Medical Image Analysis 30 (2016) 60–71

Contents lists available at ScienceDirect

Medical Image Analysis

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

Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles☆

Jocelyn Barker[c], Assaf Hoogi[a], Adrien Depeursinge[a,b], Daniel L. Rubin[a,c,*]

[a] Department of Radiology, Stanford University School of Medicine, CA, USA
[b] Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
[c] Department of Medicine (Stanford Biomedical Informatics Research), Stanford University School of Medicine, CA, USA

ARTICLE INFO

Article history:
Received 25 February 2015
Revised 2 December 2015
Accepted 7 December 2015
Available online 29 December 2015

Keywords:
Digital pathology
Computer aided diagnosis
Object classification

ABSTRACT

Computerized analysis of digital pathology images offers the potential of improving clinical care (e.g. automated diagnosis) and catalyzing research (e.g. discovering disease subtypes). There are two key challenges thwarting computerized analysis of digital pathology images: first, whole slide pathology images are massive, making computerized analysis inefficient, and second, diverse tissue regions in whole slide images that are not directly relevant to the disease may mislead computerized diagnosis algorithms. We propose a method to overcome both of these challenges that utilizes a coarse-to-fine analysis of the localized characteristics in pathology images. An initial surveying stage analyzes the diversity of coarse regions in the whole slide image. This includes extraction of spatially localized features of shape, color and texture from tiled regions covering the slide. Dimensionality reduction of the features assesses the image diversity in the tiled regions and clustering creates representative groups. A second stage provides a detailed analysis of a single representative tile from each group. An Elastic Net classifier produces a diagnostic decision value for each representative tile. A weighted voting scheme aggregates the decision values from these tiles to obtain a diagnosis at the whole slide level. We evaluated our method by automatically classifying 302 brain cancer cases into two possible diagnoses (glioblastoma multiforme ($N = 182$) versus lower grade glioma ($N = 120$)) with an accuracy of 93.1 % ($p << 0.001$). We also evaluated our method in the dataset provided for the 2014 MICCAI Pathology Classification Challenge, in which our method, trained and tested using 5-fold cross validation, produced a classification accuracy of 100% ($p << 0.001$). Our method showed high stability and robustness to parameter variation, with accuracy varying between 95.5% and 100% when evaluated for a wide range of parameters. Our approach may be useful to automatically differentiate between the two cancer subtypes.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Gliomas account for 28% of primary brain tumors and tumors of the central nervous system, and make up 80% of malignant brain tumors, affecting Americans at a rate of 5.83 per 100,000 in 2007–2011 (Ostrom et al., 2014). Of those, glioblastoma multiforme (GBM, WHO grade 4) represents 54.7%, and lower grade gliomas (LGG, astrocytomas, oligodendrogliomas, and oligoastrocytomas, WHO grade 2 and 3) represent 24.2% of the gliomas. The prognosis for GBM is dire with a 1-year survival at 36.5%, with only 5.0% surviving 5-years, whereas the 5-year survival for LGG is much higher at 51.4%. Not only does prognosis differ for patients, but the choice of therapy differs as well (Khasraw and Lassman, 2010). Many LGG subtypes have curative treatments, but only palliative treatments exist for GBM (Preusser et al., 2011). Thus it is of critical importance to differentiate GBM from LGG when making decisions on patient treatment options.

The primary way of differentiating GBM from LGG is through histopathology. The primary distinguishing features of GBM include vascular thrombosis, microvascular proliferation, and necrosis (Cavenee et al., 2007). However, these features are not always clear, and they can be difficult to find and recognize in whole slide images, likely accounting for the high inter-reader variability observed between pathologists (Coons et al., 1997).

Three common subtypes of LGG are astrocytomas, oligodendrogliomas, and oligoastrocytomas. Astrocytomas develop from the star-shaped astrocytes (Fig. 1a). While astrocytomas have some

☆ "This paper was recommended for publication by Nicholas Ayache".

* Corresponding author at: Department of Radiology, Stanford University School of Medicine, CA, USA .
E-mail addresses: jbarker2@stanford.edu, jocelyn.e.barker@gmail.com (J. Barker), ahoogi@stanford.edu (A. Hoogi), adrien.depeursinge@hevs.ch (A. Depeursinge), dlrubin@stanford.edu (D.L. Rubin).

http://dx.doi.org/10.1016/j.media.2015.12.002
1361-8415/© 2015 Elsevier B.V. All rights reserved.

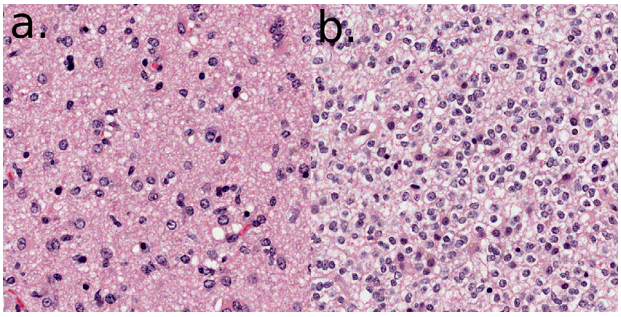

Fig. 1. Sample images of lower grade gliomas showing (a) a grade 3 anaplastic astrocytoma and (b) a grade 2 oligodendroglioma. Note the parinuclear halos in (b), seen as a decrease in the pink eosin stain surrounding the nuclei. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

molecular commonalities, there is no unifying histopathological phenotype, rather multiple subtypes exist, each with their own identifying set of features. Oligodendrogliomas develop from the branching oligodendrocytes (Fig. 1b). Like the oligodendrocytes they derive from, oligodendrogliomas have a perinuclear halo, giving the tumor cells a "fried egg" appearance. Oligoastrocytomas tumors have a mixed population with some cells showing a more astrocytoma like phenotype and others appearing more like oligodendroglioma cells. Lower grade gliomas, as defined by The Cancer Genome Atlas (TCGA), comprise grade 2 and 3 gliomas, where grade 2 tumors show only cytological atypia and grade 3 tumors also show anaplasia and mitotic activity.

Glioblastoma multiforme is a grade 4 astrocytic tumor. As the descriptor "multiforme" in the name suggests, GBM varies in appearance. While all tumors are highly cellular with high degrees of anaplasia and mitotic activity, some tumors have nuclei that are highly polymorphic, containing many giant multinucleated cells, whereas others are monotonous. While astrocytic features are easily identified in some tumors, they are difficult to distinguish in others due to extreme anaplasia. The primary distinguishing features of GBM include vascular thrombosis, microvascular proliferation, and necrosis (particularly pseudopalisading necrosis) (Fig. 2). The remarkable regional heterogeneity of the tumors makes them challenging to diagnose if these distinguishing features are not present in regions of the pathology image analyzed.

We believe that computer aided diagnosis applied to digital pathology images can help make the histopathological designation of glioma subtypes clearer by providing reproducible and exhaustive image analysis (Hamilton et al., 2014). In addition, adoption of such computerized methods may improve the pathologists' workflow, and reduce in inter-reader variability. Moreover, in order to develop computer aided diagnosis systems for pathology, it is important not only to analyze the data present in the entire WSI, but also be able to filter out the important signal (i.e., relevant regions in the WSI) from the background. In this paper we describe a completely automated method for classifying whole slide digital pathology images of brain tumors into the GBM and LGG diagnostic categories. It allows for the analysis of the whole tumor through a coarse profiling stage while reducing the noise present in the sample by selecting a smaller number of representative regions. The final step consists of weighted region voting, which allows image regions with strong signal to outweigh signals arising from less relevant (i.e., non-signal containing) regions present in the slide.

