Training a cell-level classifier for detecting basal-cell carcinoma by combining human visual attention maps with low-level handcrafted features

Germán Corredor
Jon Whitney
Viviana Arias
Anant Madabhushi
Eduardo Romero

Training a cell-level classifier for detecting basal-cell carcinoma by combining human visual attention maps with low-level handcrafted features

Germán Corredor,^{a,b} Jon Whitney,^b Viviana Arias,^c Anant Madabhushi,^b and Eduardo Romero^{a,*}

^aUniversidad Nacional de Colombia, Computer Imaging and Medical Applications Lab, Department of Medical Imaging, Bogota, Colombia

^bCase Western Reserve University, Center of Computational Imaging and Personalized Diagnostics, Department of Biomedical Engineering, Cleveland, Ohio, United States

^cUniversidad Nacional de Colombia, Patología Molecular Research Group, Department of Pathology, Bogota, Colombia

Abstract. Computational histomorphometric approaches typically use low-level image features for building machine learning classifiers. However, these approaches usually ignore high-level expert knowledge. A computational model (M_{im}) combines low-, mid-, and high-level image information to predict the likelihood of cancer in whole slide images. Handcrafted low- and mid-level features are computed from area, color, and spatial nuclei distributions. High-level information is implicitly captured from the recorded navigations of pathologists while exploring whole slide images during diagnostic tasks. This model was validated by predicting the presence of cancer in a set of unseen fields of view. The available database was composed of 24 cases of basal-cell carcinoma, from which 17 served to estimate the model parameters and the remaining 7 comprised the evaluation set. A total of 274 fields of view of size 1024 × 1024 pixels were extracted from the evaluation set. Then 176 patches from this set were used to train a support vector machine classifier to predict the presence of cancer on a patch-by-patch basis while the remaining 98 image patches were used for independent testing, ensuring that the training and test sets do not comprise patches from the same patient. A baseline model (M_{ex}) estimated the cancer likelihood for each of the image patches. M_{ex} uses the same visual features as M_{im}, but its weights are estimated from nuclei manually labeled as cancerous or noncancerous by a pathologist. M_{im} achieved an accuracy of 74.49% and an *F*-measure of 80.31%, while M_{ex} yielded corresponding accuracy and *F*-measures of 73.47% and 77.97%, respectively. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.4.2.021105]

Keywords: digital pathology; cancer detection; basal-cell carcinoma; classification; visual attention map; implicit relevance feedback; graphs.

Paper 16181SSRR received Aug. 23, 2016; accepted for publication Feb. 22, 2017; published online Mar. 11, 2017.

1 Introduction

The development of computer-assisted histomorphometric tools for interrogating digital pathology slide images allows for the discovery of patterns and objective measurements associated with disease aggressiveness, grade, and patient outcome.^{1–4} Nevertheless, the size of a whole slide image (WSI) that typically runs in the order of several gigabytes poses computational and diagnostic challenges.⁵ Despite the large size of a sample, in general, pathologists require very little time to examine samples and reach a diagnosis.⁶ They rapidly identify regions of potential diagnostic interest, i.e., areas that usually contain abnormal patterns that characterize a particular set of pathologies.^{7,8} Although every case is different, it is widely acknowledged that diseases present certain characteristic patterns and pathologists are educated to recognize them, even under very noisy conditions.^{9,10}

There is a growing demand in pathology to use objective measurements that support treatment decisions, but few systems in use can meet this need.¹¹ The ability to identify the regions of interest (ROIs) that contain pertinent diagnostic information could (1) reduce the time pathologists need to dedicate to

each sample by focusing their attention on relevant areas¹¹ and (2) improve computational image analysis tools, thus enabling improved objective measurements of histological data.^{11,12} Different approaches^{13–16} have attempted to identify relevant information on medical imaging databases using manual annotations, a very time consuming, laborious, and impractical approach when processing vast image collections. Other approaches have used automated ROI identification by handcrafted features (e.g., color, shape, and texture). For instance, Doyle et al.¹⁷ applied a boosted Bayesian multiresolution classifier along with texture features to identify areas of adenocarcinoma in prostate biopsy slide images. Low-resolution ROIs were identified by a texture classifier and then mapped onto the next higher resolution to generate high-resolution ROIs. Another method to automatically detect ROIs in WSIs involves the use of handcrafted graph features. Graphs are mathematical constructions composed of finite sets of connected objects (nodes) that capture global and local relationships via pairwise connections between their members (edges). These structures can be used to quantitatively represent spatially distributed information, such as neighborhood relationships and the spatial arrangement of structural tissue primitives (e.g., nuclei, lymphocytes, and glands).^{18–21} Different graph-based features such as

*Address all correspondence to: Eduardo Romero, E-mail: edromero@unal.edu.co

Delaunay triangles, Voronoi diagrams, and minimum spanning trees have been extracted and combined with other visual features to effectively characterize histopathologic images.^{1,20,22–24}

While these approaches turn out to be simple, accurate, and efficient, in practice they suffer from different drawbacks. One of these drawbacks is the semantic gap, a concept related to the impossibility of fully describing the global meaning of the image by using low-level image features.^{25–27} Figure 1 shows a situation in which two visually very similar images (shape, texture, and color) have very different meanings: Fig. 1(a) is a hair follicle, a normal structure, and Fig. 1(b) is a basal-cell carcinoma (BCC) nodule, a cancerous structure. While an expert can easily identify and differentiate such structures, this is not an easy task for a computer program that uses only information coming from visual primitives.

The semantic gap might be minimized by introducing prior knowledge from the pathologist as part of the learning workflow. This prior knowledge could be used to inform low-level feature algorithms where in the images to initially “look.” These algorithms could combine pathologists’ knowledge with features such as (1) low-level image cues such as color and texture of the epithelium and the stroma and (2) midlevel structural motifs such as nuclear architectural patterns, collagen patterns, and gland morphology.

