[1] to a feature w_k) with a high positive value indicates that the corresponding feature is associated with cancer images, in the same way a large negative value indicates that the feature is associated with normal images. This fact is exploited by the new method to build a *digitally stained* version of the image, one where cancer (or non-cancer) related features are present. The process is as follows: each feature map in the convolutional layer is multiplied by the corresponding weight (Θ) in the softmax layer, all the weighted feature maps are combined into an integrated feature map. A sigmoid function is applied to each position of the resulting map and finally the map is visualized by applying a colormap that assigns a blue color to values close to 0 (non-cancer) and a red color to values close to 1 (cancer).

3 Experimental Setup

Histopathology basal cell carcinoma dataset description (BCC dataset): The BCC dataset comprises 1417 image patches of 300×300 pixels extracted from 308 images of 1024×768 pixels, each image is related to an independent ROI on a slide biopsy. Each image is in RGB and corresponds to field of views with a 10X magnification and stained with H&E [1,3]. These images were manually annotated by a pathologist, indicating the presence (or absence) of BCC and other architectural and morphological features (collagen, epidermis, sebaceous glands, eccrine glands, hair follicles and inflammatory infiltration). The Figure 1 shows different examples of these images.

Learned image feature representations: The main focus of the experimental evaluation was to compare the image representations learned from the histopathology data, generated by the DL-based proposed method, against two standard canonical features (discrete cosine transform (DCT) and Haar-based wavelet transform (Haar)). Since the focus of this experimentation was the evaluation of different image representations, the same classification method was used in all the cases (steps 2 to 4 of the new architecture, Figure 2). Also a comparison against BOF image representation was included with the same number of features and patch size employing the same local feature descriptors (DCT and Haar) based on previous work [3,5]. In the case of canonical image representations, the feature weights $\{w_1, \ldots, w_k\}$ were replaced by the basis vectors that define either DCT or Haar. In addition to the image features learned from the BCC dataset, two other sets of image features were learned from different data sets: a histology data set composed of healthy histological tissue images (HistologyDS[2] [4]) and a natural scene image data set commonly used in computer vision research (STL-10 dataset[3]). In order to choose an appropriate parameter configuration for our method, an exhaustive parameter exploration was performed. The parameters explored were: image scales (50%,20%), number of features (400, 800), pool dimension (7, 13, 26, 35, 47, 71, 143)

---

[1] In general, a feature has as many weights associated with it as the pool dimension. When the pool dimension is larger than one, the average weight is used.

[2] Available in: `http://www.informed.unal.edu.co/histologyDS/`

[3] Available in: `http://www.stanford.edu/~acoates//stl10/`

and pooling function (average or sum). The best performing set of parameters were a image scale of 50%, 400 features, a pool dimension of 71 with average pooling function, and all the reported results correspond to this configuration. A patch size of 8×8 pixels was used since this was ascertained to be the minimum sized region for covering a nucleus or cell based off the recommendations from previous related work using the same datasets [3,5].

Cancer detection performance: Once the best parameter configuration for the DL architecture is selected, each representation was qualitatively evaluated for comparison using a stratified 5-fold cross-validation strategy on a classification task (discriminating cancer from non-cancerous images) in the BCC dataset. The performance measures employedwere accuracy, precision, recall/sensitivity, specificity, f-measure and balanced accuracy (BAC).

4   Results and discussion

Learned representations vs canonical representations: Figure 3 shows a set of 200 image features learned from three different data sets, (a) BCC, (b) HistologyDS and (c) STL-10, along with the set of features corresponding to and obtained from the DCT representation, (d). In all the cases, features were sorted according to their frequency when representing the BCC data set. As expected features learned from BCC and HistologyDS images better capture visual patterns related to dyes, edges of large nuclei in different orientations and perhaps most interestingly small dots related to common/healthy nuclear patterns that do not appear in the other feature sets. Features learned from the natural image dataset (STL-10), also capture visual patterns such as edges, colors, and texture but are less specific. The features associated with DCT representation are even less specific and only capture general color and texture patterns.
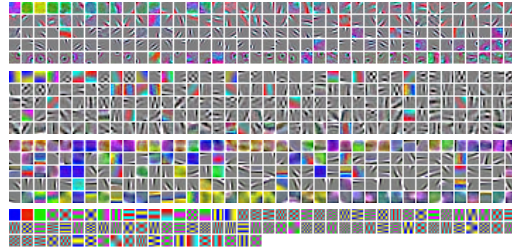

Fig. 3. Comparison of learned features (a.k.a. dictionaries or basis) by autoencoders from: a) BCC (histopathology), b) HistologyDS (healthy tissues) and c) STL-10 (natural scenes) datasets, and d) DCT basis.

Automatic cancer detection performance: Table 1 presents the classification performance results in terms of accuracy, precision, recall, specificity, f-measure and balanced accuracy (BAC). The results show a clear advantage of learned features over canonical and BOF representations. A *t*-test showed that differences among DL models with learned features was not significant ($p > 0.05$) and that DL models were significantly better ($p < 0.05$) than canonical features and BOF representations. The fact that features learned from histology images produced the best results for histopathology image classification is an interesting result, suggesting that the proposed approach is learning important features to describe general

visual pattern present in different histopathological images. This is consistent with the findings in [9], which had shown that the strategy of learning features from other large datasets (known as self-taught learning) may produce successful results .