The paper's organization is as follows. Section 2, Background, discusses the strengths and weaknesses of related work. Section 3, Materials and methods, describes the dataset and preprocessing in Section 3.1, the feature collection and selection of profiled regions in Section 3.2, and the modeling and weighted voting in Section 3.3. Section 4, Results, describes the results, in terms of prediction accuracy in Section 4.1, investigates the different feature groups importance in Section 4.2, the impact of representative regions on final WSI decision in Section 4.3, and model stability in Section 4.4. Finally, Section 5, Discussion and conclusions, describes our interpretation of the success of the model and its implications for the future of computer aided diagnosis of digital pathology images.

2. Background

While computer aided diagnosis has become common in radiology (Doi, 2007), there have been challenges in realizing similar successes in pathology. Only a fraction of pathology practices digitize their slides (Onega et al., 2014). Even if digital images were more widely available, there are two key challenges to computerized analysis of whole-slide pathology images (WSI). First, pathology images are large with each image consisting of around 10^{10} pixels, and processing this large amount of information can be computationally expensive. Second, large portions of pathology images contain non-tumor tissue, which is not necessarily relevant to the diagnosis. Signals arising from substantial tissue regions not directly relevant to the diagnosis may mislead computerized diagnosis algorithms. Two approaches have been undertaken to attempt to overcome these obstacles, subsetting the image and tiling the image.

2.1. Image subsetting methods

Subsetting the image is the most common approach, where one selects a small region of the image for processing, simultaneously reducing the image size and potentially extraneous tissue present in the image. Even before WSI was practical, researchers captured a single frame of an entire slide using a standard microscope creating a subset of the image for computational analysis (Adiga et al., 2006; Aiad et al., 2009; Altunbay et al., 2010; Baak et al., 1981; Brook et al., 2007; Esgiar et al., 2002, 2002; Farjam et al., 2007; Fatima et al., 2014; Glotsos et al., 2008; Gunduz et al., 2004; Gupta et al., 2001; Hall et al., 2008; Huang and Lai, 2010; Huang and Lee, 2009; Jafari-Khouzani and Soltanian-Zadeh, 2003; Kong et al., 2007; Land et al., 2008; Lessmann et al., 2007; Meng et al., 2010; Ozolek et al., 2014; Qureshi et al., 2008, 2007; Sboner et al., 2003; Schnorrenberg et al., 1997; Sudbø et al., 2000; Tahir and Bouridane, 2006; Tasoulis et al., 2003; Thiran and Macq, 1996; Tsai et al., 2009). For many researchers doing pathology image analysis today, this is still the best option, either because they do not have access to a WSI system, or because they can capture data not possible with a WSI system, such as confocal microscopy images.

A similar approach is to manually designate a region in the WSI for analysis (Al-Kadi, 2010; Basavanhally et al., 2010, 2008; Doyle et al., 2012a, 2008, 2007; Dundar et al., 2011, 2010; Qureshi et al., 2008; Sertel et al., 2010). This is the digital equivalent of the subsetting process described above: in both procedures, a human uses their knowledge and time to reduce the data available for analysis based on their evaluation of what is informative.

In 2014, the MICCAI Pathology Classification Grand Challenge allowed methods to compete in the task of classifying LGG and GBM images. Three methods used manual subsetting to select regions for analysis, though each method differed greatly in their analysis procedure. Bueno et al. (2014) used 1st and 2nd order Haralick features obtained from texton maps of the Hb, Luv, and SCT color space, and modeled them using an AdaBoost machine. Chang et al. (2014) morphologically characterized and modeled the nuclei using a spatial pyramid matching framework. Xu et al. (2014a, 2014b) characterize each image type using convolutional

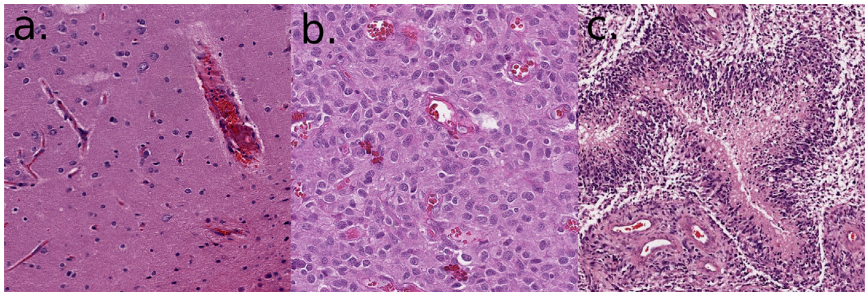

Fig. 2. Three features distinguishing glioblastoma multiforme, (a) vascular thrombosis, (b) microvascular proliferation, and (c) pseudopalisading necrosis.

neural networks. Since convolutional neural networks require large amounts of training data, this method went beyond the designated 45 training data images, using additional 3076 images.

Another group has specifically applied this tactic to the problem of distinguishing lower grade gliomas from glioblastoma multiforme (Mousavi et al., 2015) in a larger dataset of 132 images. In their method, a pathologist selects multiple regions of interest from each histopathological slide. Classification identifies regions of interest containing two hallmarks of GBM: microvascular proliferation and pseudopalisading necrosis. It may be difficult to generalize the method to other problems as the method requires a specific tissue type to identify that defines each class (i.e. microvascular proliferation and pseudopalisading necrosis). To generalize the method for new cancers, the researchers would have to determine a new tissue type before being able to train the method for this new application. If an appropriate tissue type is not known for the disease, the method would not be applicable. Additionally, for a problem like distinguishing GBM from LGG, Mousavi et al. require selection of region of interest for analysis (e.g. pseudopalisading necrosis). If the target features are not present in the image, a random region of interest representative of the slide is input instead. This has three key disadvantages. (1) It increases the pathologist's workload due to the fact that in a common clinical procedure regions of interest wouldn't be defined. (2) The process is inefficient as the pathologist has already decided visually if the features are present in the slide in their selection of the region of interest. Therefore, an additional procedure is redundant and time consuming. (3) It can be highly biased in the pathologist's decision to select a targeted versus random region of interest.

An alternative method for subsetting the data uses automatic segmentation to identify regions of interest in the image for analysis. Since these methods are automatic, they are not subjective and do not require a time commitment from an individual with pathology expertise. This method has most commonly been applied to assigning a Gleason grade to prostate cancer samples (Doyle et al., 2012b; Monaco et al., 2008; Monaco et al., 2010; Naik et al., 2008, 2007; Sparks and Madabhushi, 2013) though it has found applications in other cancers as well (Petushi et al., 2006). The Gleason scoring system is particularly well suited to this type of analysis because the system characterizes how ordered the glands are in a prostate tumor. In other applications where the task is not so rigidly defined, determining a target for segmentation may be a limiting factor in the method's applicability.

All of these approaches to subsetting WSI share a common limitation; a human must decide what data is valuable for analysis. This limits the analysis' ability to discover signal in previously unappreciated regions of the slide.

2.2. Image tiling methods

A second approach to tackling the challenges to computerized analysis of WSI uses tiling to break the image into small, manageable pieces (Kong et al., 2013, 2009; Ruiz et al., 2007; Sertel et al., 2009), analyzing each image piece, as opposed to just examining a few selected regions. However this increase in data means an increase in computation complexity. In order to characterize the oligodendroglioma versus astrocytoma components of 117 GBM samples, Kong et al. (2013) segmented and classified over 200 million nuclei. Based on this high computational complexity, it is easy to understand why Ruiz et al. (2007) focused their work on using a GPU to reduce the execution time of their stromal classification in neuroblastoma. An alternative approach to reduce computational load while still examining the entire WSI is to reduce the image resolution. For their stromal classification in neuroblastoma, Sertel et al. (2009) used a multi-resolution approach to classify their slides, starting with low resolution and continuing to with higher one in the case of a weak classifier decision.