There are several methods for either predicting or identifying the interesting ROIs in a WSI based on the pathologist knowledge. For example, Peter et al.²⁸ described a method using a random forest algorithm to learn and identify potential ROIs within an image. These regions were then shown to an expert who interactively flagged the regions that were actually relevant. This in turn allowed a continuous updating of the scoring function for different ROIs. This approach requires substantial manual interaction that is time consuming and therefore difficult to implement in clinical practice. Alternatively, this high-level expert domain knowledge might be captured by implicitly extracting information from the pathologist actions with the virtual microscope during a diagnosis task.^{6,29,30} One manner consists of passively recording the pathologists navigations during a diagnostic task, a strategy perfectly suited for real clinical scenarios. In such an approach, the idea is to identify and characterize the most frequently visited areas, since very likely they contain relevant diagnosis information. In a previous work,³¹ a Bayesian framework predicted regions an expert might visit based on the visual data and previous visits of other experts, attempting to improve the cache performance for different navigation tasks. This approach begins by learning a set of candidate relevant regions using salient information coming from visual features such as color, shape, and orientation. These

candidate regions are then pruned based on feedback information from the navigation trends of the pathologist. This approach is limited because each image needs at least one navigation profile to generate the initial prediction, and thus cannot function without initial user input.

This paper introduces a model that aims to integrate low-, mid-, and high-level information in order to determine whether a nucleus is cancerous. The model exploits the implicit knowledge extracted from actions performed by pathologists while navigating WSIs. Briefly, the method starts by segmenting the WSI nuclei and computing a set of nuclear visual features typically altered in regions with cancer, namely color, size, and spatial distribution.¹⁰ Then, each nucleus is assigned a likelihood of being cancerous based on the number of times said nucleus is visited by a group of pathologists while performing a diagnostic task. Finally, the relative importance of each of the visual features is calculated by performing a multilinear regression between the visual features associated with each nucleus and the corresponding likelihood of cancer presence. The primary contribution of this work is a simple and adaptable integration of the high-level expert knowledge (implicitly extracted) with handcrafted visual features to create a classifier for cancer detection on a patch-by-patch basis within WSIs. The underlying rationale here is that prior information, derived from pathologists’ interactions with the WSI, improves the predictive performance of handcrafted features. The model is trained offline with a set of WSIs and then applied to classification/prediction tasks in new images (previously unseen). A second contribution of this method is that it may provide an alternative approach to defining expert ground truth since this method enables identification of areas that represent a high level of interest. Figure 2 shows the workflow of the new approach to identify suspicious ROIs from WSIs of BCC.

The presented method was evaluated in terms of its ability to determine if a set of patches or fields of view (FOVs), extracted from WSIs, contained cancerous regions. The validation of our approach was performed on a use case involving the detection of BCC from WSIs. BCC is the most common malignant skin cancer.³² Although skin cancer is a threat to human life, fortunately it can be cured if it is detected and removed before it makes metastasis.³³ However, detecting BCC is challenging because it is difficult to distinguish the actual tumor from its surrounding noncancerous tissue.³⁴ If BCC remains undiagnosed, the tumor can grow, requiring more extensive resection that can result in visual or functional deformities. Currently, diagnosis relies on the experience and subjective judgment of each pathologist.³⁴ The development of objective measures for BCC detection could help to improve clinical decision

Fig. 1 Two very visually similar structures obtained from skin histological samples. (a) A hair follicle, a normal structure. (b) A basal-cell carcinoma, a cancerous structure.

Fig. 2 Illustration of the process of identifying cancer ROIs by fusing information from handcrafted features and interactions with the pathologist. In the learning phase, two types of information are extracted from a set of WSIs: low- and high-levels. Low-level features originate from the visual content of the image, specifically from nuclei arrangements and their sizes and colors. High-level information, such as magnification-weighted location focus, comes from the visual attention of the experts. These two information sources are then integrated by a multilinear model that learns the relative importance (weights) of each handcrafted feature for predicting ROIs, i.e., cancer areas. In the prediction phase, low-level features are extracted from new WSIs. Using the learned weights, a nuclei cancer likelihood is calculated to predict whether a specific patch is either cancerous or not.

support approaches for disease diagnosis and could also potentially facilitate better strategies for treatment planning. This paper shows how integration of navigations and low-level image features could help to identify cancerous regions in BCC WSIs.

The organization of the rest of this paper is as follows. In Sec. 2, we describe the process to extract the low- and high-level information from WSIs and how the likelihood score for each nucleus is determined. Section 3 details the dataset, as well as the experiments performed and their respective results. In Sec. 4, we present a discussion of our experimental results and findings and also briefly discuss limitations and directions for future work. Finally, Sec. 5 presents some closing thoughts and conclusions of the presented work.

2 Description of Methods

2.1 Low-Level Model

2.1.1 Nuclear segmentation

The low-level image features employed in this work are based on spatial nuclear architecture and arrangement. Consequently,

the first challenge in extracting these measurements is carefully identifying these nuclei on digitized pathology images.

While a number of different nuclear segmentation methods for digitized pathology images have been previously presented,^{35–38} we opted to go with the method of Wienert et al.³⁹ This strategy, whose major advantage is its simplicity, involves the use of an adaptive tracing technique that detects contours that are filtered based off the strength of the image gradient. The convex-hull of each detected region is employed to separate out nuclei clusters. Nuclei classification is, subsequently, performed on these clusters in order to (1) distinguish between overlapping and conjoined nuclei and (2) distinguish nuclei from similar appearing primitives such as lymphocytes. According to the authors,³⁹ with a validation set containing 7931 manually annotated cells from 36 images of different organs, this segmentation method presents a precision = 0.908 and recall = 0.859. After these results, one could expect that about 15% of the population might be missed; however, the fact that the segmentation process is applied to the complete WSI means that in practice the probability distribution function of these features would change very little by the misdetected nuclei.