Table 1. Classification performance for each learned features from different image datasets using the new DL based method, canonical features, along with different configurations of BOF. The best results are in bold typeface.

	Accuracy	Precision	Recall/Sensitivity	Specificity	F-measure	BAC
(DL) BCC	0.906 +/- 0.032	0.876 +/- 0.030	0.869 +/- 0.049	0.927 +/- 0.028	0.872 +/- 0.037	0.898 +/- 0.034
(DL) HistologyDS	**0.921 +/- 0.031**	**0.901 +/- 0.041**	**0.887 +/- 0.033**	**0.941 +/- 0.034**	**0.894 +/- 0.032**	**0.914 +/- 0.029**
(DL) STL-10	0.902 +/- 0.027	0.871 +/- 0.046	0.867 +/- 0.024	0.922 +/- 0.033	0.868 +/- 0.024	0.895 +/- 0.022
(DL) DCT	0.861 +/- 0.036	0.824 +/- 0.042	0.794 +/- 0.059	0.900 +/- 0.037	0.808 +/- 0.037	0.847 +/- 0.037
(DL) Haar	0.841+/- 0.032	0.787+/- 0.039	0.785 +/- 0.061	0.873 +/- 0.048	0.784 +/- 0.027	0.829 +/- 0.030
(BOF) Haar-400	0.796 +/- 0.026	0.796 +/- 0.059	0.680 +/- 0.067	0.864 +/- 0.048	0.708 +/- 0.031	0.772 +/- 0.026
(BOF) GrayDCT-400	0.880 +/- 0.026	0.880 +/- 0.033	0.834 +/- 0.042	0.908 +/- 0.019	0.836 +/- 0.028	0.871 +/- 0.027
(BOF) ColorDCT-400	0.891 +/- 0.023	0.891 +/- 0.026	0.851 +/- 0.033	0.914 +/- 0.017	0.851 +/- 0.027	0.883 +/- 0.024

Digital staining results: Table 2 illustrates some examples of the last and most important stage of the new DL method– digital staining. The table rows show from top to bottom: the real image class, the input images, the class predicted by the model, the probability associated with the prediction, and the digital stained image. The digitally stained version of the input image highlights regions associated to both cancer (red stained) and non-cancer (blue stained) regions. These results were analyzed by a pathologist, who suggested that our method appeared to be identifying cell proliferation of large-dark nuclei. A caveat however is that this feature also appears to manifest in healthy structures where the epidermis or glands are present. Nonetheless, this enhanced image represents an important addition to support diagnosis since it allows the pathologist to understand why the automated classifier is suggesting a particular class.

Table 2. Outputs produced by the system for different cancer and non-cancer input images. The table rows show from top to bottom: the real image class, the input image, the class predicted by the model and the probability associated to the prediction, and the digital stained image (red stain indicates cancer regions, blue stain indicates normal regions).

True class	Cancer	Cancer	Cancer	Non-cancer	Non-cancer	Non-cancer
Input image						
Pred/Prob	Cancer (0.82)	Cancer (0.96)	Cancer (0.79)	Non-cancer (0.27)	Non-cancer (0.08)	Non-cancer (0.03)
Digital staining						

5 Concluding Remarks

We presented a novel unified approach for learning image representations, visual interpretation and automatic BCC cancer detection from routine H&E histopathology images. Our approach demonstrates that a learned representation is better than a canonical predefined representation. This representation could be learned from images associated with the particular diagnostic problem or even from other image datasets. The paper also presented a natural extension of a DL architecture to do digital staining of the input images. The inclusion of an interpretability layer for a better understanding of the prediction produced by the automated image classifier.

Acknowledgements

References

1. Gloria Díaz and Eduardo Romero. Micro-structural tissue analysis for automatic histopathological image annotation. *Microsc Res Tech*, 75(3):343–358, 2012.
2. Alex Krizhevsky et al. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1106–1114, 2012.
3. Angel Cruz-Roa et al. Automatic Annotation of Histopathological Images Using a Latent Topic Model Based On Non-negative Matrix Factorization. *J Pathol Inform*, 2(1):4, 2011.
4. Angel Cruz-Roa et al. Visual pattern mining in histology image collections using bag of features. *Artif Intell Med*, 52(2):91–106, June 2011.
5. Angel Cruz-Roa et al. A visual latent semantic approach for automatic analysis and interpretation of anaplastic medulloblastoma virtual slides. In *MICCAI*, pages 157–164, 2012.
6. Baochuan Pang et al. Cell nucleus segmentation in color histopathological imagery using convolutional networks. In *CCPR*, pages 1–5. IEEE, 2010.
7. Lei He et al. Histology image analysis for carcinoma detection and grading. *Comput Meth Prog Bio*, 2012.
8. Quoc V Le et al. Building high-level features using large scale unsupervised learning. In *ICML*, 2011.
9. Rajat Raina et al. Self-taught learning: transfer learning from unlabeled data. In *ICML '07*, pages 759–766.
10. Yoshua Bengio et al. Representation learning: A review and new perspectives. *Arxiv*, 2012.
11. Grégoire Montavon. A machine learning approach to classification of low resolution histological samples. Master's thesis, 2009.