While these approaches have managed to leverage the vast quantities of information available in the WSI, there is room for improvement in determining the final decision. Both stromal classification and oligodendroglioma versus astrocytoma characterizations classify the slide based on the area occupied by a particular tissue type. The methods classify individual tiles as the clinically relevant tissue type, and then classify the entire image by determining the dominant tissue type in the slide. This appears to work well for diseases where the tissue is homogeneous, and all affected tissue shares a common appearance distinct from phenotypes seen in alternative diagnoses. However, these approaches are unsuitable for diseases where a local feature determines the diagnosis for the slide or where the disease is heterogeneous, such as in the LGG versus GBM classification. In these cases, the majority of the tissue may be more similar to the incorrect classification, but a small region of the slide containing a key disease feature (e.g., pseudopalisading necrosis) determines the diagnosis. Additionally, many of them do not perform well in images containing a large amount of extraneous tissue, as this tissue can bias their classification.

3. Materials and methods

Fig. 3 depicts an overview of our image processing pipeline, which is further detailed in Sections 3.2 and 3.3. The pipeline uses a "coarse-to-fine" approach, where "coarse profiling" assesses image content diversity, followed by "fine profiling" to produce in the final evaluation. First, the pipeline tiles input WSIs and coarsely characterizes them by extracting shape, color, and texture features (i.e., "coarse features") from each tile. It clusters all tiles contained within an individual image and identifies tiles nearest the cluster center to select a representative subset of tiles while fostering their diversity. In a second step, the pipeline carries out a deeper characterization of the representative tiles by extracting a more comprehensive set of image features from them (i.e., "fine features"). This two-stage approach avoids having to extract all features from all tiles and the corresponding intractable computational load. An Elastic Net classifier (Zou and Hastie, 2005) provides a decision

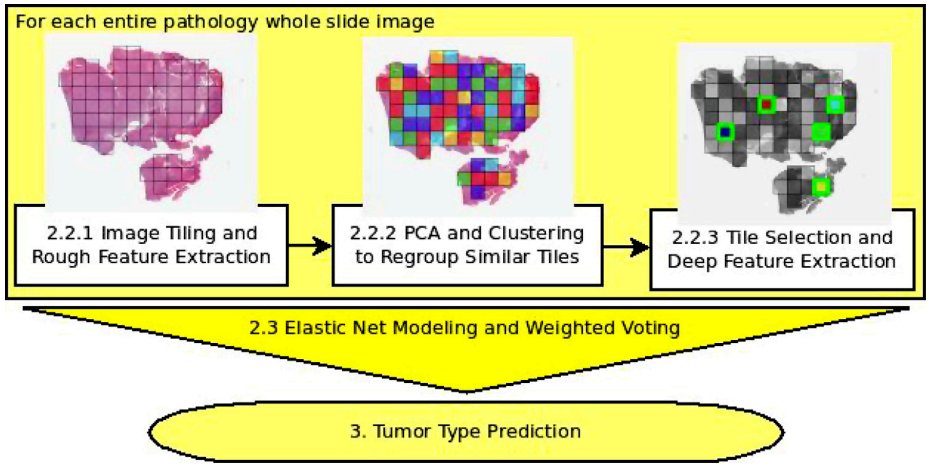

Fig. 3. Pipeline for pathology analysis. Images show a scaled down representative whole-slide image with boxes representing tiles. Unique colors represent clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

value for each tile. A weighted voting scheme aggregates these decision values to obtain a global decision at the whole-image level.

3.1. Dataset and preprocessing

3.1.1. Dataset description

Whole slide, diagnostic images were obtained from The Cancer Genome Atlas (TCGA). All slides were stained with hematoxylin and eosin. TCGA data derives from multiple institutions and is collected over many years, so results found in these data are expected to hold for other studies. Each slide had been scanned at multiple resolutions ranging from 2.5 to 40×, with all images containing a maximum resolution scan of at least 20×. For slides where the maximum resolution was 40×, bicubic interpolation (Hou and Andrews, 1978) resized the images to 20×. The ground truth is that designated by The Cancer Genome Atlas based on patient records.

Two datasets each evaluated of different aspects of the method. The first dataset comprised 45 images from 2014 MICCAI Pathology Classification Grand Challenge. This dataset allowed for direct comparison of the method to existing methods that were evaluated in that challenge. These 45 images included two types of brain cancer: 23 images of glioblastoma multiforme (GBM) and 22 images of lower grade glioma (LGG). For the second dataset all 604 images (364 GBM, 240 LGG) from TCGA, which had a complete complement of pathology and molecular data were selected to evaluate performance of our method in a larger dataset. No more than one whole slide image came from a single patient.

3.1.2. Tiling the images

WSI tiling created 1024 × 1024 pixel images at 20× resolution. This is the same size and resolution used for tissue microarrays, on which a trained pathologist can use to make informed opinions about a whole tumor. Tiles were adjacent to one another covering the entire tissue region of the slide. Many tiles contain very little tissue, as tissue generally occupies only a small portion of the glass slide. Tiles of the slide containing a minimum of 0.237 mm^2 tissue area were used for further analysis. A tissue area of 0.237 mm^2 comprises 90% of the tile, making tiles with this amount of tissue unlikely to have many artifacts derived from the background glass slide. The following procedure distinguished tissue from background: (1) convert the lowest resolution scan of the image to grayscale. (2) Apply automatic contrast enhancement (Divakar, 2009). (3) Take the 8-bit depth complement. (4) Perform hysteresis thresholding with an experimentally-chosen high threshold of 100 and a low threshold of 50.

3.1.3. Unmixing the stains

In order to better represent the biological information of a sample, the hematoxylin and eosin stains were separated from the original image using the color deconvolution method used in CellProfiler (Kamentsky et al., 2011). This method for color deconvolution uses pseudo inverse matrices as described by Ruifrok and Johnson (Ruifrok and Johnston, 2001). The RGB values used for hematoxylin were [0.644, 0.717, 0.267] and for eosin were [0.093, 0.954, 0.283]. To obtain the hematoxylin and eosin stain values we find

$$\exp\left(\ln\left(RGB_{i,j}\right)\begin{bmatrix}0.644, 0.171, 0.267\\ 0.093, 0.954, 0.283\end{bmatrix}^{+T}\right) = HE_{i,j}$$

where $RGB_{i,j}$ are the red, green, and blue channel values for pixel i,j and $HE_{i,j}$ are the deconvoluted hematoxylin and eosin values for pixel i,j. The symbol $+T$ indicates the transpose of the pseudo-inverse of the matrix. Since hematoxylin binds to nucleotides and eosin to proteins, unmixing the stains allows the method to better probe the roles of these biologically important molecules.

3.1.4. Tissue and nuclei segmentation

The process described in Section 3.1.2 identified tissue regions in the image tiles. A method we derived from Gurcan (Gurcan et al., 2006) was used to segment the nuclei. In the original method, the red channel of the RGB image was complemented and transformed using morphological top-hat reconstruction (Meyer, 1979). Nuclei are segmented from this transformed image using hysteresis thresholding, and clumps of nuclei are split using watershedding. In our modification of their work, the hematoxylin stain was transformed using morphological top-hat reconstruction (Meyer, 1979). An iterative series of hysteresis thresholds defined candidate nuclei. The upper threshold ranged from 150 to 50 and the lower threshold was 0.2× the upper threshold with each iteration, reducing the upper threshold by 5 intensity units. At each iteration, objects meeting size specifications (30–200 pixels) were identified as nuclei. This allows for the identification of a large variety of nuclei with different staining properties.