2.1.2 Nuclear feature extraction

Experts frequently base their diagnosis of disease presence or absence on nuclear features.¹⁰ Consequently, nuclei represent critical visual information for analysis in digital histopathology. For this reason, we attempted to determine the degree to which a nucleus might represent cancer. Identifying individual nuclei that represent cancer might in turn enable identifying regions involved with cancer. For this reason, each nucleus was assigned a likelihood of being cancerous or not based on its visual properties; hence, features that are characteristic of morphologic disruption in cancer were extracted, namely attributes pertaining to spatial distribution, size, and color of the nuclei.

Since cancer is characterized by rapid cell proliferation and the formation of cell clusters, our model was designed to assign a likelihood to each nucleus according to its grade of grouping. A high value for an individual nucleus means that it is close to other nucleus and has a high probability of being cancerous. The model is defined as follows: once nuclei are segmented, the center of gravity of each individual cell is associated with a node in a graph that represents nuclei spatial relationships. The weights of the graph edges correspond to the inverse of the Euclidean distances between the connected nuclei. The likelihood of a node being cancerous is then computed by adding the corresponding weights of every edge connecting to that node. In this way, this feature captures the local spatial distribution of cells. Particular attention was paid to setting a threshold value that prevented large dense clusters from dominating the likelihood distribution. If the analysis is performed without such a threshold, the probability mass of the likelihood is basically concentrated into very few foci, typically one or two. The idea behind this threshold is to achieve a likelihood distributed as homogeneously as possible. For this reason, this threshold was set to 0.02; a value that ensured a number of foci reached a number of Gaussians with similar probability mass. This value was established after an exhaustive search for the optimal threshold values on a subset of images from the training set. This model is defined as

$$x_{1,i} = \sum_{j=1}^{n-1} \frac{1}{D_{i,j}} \times b_{\theta}, \quad (1)$$

where $x_{1,i}$ is the likelihood of the spatial distribution-based feature for the i 'th nucleus, n is the number of nuclei, $D_{i,j}$ is the Euclidean distance between nuclei i and j , θ is the threshold value, and b_{θ} is a binary value that is 0 when $D_{i,j} > \theta$ and 1 otherwise. This model generates very high values if nuclei are close together or overlapping, a useful condition for the model that detects clustered nuclei. It should be noted that these values are calculated between centroids, therefore, the minimum distance between two nuclei is at least the nucleus radius, thus this quantity will never diverge.

Cancer cells are also characterized by the proliferation of cells with large nuclei. For this reason, our method uses a model that clusters cells based on their area in order to highlight homogeneous neighborhoods. In this process, the area of each nucleus is computed and modeled as a graph in which nodes are the nuclei and weights are the inverse of the differences between areas. Here, the likelihood of each nucleus being cancerous corresponds to the sum of the weights of the nodes connecting to it. High likelihood values are typically ascribed to cells with large areas surrounded by other similar cells.

It is also known that cancer cells and their environment (surrounding area) present slight differences in color when compared to normal cells. Taking this into account, the averaged *YUV* color values of each nucleus and its surrounding area are also extracted. The area surrounding the nucleus is defined as a circumference with radius $\times 1.5$ the nucleus major axis length but excluding, of course, the nucleus itself and any other intersecting nuclei. This value aims to model a nucleus–cytoplasm relationship that usually is 1:4, but in the present investigation was set to 1:3 trying to take into account the shrinkage effect of the whole histological procedure ($\sim 25\%$ according to Ref. 40).

Finally, after applying these models and the subsequent feature extraction, a set of eight feature values was obtained for each nucleus. These included spatial distribution, size, mean nuclear *Y*, *U*, and *V* values, and mean *Y*, *U*, and *V* values of the region surrounding each nucleus.

2.2 High-Level Model

We had a set of pathologists perform a series of diagnostics tasks on WSIs and captured the coordinates of the window of interest they visited and the type of action the expert performed (panning, zooming in or out). All of this information was recorded for each reader and used to generate the respective visual attention maps.

2.2.1 Visualization strategy and tracking system

Given the size of the WSIs, the most common visualization strategy consists in constructing a multiresolution pyramid of the image, compressed in JPEG format; however, this approach presents important limitations regarding management, quality, flexibility, and performance.^{41–43} However, the JPEG2000 (J2K) standard, characterized by a natural multiresolution decomposition, lossless compression, and random spatial access,^{44,45} better suits the requirements for this type of interaction. In Ref. 43, the J2K model was adapted to improve the dynamic interaction between WSIs and a custom virtual microscope was enabled to run on small devices such as smartphones or tablets. The principle of this navigation is that the client requests ROIs instead of the entire image, thereby reducing the amount of information requested. When users navigate a particular WSI, they have as reference a window of interest that is moved along a navigation path, and whose coordinates and magnification are encoded as spatial coordinates. These coordinates define the J2K packets needed to construct the specific requested region. The packets are then extracted from the compressed file, combined with the region parameters and the image main header, decompressed using a J2K decoder, and used by the client to reconstruct the original ROI. Here, this strategy was adapted and used to enable pathologists to visualize different WSIs.

Furthermore, the visualization system was tuned to register the different navigation movements and actions performed by the pathologists during the diagnostic task. The tracking system records the time, location, and magnification activity along the entire diagnostic path. This information is subsequently stored for offline analysis.

2.2.2 Visual attention map

Visual attention maps are a summary of the most visited regions of an image by a set of experts during a diagnosis task. In this

work, the visual attention map is built using the frequency with which each image pixel is visited; this approach assumes that regions most visited by a set of pathologists are those likely containing cancer. Information related to the time pathologists spend at examining a particular region was ignored by this model since it could be highly noisy. Generation of this map starts by setting to zero a visit counter for every pixel. Each time a pathologist visits a certain region, the visit counter of each pixel belonging to such a region increases (see Fig. 3 for illustration). Consequently, pixels from regions more frequently visited will have higher visit counter values. The final visual attention map is generated by summing all the visits from every pathologist, followed by a normalization process that sets 1 to the highest value and 0 to the lowest. From this map, we infer the likelihood of a cancerous nucleus depending on its location, i.e., a nucleus in a frequently visited area will have a higher likelihood value than a nucleus in a nonvisited region.

2.3 Learning Feature Relevance

As previously stated, traditional approaches usually employ only low-level features for building machine learning classifiers for disease diagnosis. While these strategies have shown good results at discriminating among classes, they disregard the concepts behind the image and ignore the expert domain knowledge.

In this work, we employ a mathematical formulation that enables the integration of higher-level expert knowledge with lower-level image-derived features. The set of low-level features used to determine the degree to which a nucleus is cancerous can be thought of as the prior class conditional information, invoking the Bayesian formulation. Additionally, information about which nuclei define a diagnosis can be implicitly extracted from the visual attention map, which can be thought of as evidence or class prior. Thus, the likelihood function of an

individual nucleus was defined as a linear combination of its visual features as

$$y = \sum_{k=0}^f w_k x_k, \quad (2)$$

with x_k the k 'th feature vector, $x_0 = 1$ (the bias factor) and w_k the corresponding weight (the model parameters to be learned), f the number of different features, and y the resulting likelihood. The bias term is usually introduced in multilinear regression models to avoid the regression line that is forced to pass through the origin. In other words, if the bias is not used, when all of the feature regressors or predictors are zero, the predicted likelihood value should be zero, a meaningless statement in this context since nuclei relevance is always nonnull by other features than can have any influence on such a likelihood value. So, a bias term independent of the regressors was added to allow the hyperplane described by the learned weights to naturally capture the statistical relations.

In order to integrate this information with the expert knowledge, the information extracted from the visual attention map was used as the objective function in this linear model. The weight or relative importance of each feature is then learned by performing a multilinear regression using the least squares method.⁴⁶

2.4 Model Validation

Validation of the presented approach was carried out using 24 patient studies as described below (see Sec. 2.6). First, the feature weights were estimated using the model described in Sec. 2.3. This process was carried out using the 70% of the cases (17 randomly selected) (see Fig. 4). Second, each of the images within the remaining 30% (seven images) was

Fig. 3 Generation of the WSI visual attention map. The visual attention map is generated at a pixel level. Every pixel belonging to the navigation window is set to the relevance level, defined as the number of times said pixel was included within one of the observation windows. In this example, a pathologist exploring the WSI changed the observation window. Each time this happened, the number of visits was continuously updated for every pixel within the observation window. In step 1, all image pixels were initialized with a visit counter of 0. Steps 2 to 4 show visits to different image regions (red squares); in these cases, the visit counter of the pixels within the visualized regions increases. In step 5, a region is revisited and the visit counters are incremented for every pixel within the observation window. Finally, a visual attention map is obtained from the total number of pixel visits. The more frequently a region is visited, the more likely it is cancerous (brighter).

Fig. 4 Dataset partition.

split into smaller patches or FOVs with a size of 1024×1024 pixels. This FOV typically contains most of the relative primitives and surrounding structures needed to render a diagnosis by a pathologist. This FOV size was empirically determined by the pathologists. FOVs with no tissue were discarded. Next, each FOV was labeled by an expert pathologist as cancer or not-cancer. The number of FOVs is dependent on the WSI size, i.e., the larger the WSI the higher the number of FOVs. Once all the FOVs are collected, the number of samples within the cancer and noncancer classes was balanced by randomly removing a set of FOVs with the most represented class. Then 176 FOVs corresponding to five patients were used to train a classifier to predict the presence of cancer on a patch-by-patch basis while the remaining 98 FOVs, from two patients, were used for independent testing.

The first step to predict the cancer presence in an FOV was to calculate the likelihood of being cancerous for each of its nuclei. For this purpose, each nuclear feature was multiplied with its corresponding weight in the weighting vector (previously learned). Features that were positively correlated with cancer were assigned larger positive weights, features inversely correlated with cancer were given negative weights, and uncorrelated features were assigned weights of 0. Thus, the numerical result of multiplying the feature and weighting vectors was a direct indication of the likelihood that each nucleus was cancerous. Finally, the likelihood of an FOV being cancerous is calculated from the average likelihood of its individual constituent nuclei.

2.5 Baseline

The model described in this work (M_{im}) [defined by Eq. (1)] attempts to infer the relevance of visual features from implicit knowledge, extracted from interactions between a group of pathologists and WSIs. The comparative strategy or baseline model (M_{ex}) uses the very same visual features as M_{im} (see Sec. 2.1.2), but the estimation of weights is performed from explicit knowledge, i.e., nuclei manually labeled by an expert as cancerous or noncancerous (see Fig. 5). The baseline feature weights are used to predict the likelihood of cancer on a per-nucleus basis, across the different WSIs. Finally, from these WSIs, a set of FOVs is extracted and a classifier is trained to predict the presence or absence of cancer at the FOV level.

2.6 Dataset

The dataset consisted of hematoxylin–eosin (H&E) slides from patients diagnosed with BCC. These samples exhibit unique and mixed BCC types, namely superficial, nodular, micronodular, morpheaform, and trabecular. They were collected from the Pathology Department of Universidad Nacional de Colombia between 2009 and 2014 and were randomly selected from a set of 98 patients previously diagnosed with BCC and for whom slides were available. All the cases were anonymized. The study cohort was limited to 24 cases, each from a different patient, owing to the time constraints involving digitizing, manually annotating the samples, and recording the navigations of four different pathology readers. The slides were digitized at $40\times$ using a triocular CARL ZEISS Axiostar plus microscope coupled to a DXM1200 Nikon color digital camera, controlled by a custom-motorized scanner.