3.2. Tile-based local image characterization

In order to profile the entire tumor and to sample image diversity, the whole slide image is coarsely profiled, and similar regions are clustered together. The tile deemed most representative of each

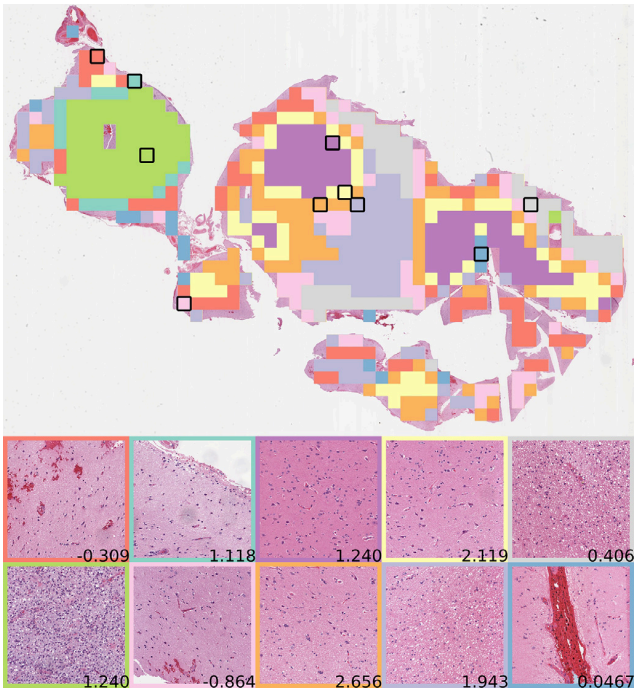

Fig. 4. Sample image showing the results of the image processing pipeline. Unique colors represent clusters. Black boxes represent tiles chosen for "fine" profiling. Images at the bottom show the tiles selected from each cluster, with the color on the border of the tiles indicating the cluster from the image. Numbers at the bottom are the decision values the tiles received after machine learning analysis with positive values indicating a correct decision and negative values indicating an incorrect one. The strong positive values from some tiles (e.g., 2.656 in the orange tile) are able to overcome weak incorrect decisions from others (e.g., −0.309 in the red tile). Additionally, the small decision value for the bleeding artifact in the blue tile in the bottom left, demonstrates the pipeline's ability to minimize the impact of artifacts on the final decision.

cluster receives a more through characterization for tissue modeling. Using this strategy ensures that the many varied regions of the tumor are represented in the model. Since both GBM and LGG have heterogeneous phenotypes, sampling the image in this way can aid in classification. Fig. 4 shows a sample image with clusters as well at the representative tiles and decision values from the model.

3.2.1. "Coarse" feature extraction

Extracting fine features for all tiles may highly increase the computational load. As a first step, coarse feature collection and profiling are computed on all tiles to select a subset of tiles that may serve as representatives of the WSI. Since pathology images may contain a variety of tissue types and other features of interest, the coarse analysis procedure enables the best representation of the tissue diversity in the WSI. Examining cell nuclei is one of the primary ways that pathologists establish diagnoses in pathology images. Therefore, segmenting cell nuclei and collecting features which characterize them is fundamental to quantitative digital pathology image analysis (Gil and Wu, 2003). For this first pass, a simple threshold on the hematoxylin stain segments the nuclei rather than using the iterative hysteresis thresholding described in Section 3.1.4 to save computational time. Then features were extracted from the whole tissue segmentation, from the nuclear segmentation, and from the non-nuclear regions, independently for both the hematoxylin and eosin stains. Specifically, shape, color (Gurcan et al., 2009), Haralick texture features (Haralick et al., 1973), and second order unaligned Riesz features (Depeursinge et al., 2014) were extracted from each tile. The Riesz features correspond qualitatively to a multi-scale Hessian filterbank (Fig. 5) and can describe local organizations of image directions (i.e., for a fixed scale) that are specific to a given texture pattern (i.e., from single oligodendroglial cell appearances to tissue organization like pseudopalisadation). Haralick features were computed from a one pixel displacement, symmetric gray level co-occurrence matrix at 0 and 90 degrees. The 96-feature set comprised 15 shape, 12 color, 26 Haralick, and 36 Riesz features. There were 38 features derived from the hematoxylin stain and 38 features from the eosin stain. There were 13 features derived from the nuclei, 2 from the cytoplasm (regions designated as tissue, but not nuclei), and 81 from the tissue as a whole. On average, it takes ~6.2 s to calculate coarse features from 1024 × 1024 pixel tile.

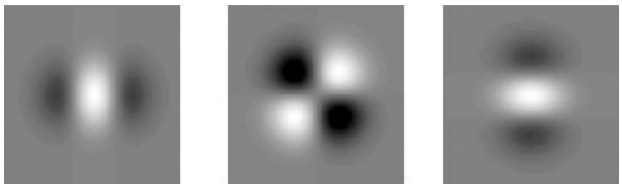

Fig. 5. Second order Riesz filterbank at a fixed scale.

3.2.2. PCA and clustering to regroup similar tiles

After collecting the features from the "coarse analysis", principal component analysis (PCA) reduced the substantial number of "coarse" features, decreasing computational complexity of the classification algorithm. The first few components of the PCA, which are the components with the largest eigenvalues (i.e., capturing most of the variance in the data), serve as a basis for the information description. Three PCA components represented each tile. K-means clustering defined similar tiles based on the reduced-feature collection. The K-Means++ algorithm (Arthur and Vassilvitskii, 2007) allowed for more stable clustering. K-Means++ outperforms classic k-means with random selection of initial seeds. It results in much more stable clustering by reducing the distance from the points to the nearest cluster center in $O(\log(k))$ time. Ten clusters represented various groups of similar tiles. For each cluster, a single representative tile whose values were the closest to the centroid of the cluster were selected to use in the "fine analysis" (Section 3.2.3). Selecting tiles after the clustering procedure ensures that diverse tissue elements in the slide are included in modeling.

3.2.3. "Fine" feature extraction

For the more detailed "fine" analysis, all the feature types collected in the "coarse" analysis (Section 3.2.1) are collected from the representative tiles, with a few changes that increased feature quality at the cost of computational time. Most notably, the features derive from the more detailed nuclear segmentation described in Section 3.1.4. Skeleton features for the nuclei were computed along with the other shape feature to better define nuclear morphology (Zhao and Daut, 1991). Additionally, aligned second-order Riesz features with six dyadic scales are computed to characterize texture in the fine analysis, rather than using the unaligned features used in the coarse analysis (Depeursinge et al., 2014). Alignment of the Riesz wavelets allows for rotation invariance, and avoids imposing arbitrary directionality on the tissue analysis. As in Section 3.2.1, features derive from within the nuclei segmentation, in the cytoplasm segmentation (regions designated tissue but not nuclear), and from the tissue as a whole, as well as from both the hematoxylin and eosin stain. A total of 227 features are extracted, comprising 19 shapes, 15 colors, 78 Haralicks, and 108 Riesz features. There were 101 features derived from the hematoxylin stain and 101 from the eosin stain. Eighty-one

features derived from the nuclei, 64 from the cytoplasm (regions designated as tissue, but not nuclei), and 83 from the tissue as a whole. On average it takes ~185.2 s to calculate fine features from a 1024 × 1024 pixel tile.