3 Experimental Results

3.1 Visual Attention Map

Four pathologists with at least 10 years of professional experience navigated 17 WSIs (70% of all the cases) using a customized virtual microscope (see Sec. 2.2.1). Each pathologist was asked to examine each WSI until a diagnosis of disease presence or absence had been made. The set of WSIs used in the present investigation were determined to be of intermediate diagnostic difficulty, as defined by our dermatopathologist (see details on the dataset in Sec. 2.6). The test demanded only a general diagnosis and all four pathologists reached the same diagnosis in each case. Overall, the four pathologists mostly explored the regions with cancer and usually the difference in diagnostic times was due to the difficulty in finding the cancerous region. Once this region was located, pathologists frequently focused on local details that supported a refinement of the diagnosis. This navigation profile was not strongly impacted by the type of carcinoma in each slide; i.e., whether it was any one of the unique and mixed BCC types, namely superficial, nodular, micronodular, morpheaform, and trabecular.

Navigation patterns presented by pathologists during diagnosis tasks present a very high variability associated with different factors such as the experience level or the complexity of the visual patterns to identify. Although patterns may be highly variable, previous works have shown that a group of experts

Fig. 5 Validation methodology.

exploring the same histological slide tend to visit similar locations.³⁰ Different pathologists navigate the same WSI assuming that the intersection of their navigations mitigates the noise generated by a single navigation. While four pathologists were part of the experiment, results of previous works suggest that including more pathologists has the potential of improving these results even more.³¹

3.2 Learning Feature Relevance

As defined in Sec. 2.3, the likelihood a nucleus is cancerous is given by the linear combination of its features and the weights of each feature. In order to learn these feature weights, the least squares method was applied to approximate the likelihood of each nucleus by employing the visual attention map. This

Fig. 6 Examples of the predicted likelihood for representative FOVs. The color of each nucleus (dot) represents the relative likelihood of each nucleus being cancerous (a heat map color palette was used, in which blue and red represent low- and high-relevance, respectively). Column A presents the original image, column B shows the likelihood based on low-level features, column C displays the likelihood inferred from navigations of pathologists, and column D illustrates the final likelihood calculated with the presented approach. First row is a visual field with two structures: a carcinoma nodule (bottom) and epidermis (top), second row corresponds to a cancerous area, and third row corresponds to a follicle, a normal/healthy structure.

Table 1 Learned weights for each visual feature.

Feature	Weight
Bias factor	0.3623
Spatial distribution	0.2272
Size	0.0231
Mean nuclear Y	0.1142
Mean nuclear U	-0.0354
Mean nuclear V	-0.0292
Mean Y of the region surrounding a nucleus	-0.0624
Mean U of the region surrounding a nucleus	-0.1355
Mean V of the region surrounding a nucleus	0.0146

estimation process was performed using the “learning set,” i.e., 70% of all the cases.

The least squares method was used to determine the feature weights; the average mean square error for the regression fit was 4.88%. Interestingly, the spatial distribution-based feature (see Sec. 2.1.2) had the highest weight among all the features. This makes intuitive sense given that cancer is characterized by high cellular proliferation and cluster formation. Figure 6 shows some visual results of images whose relevance values were predicted using the learned weights. Table 1 shows the weights learned for each of the image features considered in this study.

3.3 Model Validation

From the “evaluation set,” i.e., 30% of all the cases (seven patients), 274 FOVs were extracted. Then 176 FOVs (from five patients) were used to train a support vector machine (SVM) to predict the presence of cancer on a patch-by-patch basis while the remaining 98 FOVs (from two patients) were used for independent testing.

The accuracy, area under the receiver operating characteristic (ROC) curve, and model performance were compared against the predictions made by a model trained with explicit manual annotations (see Sec. 2.5).

Results, presented in Table 2, show that classification performance is slightly higher when using nuclear features weighted by visual attention information, in contrast to using a model trained with manual annotations.

Figure 7 shows the ROC curve corresponding to the SVM. The solid line corresponds to the ROC curve for implicit model (M_{im}), where the area under the curve (AUC) is equal to 0.7771. In contrast, the dashed line represents the baseline (M_{ex}) ROC curve with an $AUC = 0.7750$. Although differences are not meaningful, these results demonstrate that

Fig. 7 ROC curves for correctly identifying cancer in the test set of FOVs using the integrated and the baseline.

the presented approach, based on knowledge implicitly extracted from pathologists, can be used instead of the classical approach that requires specific manual annotations.

4 Discussion

In this work, we introduced a new computational model that takes advantage of low-, mid-, and high-level image information to predict the likelihood of cancer presence in WSIs. Low-level information was extracted from nuclear visual properties (spatial distribution, area, and color), and high level information was extracted from visual attention maps, generated from pathologists interactions during diagnostic tasks. Our approach was able to identify ROIs within the WSIs that seemed to be critical in predicting the presence of cancer; areas on which the pathologists tend to focus when making their diagnostic decision. As previously stated, there is a wide body of evidence that abnormal tissue is characterized by nuclear patterns that can substantively inform the pathologists’ decision.⁶ These patterns are interpreted by expert pathologists in ways that cannot easily be captured by simple models. As a first step, our experiments used nuclear features extracted from H&E slides of BCC samples to train a classifier to identify whether or not a given FOV contained cancer. This model was informed using an automated information gathering process that identified what ROIs in tissue slides a pathologist typically tends to focus on, and used that information to interpret the significance of nuclei features. While in this work we demonstrated the applicability of the new model in the diagnosis of BCC alone, this approach may be able to help incorporate expert knowledge into the interpretation of a wide variety of disease types.