3.3. Whole slide classification using Elastic Net classification and weighted voting

The Elastic Net linear regression model (Zou and Hastie, 2005) is used to generate the predicted diagnosis from our data. Elastic Net combines the l1 and l2 penalties from the LASSO and ridge regression model. More specifically, Elastic Net seeks to minimize,

$$\underset{(\beta_0,\boldsymbol{\beta})\in R^{p+1}}{\arg\min}\left[\frac{1}{2}N\sum_{i=1}^{N}\left(y_i-h_{\beta_0,\boldsymbol{\beta}}(\mathbf{x}_i)\right)^2+\lambda P_\alpha(\boldsymbol{\beta})\right]$$

where

$$P_\alpha(\boldsymbol{\beta})=(1-\alpha)\frac{1}{2}\boldsymbol{\beta}_{l_2}^2+\alpha\boldsymbol{\beta}_{l_1}$$

and

$$0\le\alpha\le 1.$$

β_0 is the intercept for the linear regression model, $\boldsymbol{\beta}$ represents the feature weights, $\mathbf{x}_i$ is the feature vector for image i, y_i is the outcome for image i, N the number of training examples, and $h_{\beta 0,\boldsymbol{\beta}}(\mathbf{x}_i)$ is the generalized linear regression hypothesis function. As in other regression models, λ is penalizing the feature coefficients $\boldsymbol{\beta}$. Additionally, Elastic Net includes the parameter α, which balances between the l1 and l2 penalties.

All of the features from the tiles selected in Section 3.2.3 are normalized, and each tile becomes an independent instance. The glmnet implementation (Friedman et al., 2010) of a binomial Elastic Net logistic regression machine with an α value of 0.5 was used to model the data. A five-fold cross-validation on the training set was used to select the λ parameter with the lowest mean cross-validated error.

In order to estimate the generalized classification performance using all images, a 5-fold cross-validation is used. For each iteration of 5-fold cross-validation, we define our training set T for test images $j-j+n$ such that

$$T=\{\mathbf{x}_{i,c},y_{i,c}\}\ \text{for}\ i=1,\ldots,j-1,j+n+1,\ldots I\ \text{and}\ c=1,\ldots,C$$

where $\mathbf{x}_{i,c}$ is the feature vector for image i, n is the number of images required for 5-fold cross validation, cluster c and $y_{i,c}$ is the ground truth for image i, cluster c, I is the number of images and C the number of clusters. Note that in this method of cross-validation, the feature vectors for all clusters from an image are either left out or included in each fold to avoid training and testing the model with tiles belonging to the same image. 5-fold cross-validation produced a total of 5 models and classified the entire dataset.

In order to get the final, aggregated, decision value for a whole image i, we compute the value $\hat{y}$.

$$\hat{y}=\sum_{c=1}^{C}h_{\beta_0,\boldsymbol{\beta}}(\mathbf{x}_{i,c})$$

An appropriate threshold, $\hat{t}$ for the final LGG-versus-GBM classification is selected from the training set such that:

$$\hat{t}=\underset{t}{\text{argmax}}\left[sensitivity(t,\hat{y}_i)+specificity(t,\hat{y}_i)\right]$$
$$\text{for}\ i=1,\ldots,j-1,j+n+1,\ldots I.$$

Final class designation for the test image j was made as $\hat{y}_j>\hat{t}$.

Table 1
Accuracies of the classification methods for the same group of 45 whole slide pathology images. While all methods perform well, only the method proposed here had 100% accuracy.

Method	Accuracy (%)
Barker et al.	**100**
Bueno et al. (2014)	98.1
Chang and Parvin (2014)	85.83
Xu et al. (2014a, 2014b)	97.8

4. Results

4.1. Classification accuracy

4.1.1. Comparison with MICCAI challenge results

The predicted classes of the test images matched the ground truth in all 45 tissue slices, out-performing the methods from the 2014 MICCAI Pathology Classification challenge (Table 1). All other methods used in this comparison included a manual step where regions of the slide were selected for analysis. Since this method was able to out-perform the others, it indicates strong potential for the tile selection methods employed.

4.1.2. Classification of a larger data set

In order to test the robustness of our method, we applied it to the larger dataset of 604 images from TCGA. Half of the WSIs were chosen randomly as a test set, while the remaining half of the cases were used as a training set. The classification accuracy of our method was 93.1%. A one sided binomial test with a no information rate of 0.6 determined this accuracy was highly unlikely to be accounted for by chance ($p << 0.001$). This result supports the robustness of our method and suggests that the high accuracy of our method seen in the MICCAI challenge was not due to sample bias (Fig. 6). Additionally, the ROC curve showed high performance, with an AUC 0.96.

4.2. Importance of the feature groups

4.2.1. Impact of fine features

The impact of the fine features in the final classification was tested. Instead of using the fine features from the selected tiles, the same coarse features used for PCA and clustering were applied. This modified version of the pipeline classified the 604 previously described tumor samples as described in Section 4.1.2. Using only coarse features decreased the classification accuracy from 93.1% to 88.4%. Both the Net Reclassification Improvement (NRI = 0.100, p = 1.12e − 3) and the Integrated Discrimination Improvement (IDI = 0.100, 1.17e − 3) suggest improvement in the model when adding the fine features (Pencina et al., 2008), justifying the computational time.

4.2.2. Use of available feature space

The feature space diversity was well exploited, where shape, color, Haralick, and Riesz features, as well as features from the nuclei and the cytoplasm, and both the hematoxylin and eosin stain were selected by every model built in the cross-validation. One hundred and ninety-one features represented the data in all five folds of cross-validation including 22 shape, 10 color, 70 Haralick, 89 Riesz (Fig. 7). As would be expected based on current histopathological convention where the nuclei are the primary focus, the hematoxylin stain and nuclear-derived features were well represented in the models with 88 hematoxylin and 76 nuclear features. More surprisingly, eosin and cytoplasm features were nearly as well represented with 81 eosin and 49 cytoplasm features, demonstrating the model's ability to discover previously

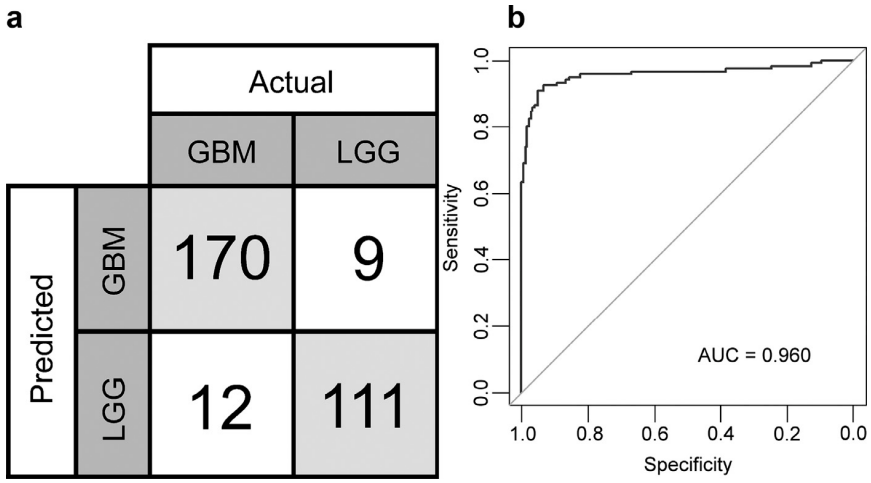

Fig. 6. Accuracies LGG/GBM classification in a larger dataset. (a) A confusion matrix summarizing the dataset classification results. (b) ROC curve demonstrating that the values generated by the model match the predicted classes.

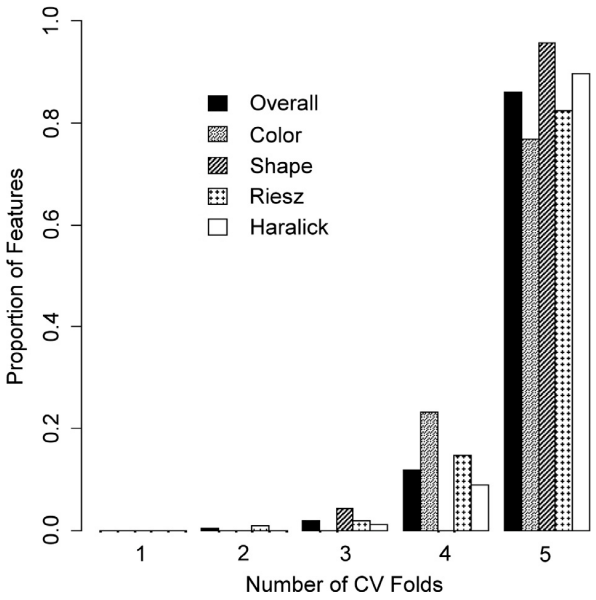

Fig. 7. Histogram distribution of features used in Elastic Net models, showing the number of models in which features of a given class appear. The number of features is normalized based on the total number of features represented for each class. Bar patterns represent the feature class. Most features for each class appears in all, or nearly all models, as would be expected if they have diagnostic value.

unappreciated signal. Additionally, it was observed that the majority of the features appeared in all models indicating stability with respect to the training set.

4.2.3. Contribution of Riesz texture features

The contribution of the Riesz texture features to the accuracy of the model was tested using two experiments. The pipeline classified the 604 TCGA images that were divided as described in Section 4.1.2 two ways: (1) excluding the Riesz features and (2) using only the Riesz features.

Classifying the images without using Riesz features reduced the accuracy to 90.4% (NRI = 0.067, $p = 1.56e-2$; IDI = 0.067, $p = 1.60e-2$) from the original pipeline. Using the Riesz features alone reduced the accuracy to 90.0% (NRI = 0.058, $p = 1.95e-2$; IDI = 0.058, $p = 1.62e-2$). In order to evaluate if the better fit was simply a product of increasing the number of features, two models were compared using the Rao's efficient score test. This test determines if adding features to a model reduces the deviance residuals in the model. The test requires the models to be nested, so features for the models were selected as follows. The patient samples were divided and trained an Elastic Net model as previously described. Features included in the model were recorded. This was done three times, once using all features, once excluding the Riesz features, and once including only the Riesz features producing three following feature sets:

$$A = xa_1, xa_2, xa_3, \ldots, ra_1, ra_2, ra_3 \ldots$$

$$B = xb_1, xb_2, xb_3, \ldots$$

$$C = rc_1, rc_2, rc_3, \ldots$$

where A is the set of features used in the model including all features, B is the set of features used in the model excluding the Riesz features, C is the set of features used in the model including only the Riesz features, xn$i$ is the ith non-Riesz feature in feature set N, and rn$i$ is the ith Riesz feature in feature set N. Four binomial general logistic regression models (f) were made on the testing data such that:

$$M_{OS} = f(xb_1, xb_2, xb_3, \ldots)$$

$$M_{PR} = f(xb_1, xb_2, xb_3, \ldots\ldots ra_1, ra_2, ra_3 \ldots)$$

$$M_{OR} = f(rc_1, rc_2, rc_3, \ldots)$$

$$M_{PS} = f(xa_1, xa_2, xa_3, \ldots, rc_1, rc_2, rc_3, \ldots)$$

The test found that adding both the Riesz features to the standard features (Rao(M_{OS}, M_{PR}) $<<$ 0.001) and adding the standard features to the Riesz features (Rao(M_{OR}, M_{PS}) $<<$ 0.001) reduced the residuals in the model. Combining the Riesz and traditional quantitative histopathology features gives a higher accuracy and reduced residuals compared to either models with only one of these feature sets, indicating that the two feature sets are complementary rather than redundant, demonstrating the synergy between the feature sets.

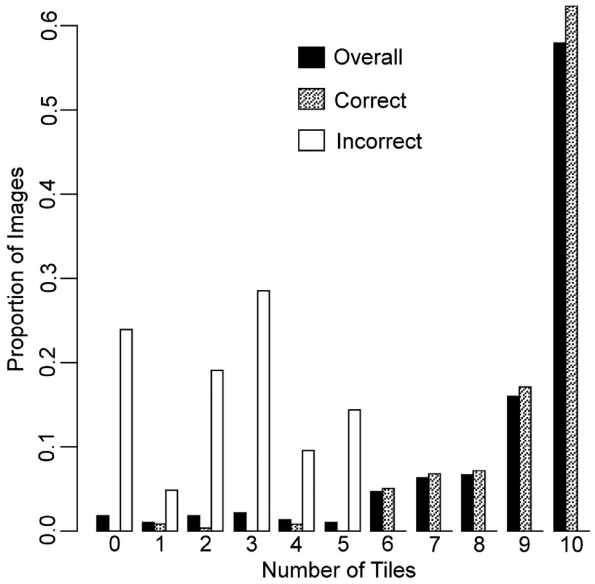

Fig. 8. Histogram of the number of tiles correctly classified for the WSI overall, the WSI that were correctly classified, and the WSI that were incorrectly classified. For example, in the data set as a whole, we see that ~58% of the images had all 10 tiles classified correctly, among the WSI that were correctly classified, ~62% had all 10 tiles classified correctly, and as would be expected, among the WSI that were incorrectly correctly classified, none had all 10 tiles classified correctly.

4.3. Analysis of tile classification

In addition to the classification of the WSI, we also tested how well the classification of the individual representative tiles and that of the WSI classification are correlated. In more than half the cases, all tiles were correctly classified (Fig. 8), indicating that the method is capable of finding signal in nearly all regions of the tumor. Even when tiles were incorrectly classified, in most cases, the pipeline was capable of making the correct decision (whole WSI classification) based on the correct classification of the other tiles. In two cases, the method made the correct decision for the WSI based on a strong correct decision for a single tile. On the other hand, when a WSI was incorrectly classified, fewer than half of the tiles were classified correctly, meaning that taking a simple majority would not have improved the classification of any images. This demonstrates the importance of the final weighted voting to detect strong signal in the data and make correct classifications.

4.4. Model stability with respect to parameter variation

4.4.1. Stability with respect to clustering parameters

The stability of the model with respect to the number of clusters and the number of PCA components used to select representative tiles was tested. All combinations of 5, 10, 15, and 20 clusters and 3, 4, 6, 8, 10, or 12 principal components have been examined. The accuracy of the classification remained stable and yielded accuracies between 95.5% and 100% (Fig. 8a). Additionally, when we examined the decision values used to assign the classification, we found that the mean decision values per number of clusters generated by the Elastic Net model was also similar across all combinations of parameters, with the sole exception of the 5 cluster models (Fig. 8b). This suggests that with only 5 clusters, the tissue diversity is not adequately sampled. In the instances of misclassified images, the magnitude of the decision value was well below the mean, indicating a lack of confidence in the assignment. This demonstrates that the model provided reproducible predictions with respect to the parameters used in tile selection.

4.4.2. Stability with respect to machine learning parameters

To test the model stability with respect to the parameters in the Elastic Net model, we evaluated the accuracy of the prediction when varying the logistic regression penalty value λ, and α, which controls the balance between the l1 and l2 penalty. λ was evaluated at 100 different values ranging from 6e − 5 to 0.6 with a logarithmically-spaced interval between adjacent test values and α for five values (0.0, 0.3, 0.5, 0.7, 1.0) (Fig. 9a). The accuracy of the predictions with respect to λ was stable over four orders of magnitude for all α values, only degrading at the extreme high end of the range when only a very small number of features were included in the model. The accuracy of predictions with respect to α remained stable at 100% when evaluated at the λ with the minimum cross validated mean square error. To evaluate the impact that α had on the model at a specific λ value, we looked at the distances of the final decision values from the threshold to estimate of the certainty of the predictions (Fig. 9b). The distance distribution was very similar for all α. Together our results indicate that the model can provide fairly robust classification predictions.