The method was evaluated in terms of its ability to determine whether a set of FOVs extracted from BCC WSIs contained cancerous regions. For this purpose, the whole dataset was randomly split into a learning set and an evaluation set. The learning set was used to estimate the weights of a multilinear regression (the least squares method) of visual features, aiming

Table 2 Performance metrics for identifying the presence of cancer in a set of FOVs.

	Accuracy	Precision	Recall (sensitivity)	F -measure	Specificity
Implicit model	74.49	75.00	86.44	80.31	56.41
Baseline	73.47	77.97	77.97	77.97	66.67

to approximate the visual attention maps generated by interactions of a pathologist. Subsequently, the WSIs in the experimentation set were split into 274 FOVs. Then 176 FOVs (from five patients) were used to train an SVM learning classifier to predict which FOVs were cancerous or noncancerous. The remaining 98 FOVs (from two patients) were employed as the test set.

The SVM yielded an accuracy of 74.49% and an *F*-measure of 80.31%, in turn representing an improvement of 1.02% and 2.34% over a baseline representation using a model trained with manual annotations. The ROC curve of this SVM classifier yielded an $AUC = 0.7714$ for the integrated model and $AUC = 0.7505$ for the baseline.

Our approach aims to identify diagnostically relevant information from the WSI by explicitly modeling and integrating attributes pertaining to the interaction of an expert with a WSI. The method may potentially be prone to different types of noise. An example may be where the cancer region is too small, with many healthy cells surrounding a region of interest, e.g., micronodular carcinoma. The results presented herein suggest that even if different subtypes of skin cancer manifest on the slides, at the level of the cell, they may still be similar. The model aims to estimate the primary relationships in the feature space and uses these learned relationships to then predict the tissue class. In this particular case, since some data might not be the most class representative, linear regression was performed on about 830,000 different nuclei from 17 different WSIs. Our results suggest that the use of a large number of training exemplars might help to offset concerns regarding image noise.

Experimental results appear to suggest the following two issues. First, handcrafted nuclear features (spatial distribution, area, and color) were found to be independently discriminating of BCC and benign regions. Second, the presented approach, using implicit knowledge from pathologists, has a slightly better performance than the very same model but trained using manual annotations of individual nucleus. Consequently, this approach shows a high potential of being used in real scenarios since experts are not asked to manually annotate the relevant structures, an obviously time consuming and prone to error task. Therefore, relevant information can be passively collected during pathologists' routine tasks (e.g., performing a diagnosis or teaching) for posterior analysis.

While visual saliency is a useful clue for focusing on relevant information, these maps tend to incorrectly suppress targets and pop-out distractors.⁴⁷ Past experience suggests that most models cannot replicate the ability of expert knowledge to describe entire populations of nuclei.⁶ Different experiments^{48,49} have shown that when the visual process is task-driven, identification of ROIs requires complex interaction between two complementary sources of information. These include the bottom-up flow of information coming from low-level image features such as color, edges, intensity or texture, and the top-down flow of expert knowledge, which establishes relationships between experience and a particular application.³⁰ In this work, visual information coming from nuclear handcrafted features was combined with high-level information obtained from visual attention maps. These maps captured the focus and visual attention of expert pathologists during a focused cancer identification task, demonstrating that higher-level understanding can complement and enrich the prediction possible by low-level image features.

In a recent work,⁵⁰ authors presented an approach for breast histopathology image segmentation in which an image is represented as a graph using an Euclidean spatiocolor-texture distance-based similarity. This work utilized similar mathematical structures by introducing a graph-based model that assigns higher likelihood to clustered nodes and area-related homogeneous neighborhoods. Experimental results showed that these features were useful for tasks such as cancer detection. However, these results also indicate that this information is enriched by incorporating higher level information pertaining to the regions a pathologist tends to focus on while rendering a diagnosis, and not necessarily the rationale behind the diagnosis itself.

Our work did, however, have its limitations. A major limitation of the presented approach was the size of the dataset. Since this work required the digitization and manual annotation of WSIs, and the capture of diagnostic navigation paths of four different pathologists, it was necessary to limit the number of images due to the limited time availability of our experts. Another limitation of this work is that we focused only on selective features of nuclear morphology and architecture while ignoring potential value that could be derived from subvisual features such as from the stroma.⁵¹ Interestingly, pathologists do not currently spend a great deal of time interrogating stromal architecture, but recent results suggest that tumors adjacent normal appearing regions might contain substantial prognostic information.⁵² Future work will include validation of our approach on a larger independent test set, evaluation of our approach on other use cases apart from BCC, as well as the consideration of new visual and architectural nuclear features.

5 Conclusions

In this paper, an integrated computational model fuses low-, mid-, and high-level image information to predict the cancer likelihood of WSIs. Some state-of-the-art approaches use visual features alone for predictive tasks, which ignore the high-level knowledge from pathologists and are prone to errors due to the semantic gap. Other approaches are based on manual annotations, which are time consuming. In contrast, the presented strategy infers the nuclei relevance from the spatial focus of pathologists during a diagnostic task. Unlike other works, in this approach high-level knowledge is passively/implicitly extracted from pathologists' interactions with the WSI, negating the need for time-consuming annotations. This high-level information is used to determine the relative importance (weight) of each low-level feature to predict the likelihood that each nucleus in a new image was cancerous. Experiments show that the presented approach has a comparable performance with an approach based on explicit manual annotations. We consider that this approach has a lot of potential to be used in real clinical scenarios due to the speed and ease with which it can absorb higher-level information from physicians' routine tasks and implement that knowledge to generate objective diagnoses.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

The authors are indebted to the personnel of Department of Pathology of Universidad Nacional de Colombia, Colombia, who supplied the slides, provided advising, and helped to perform the experiments. This project was supported by project

number 23664 funded by Universidad Nacional de Colombia by means of “Convocatoria del programa nacional de proyectos para el fortalecimiento de la investigación, la creación y la innovación en posgrados de la Universidad Nacional de Colombia” and project number 850-2015 funded by Colciencias by means of “Convocatoria 711-2015 para Proyectos de Ciencia, Tecnología e Innovación en Salud.” Research reported in this publication was also supported by the National Cancer Institute of the National Institutes of Health under award numbers 1U24CA199374-01, R01CA202752-01A1, R21CA179327-01, and R21CA195152-0; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02; the DOD Prostate Cancer Synergistic Idea Development Award (PC120857); the DOD Lung Cancer Idea Development New Investigator Award (LC130463); the Case Comprehensive Cancer Center Pilot Grant VelaSano Grant from the Cleveland Clinic; the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering at Case Western Reserve University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. A. Madabhushi and G. Lee, “Image analysis and machine learning in digital pathology: challenges and opportunities,” *Med. Image Anal.* **33**, 170–175 (2016).
2. R. Bhargava and A. Madabhushi, “Emerging themes in image informatics and molecular analysis for digital pathology,” *Annu. Rev. Biomed. Eng.* **18**(1), 387–412 (2016).
3. A. Madabhushi, “Digital pathology image analysis: opportunities and challenges,” *Imaging Med.* **1**(1), 7–10 (2009).
4. Y. Xu et al., “Weakly supervised histopathology cancer image segmentation and classification,” *Med. Image Anal.* **18**(3), 591–604 (2014).
5. M. N. Gurcan et al., “Histopathological image analysis: a review,” *IEEE Rev. Biomed. Eng.* **2**, 147–171 (2009).
6. L. Roa-Peña, F. Gómez, and E. Romero, “An experimental study of pathologist’s navigation patterns in virtual microscopy,” *Diagn. Pathol.* **5**(1), 71 (2010).
7. S. Somanchi and D. B. Neill, “Discovering anomalous patterns in large digital pathology images,” in *Proc. of the 8th INFORMS Workshop on Data Mining and Health Informatics* (2013).
8. D. B. Nagarkar et al., “Region of interest identification and diagnostic agreement in breast pathology,” *Mod. Pathol.* **29**, 1004–1011 (2016).
9. G. Bueno et al., “New trends of emerging technologies in digital pathology,” *Pathobiology* **83**, 61–69 (2016).
10. V. Kumar, A. K. Abbas, and J. C. Aster, “Cell injury, cell death and adaptations,” in *Robbins Basic Pathology*, Chapter, pp. 1–29, Saunders-Elsevier (2013).
11. S. Kothari et al., “Pathology imaging informatics for quantitative analysis of whole-slide images,” *J. Am. Med. Inform. Assoc.* **20**, 1099–1108 (2013).
12. X. Moles-Lopez et al., “Whole slide imaging and analysis for biomarker evaluation in digital pathology,” in *Microscopy: Advances in Scientific Research and Education*, A. Méndez-Vilas, Ed., pp. 776–787, Formatex Research Center, Extremadura, Spain (2014).
13. T. W. Cai, J. Kim, and D. D. Feng, Chapter 4, “Content-based medical image retrieval,” in *Biomedical Information Technology*, pp. 83–113, Academic Press, Burlington (2008).
14. Q. Zhang and E. Izquierdo, “Histology image retrieval in optimized multi-feature spaces,” *IEEE J. Biomed. Health Inform.* **17**, 240–249 (2013).
15. U. Ravindran and T. Shakila, “Content based image retrieval for histology image collection using visual pattern mining,” *Int. J. Sci. Eng. Res.* **4**(4) (2013).
16. X.-Y. Wang et al., “A new SVM-based active feedback scheme for image retrieval,” *Eng. Appl. Artif. Intell.* **37**, 43–53 (2015).
17. S. Doyle et al., “A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies,” *IEEE Trans. Biomed. Eng.* **59**, 1205–1218 (2012).
18. H. Sharma et al., “A review of graph-based methods for image analysis in digital histopathology,” *Diagn. Pathol.* **1**(61) (2015).
19. G. Li et al., “Effective graph classification based on topological and label attributes,” *Stat. Anal. Data Min.* **5**(4), 265–283 (2012).
20. A. N. Basavanthally et al., “Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology,” *IEEE Trans. Biomed. Eng.* **57**, 642–653 (2010).
21. G. Lee et al., “Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients,” *PLoS One* **9**(5), e97954 (2014).
22. J. Prewitt, “Graphs and grammars for histology: an introduction,” in *Proc. of the Annual Symp. on Computer Application in Medical Care*, pp. 18–25 (1979).
23. S. Ali et al., “Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays,” *Proc. SPIE* **8676**, 86760H (2013).
24. S. Ali, J. Lewis, and A. Madabhushi, *Spatially Aware Cell Cluster (SpACCL) Graphs: Predicting Outcome in Oropharyngeal p16+ Tumors*, pp. 412–419, Springer, Berlin, Heidelberg (2013).
25. T. V. Gopal, P. P. Rani, and K. Kumar, “Relevance feedback: a proposed method for reducing semantic gap in content-based image retrieval using images with relevance level,” *Int. J. Eng. Sci. Res.* **4**(2), 842–850 (2013).
26. Y. Liu et al., “A survey of content-based image retrieval with high-level semantics,” *Pattern Recognit.* **40**(1), 262–282 (2007).
27. B. Reljin et al., “Adaptive clustering of image database (ACID) as an efficient tool for improving retrieval in a CBIR system,” *Ser. Stud. Health Technol. Inf.* **179**, 172–183 (2012).
28. L. Peter et al., “Leveraging random forests for interactive exploration of large histological images,” *Lect. Notes. Comput. Sci.* **8673**, 1–8 (2014).
29. E. Krupinski et al., “Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience,” *Hum. Pathol.* **37**(12), 1543–1556 (2006).
30. F. Gómez, D. Marín, and E. Romero, “A soft-cache strategy for pathologist’s navigation in virtual microscopy,” *Microsc. Res. Tech.* **74**(5), 401–414 (2011).
31. G. Corredor and E. Romero, “Learning histopathological regions of interest by fusing bottom-up and top-down information,” in *IEEE Int. Conf. on Image Processing (ICIP ‘15)*, pp. 3200–3204 (2015).
32. A. Cruz-Roa et al., “A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI ‘13): 16th Int. Conf., Proc., Part II*, Nagoya, Japan, September 22–26, 2013, pp. 403–410, Springer, Berlin, Heidelberg (2013).
33. D. Sharma and S. Srivastava, “Automatically detection of skin cancer by classification of neural network,” *Int. J. Eng. Tech. Res.* **4**, 15–18 (2016).
34. S.-J. Baek et al., *Detection of Basal Cell Carcinoma by Automatic Classification of Confocal Raman Spectra*, pp. 402–411, Springer, Berlin, Heidelberg (2006).
35. S. Ali and A. Madabhushi, “An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery,” *IEEE Trans. Med. Imaging* **31**, 1448–1460 (2012).
36. H. Fatakdawala et al., “Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): application to lymphocyte segmentation on breast cancer histopathology,” *IEEE Trans. Biomed. Eng.* **57**, 1676–1689 (2010).
37. A. Janowczyk and A. Madabhushi, “Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases,” *J. Pathol. Inform.* **7**(1), 29 (2016).
38. H. Xu, C. Lu, and M. Mandal, “An efficient technique for nuclei segmentation based on ellipse descriptor analysis and improved seed detection algorithm,” *IEEE J. Biomed. Health Inform.* **18**, 1729–1741 (2014).
39. S. Wienert et al., “Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach,” *Sci. Rep.* **2**, 503 (2012).
40. L. Winsor, *Laboratory Histopathology*, Churchill Livingstone, London, pp. 1–42 (1994).