5. Discussion and conclusions

We have developed a novel, automated method for classifying digital pathology images of brain tumors that recognizes GBM versus LGG. Our method is unique in combining three components: (i) coarse profiling of regions within the image to sample tissue diversity (Section 3.2.1), (ii) identifying representative regions automatically in WSI (Section 3.2.2), and (iii) allowing all the representative regions to independently impact the final diagnosis (Section 3.3). We believe that all aspects of our proposed model are unique and contribute strongly to the high accuracy we achieved. The model showed low sensitivity to parameters related to the number of clusters, PCA components (Fig. 9), thus provided reproducible predictions with respect to the parameters used in tile selection. The parameters of the Elastic Net model also provided robust classification (Fig. 10). Even with substantial changes to those parameters the accuracy of the classification remained high at 95.5–100%. We believe that obtaining the appropriate and full representation of the image characteristics is one of the reasons for the high stability of our method.

LGG-versus-GBM classification is not a trivial task. Both LGG and GBM are diverse diseases known for their varied appearance in pathology samples. In many cases, a single small region of the slide with a few distinguishing features gives the pathologist the ability to diagnose a sample. For this reason, it is not surprising that in the 2014 MICCAI Pathology Classification challenge, all of our competitors manually selected a smaller region of the slide for classification (Bueno et al., 2014; Chang and Parvin, 2014; Xu et al. 2014a, 2014b). Our model obtained 100% classification accuracy and was able to exceed other competitors performance in addition to the advantage of being fully automatic (Table 1), indicating that its ability to select regions for analysis may outperform that of a human being. An extended dataset that contains 604 images was also analyzed producing a classification accuracy of 93.1%. One other group attempted classification of a larger cohort of TCGA images (51 GBM, 87 LGG, Mousavi et al., 2015). Their classification accuracy was much lower at 84.7%, suggesting that automated classification task is not trivial. While out method shows a decrease in accuracy our results in the larger dataset are still very good and potentially promising as an aid in clinical practice.

A variety of prior works have used computational feature extraction from images for classification, but many of these methods selected smaller areas of the tumor for processing (Alexe et al., 2009; Adiga et al., 2006; Aiad et al., 2009; Al-Kadi, 2010; Altunbay et al., 2010; Baak et al., 1981; Basavanhally et al., 2010, 2008; Brook et al., 2007; Doyle et al., 2012a, 2012b, 2008, 2007;

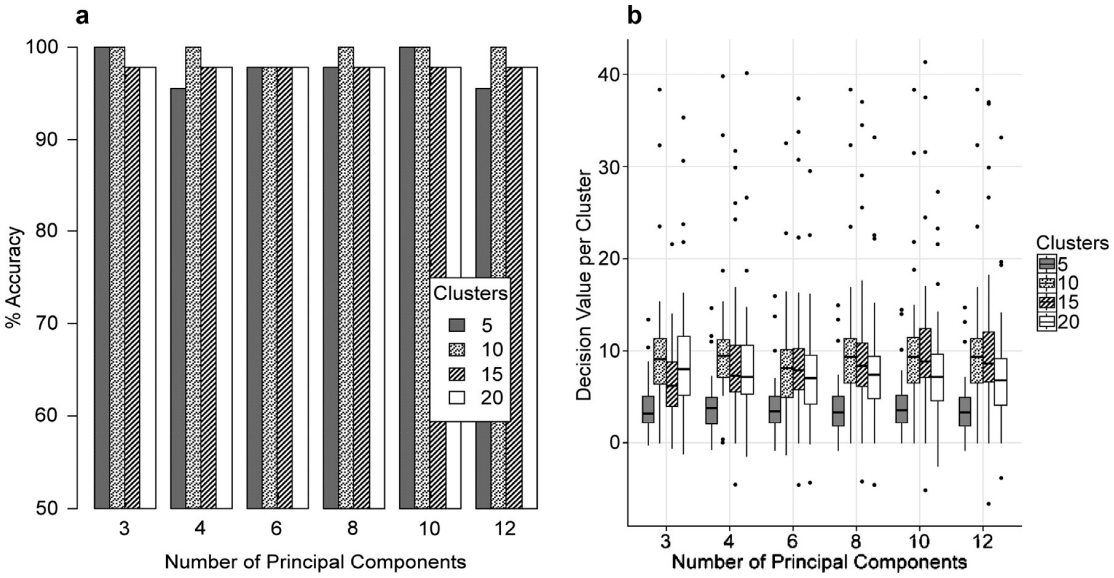

Fig. 9. Impact of tile selection parameters on model stability. (a) The accuracy of the models with differing numbers of clusters and principal components derived from the coarse features used for tile selection. (b) Mean decision value distribution for each tissue slice over all combinations of tile selection methods. Greater values indicate a stronger prediction, positive values indicate an accurate prediction and negative values indicate an inaccurate one. As both accuracy and decision values remain consistent over a wide range of parameter values, indicating that the model is robust to variation in the parameters guiding tile selection.

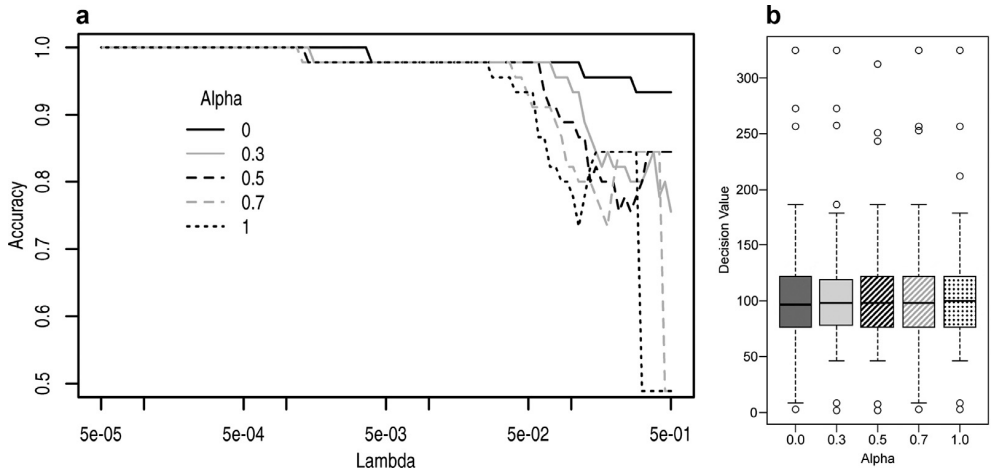

Fig. 10. The effect of the Elastic Net model parameters on classification predictions. (a) The impact of lambda and alpha on accuracy over five orders of magnitude for lambda and five different alpha values. (b) Distance distributions of the decision values from the classification threshold for alpha values (0.0, 0.3, 0.5, 0.7, 1.0). Greater values indicate a stronger prediction, positive values indicate an accurate prediction and negative values indicate an inaccurate prediction. There were no inaccurate predictions for any alpha value at lambda min. Accuracy of the model is stable over much of the parameter range, with lower accuracy only occurring where lambda forces few to no features to be included in the model.