41. L. Alfaro et al., "Compatibility of virtual microscopy equipment: an analysis of different panoramic image software," *Rev. Esp. Patol.* **44**(1), 8–16 (2011).
42. L. Pantanowitz et al., "Whole slide imaging for educational purposes," *J. Pathol. Inform.* **3**(1), 46 (2012).
43. G. Corredor, E. Romero, and M. Iregui, "An adaptable navigation strategy for virtual microscopy from mobile platforms," *J. Biomed. Inf.* **54**, 39–49 (2015).
44. M. Iregui, F. Gómez, and E. Romero, "Strategies for efficient virtual microscopy in pathological samples using JPEG2000," *Micron* **38**(7), 700–713 (2007).
45. D. Taubman and M. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Norwell, Massachusetts (2001).
46. S. Miller, *The Method of Least Squares*, Mathematics Department, Brown University, Rhode Island (2006).
47. J. Li, Y. Tian, and T. Huang, "Visual saliency with statistical priors," *Int. J. Comput. Vision* **107**(3), 239–253 (2014).
48. A. Treisman and S. Gormican, "Feature analysis in early vision: evidence from search asymmetries," *Psychol. Rev.* **95**(1), 15–48 (1988).
49. J. Duncan and G. W. Humphreys, "Visual search and stimulus similarity," *Psychol. Rev.* **96**(3), 433–458 (1989).
50. A. Belsare et al., "Breast histopathology image segmentation using spatiocolour-texture based graph partition method," *J. Microsc.* **262**(3), 260–273 (2016).
51. M. Sharma et al., "Analysis of stromal signatures in the tumor micro-environment of ductal carcinoma in situ," *Breast Cancer Res. Treat.* **123**, 397–404 (2010).
52. G. Lee et al., "Nuclear shape and architecture in benign fields predict biochemical recurrence in prostate cancer patients following radical prostatectomy: preliminary findings," *Eur. Urol. Focus* (2016).

Germán Corredor received his bachelor's and MSc degrees in computer engineering from the Universidad Nacional de Colombia,

Colombia, in 2010 and 2012, respectively. Currently, he is a PhD student at the Universidad Nacional de Colombia. His research is related to the medical image analysis field. His professional interests are software engineering, medical information systems, web and mobile development, development of videogames, and artificial intelligence.

Jon Whitney is a postdoctoral student and Hartwell fellow at Case Western Reserve University. He focuses on digital histopathology and deep learning nuclear segmentation. He has industrial development experience at Lexmark International and has developed methods of 3-D viability measurement and digital diagnosis.

Viviana Arias received her diploma as a medical doctor from the Universidad Libre, Colombia, and as a specialist in anatomical and clinical pathology from the Universidad Nacional de Colombia, Colombia. She also got an international certification in dermatology in 2015. Currently, she works as an associate professor at the Department of Pathology, Universidad Nacional de Colombia.

Anant Madabhushi is the director of the Center for Computational Imaging and Personalized Diagnostics and an associate professor at Case Western Reserve University. He received his bachelor's degree in biomedical engineering from Mumbai University, India, in 1998, and his master's degree in biomedical engineering from the University of Texas, Austin, in 2000. In 2004, he obtained his PhD in bioengineering from the University of Pennsylvania.

Eduardo Romero received his diploma as a medical doctor from the Universidad Nacional de Colombia, Colombia and his magister degree in electrical engineering from the Universidad de los Andes, Colombia, in 1995. He obtained his PhD in biomedical sciences from the Universit Catholique de Louvain, Belgium, in 2000. Currently, he is an associated professor at the Universidad Nacional de Colombia and leads both the CIM@LAB group and the Biomedical Engineering Postgraduate Program.