Dundar et al., 2011, 2010; Esgiar et al., 2002; Farjam et al., 2007; Fatima et al., 2014; Glotsos et al., 2008; Gunduz et al., 2004; Gupta et al., 2001; Hall et al., 2008; Huang and Lai, 2010; Huang and Lee, 2009; Jafari-Khouzani and Soltanian-Zadeh, 2003; Monaco and Tomaszewski, 2008; Kong et al., 2007; Kwak et al., 2011; Land et al., 2008; Lessmann et al., 2007; Meng et al., 2010; Monaco et al., 2010; Naik et al., 2008, 2007; Ozolek et al., 2014; Petushi et al., 2006; Qureshi et al., 2008, 2007; Sboner et al., 2003; Schnorrenberg et al., 1997; Sertel et al., 2010; Sparks and Madabhushi, 2013; Sudbø et al., 2000; Tabesh and Teverovskiy, 2006; Tabesh et al., 2007, 2005; Tahir and Bouridane, 2006; Tasoulis et al., 2003; Teverovskiy et al., 2004; Thiran and Macq, 1996; Tsai et al., 2009; Xu et al. 2014a, 2014b; Yang et al., 2009). These approaches also limit the analysis to regions of the image already known to contain signal and therefore limit the potential of discovering new signal in previously unappreciated regions of the tissue sample. However, since our method is completely agnostic to any predefined tissue structures, it is free to discover signal that had been previously overlooked. Additionally, many previous methods require human intervention, introducing subjectivity, as a human must select a region to be characterized. This subjective step may introduce inter-reader variability, as seen in traditional pathology (Coons et al., 1997). Since our method is completely automatic, the potential for this type of variability is greatly reduced.

Hierarchical analysis of WSI where an analysis with a reduced computational load proceeds a more computationally intensive one has been effective in the past (Sertel et al., 2009). In our analysis, we leverage this by using the coarse features to select representative regions, allowing the model to include information from the entirety of the slide, which average 2 GB. This novel use of coarse

analysis eliminates the need for information about the regions of the slide containing tumor versus normal tissue, or the high computational complexity to analyze the whole image as in the methods discussed above. The proposed method overcomes these issues by being completely naive to pathological designations and can work without input from a pathologist or requiring automated tumor segmentation, while still requiring modest computational resources. The information provided by the coarse features can then be used to appropriately select regions for fine feature analysis, increasing the accuracy of the model (Section 4.2.1).

Extracting features separately from the hematoxylin and the eosin stains allows the features to better represent biological complexity. Many other methods collect features from either the original RGB image, a converted image to other non-biologically based color spaces (e.g., Lab or HSL), or from a grayscale version of that same image (Al-Kadi, 2010; Basavanhally et al., 2010; Dundar et al., 2011, 2010; Esgiar et al., 2002; Farjam et al., 2007; Glotsos et al., 2008; Huang and Lee, 2009; Jafari-Khouzani and Soltanian-Zadeh, 2003; Kong et al., 2009; Ozolek et al., 2014; Petushi et al., 2006; Qureshi et al., 2008; Ruiz et al., 2007; Schnorrenberg et al., 1997; Tabesh et al., 2005; Tabesh and Teverovskiy, 2006; Tahir and Bouridane, 2006; Thiran and Macq, 1996; Tuzel et al., 2007; Wang et al., 2010; Wetzel et al., 1999; Weyn et al., 1998; Xu et al. 2014a, 2014b). Since hematoxylin binds to nucleic acids and eosin binds to protein, unmixing the stains allows the feature extraction to directly probe the state of these important biological molecules, whereas features from the mixed image may either miss this signal or be unable to probe them independently.

The breadth of quantitative features we extract from the images played an additional role in the model's accuracy. Though many of our features are commonly used in pathology image analysis, to our knowledge, the Riesz features are novel in the context of pathology; indeed, these features comprised more than a third of the features used in every model (Fig. 7). The multi-scale nature of Riesz wavelets allows examination of the tissue at multiple spatial scales, from individual nuclei to multicellular structures. The directional components of the Riesz features can also be oriented to locally maximize the response of the first filter at the most granular scale, which has the desirable effect of normalizing all image directions among instances (Depeursinge et al., 2014). Since pathology has no universal orientation, this allows us to directly compare features from slide to slide without imposing an arbitrary directionality. The Riesz features are complementary to the more traditional pathology features, and when used together, the model has higher accuracy than using either feature set alone (Section 4.2.3).

Allowing all the representative tiles to independently impact the final decision adds flexibility to the model, making it potentially capable of dealing with diverse pathology image data. The initial clustering produces some clusters that are very relevant to the diagnosis and others that are less relevant. By allowing independent evaluation all the clusters, relevant clusters receive a large decision value and a heavier weight in the final vote, making a stronger impact on the final decision (Fig. 4). This enables a strong consensus for the entire tissue, even if the tumor represents only a small portion of the tissue; the impact of the relevant area is not diluted by the effect of the less important tissue surrounding it. While other methods (Kong et al., 2013,2009; Ruiz et al., 2007; Sertel et al., 2009) have used tiling, they make their final decision based on the proportion of tiles assigned to a particular disease type. These previous methods have been limited to identifying diseases defined by their preponderance of particular tissue types, as they are unable to evaluate diseases where local features are key in determining a diagnosis. They frequently must discard cases having a variety of tissue types, such as samples with large quantities of normal tissue. Our approach is not too dissimilar from that used by a pathologist in evaluating WSI, though our method is beneficial since in practice, it may be difficult for the human to systematically and consistently scrutinize every tissue region within each slide. We believe this contributed to the method's high accuracy (Section 4.1).

Our work has some limitations. One limitation is that even though our two-stage, coarse-to-fine profiling significantly reduces computational time, it is still slower than would be desired for any real-time application. This could be improved by implementing in a language such as C++ or running on more powerful computer hardware. Another limitation is that we have only evaluated this method in the context of whole slide images. Other types of pathology images, such as tissue microarrays, are much smaller and a pathologist has already selected the tumor region that they feel is representative. In this context, the coarse-to-fine profiling may not add as much value as in a whole slide image.

The method's ability to capture biological information in diverse datasets has strong potential in digital pathology, not only in disease diagnosis, also in other clinical applications, such as survival and analysis of drug treatment response. In addition, we believe our approach is likely generalizable and applicable to other diseases besides brain tumors.

Acknowledgment

This work was partially supported by the Swiss National Science Foundation (Grant PZ00P2_154891).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2015.12.002.

References

Adiga, U., Malladi, R., Fernandez-Gonzalez, R., de Solorzano, C.O., 2006. High-throughput analysis of multispectral images of breast cancer tissue. IEEE Trans. Image Process. 15, 2259–2268.

Aiad, H., Abdou, A., Bashandy, M., Said, A., Ezz-Elarab, S., Zahran, A., 2009. Computerized nuclear morphometry in the diagnosis of thyroid lesions with predominant follicular pattern. Ecancermedicalscience 3, 146.

Alexe, G., Monaco, J., Doyle, S., Basavanhally, A., Reddy, A., Seiler, M., Ganesan, S., Bhanot, G., Madabhushi, A., 2009. Towards improved cancer diagnosis and prognosis using analysis of gene expression data and computer aided imaging. Exp. Biol. Med. 234, 860–879.

Al-Kadi, O.S., 2010. Texture measures combination for improved meningioma classification of histopathological images. Pattern Recognit. 43, 2043–2053.

Altunbay, D., Cigir, C., Sokmensuer, C., Gunduz-Demir, C., 2010. Color graphs for automated cancer diagnosis and grading. IEEE Trans. Biomed. Eng., 57, pp. 665–674.

Arthur, D., Vassilvitskii, S., 2007. K-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA'07. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 1027–1035.

Baak, J.P.A., Kurver, P.H.J., Overdiep, S.H., Delemarre, J.F.M., Boon, M.E., Lindeman, J., Diegenbach, P.C., 1981. Quantitative, microscopical, computer-aided diagnosis of endometrial hyperplasia or carcinoma in individual patients. Histopathology 5, 689–695.

Basavanhally, A.N., Agner, S., Alexe, G., Bhanot, G., Ganesan, S., Madabhushi, A., 2008. Manifold learning with graph-based features for identifying extent of lymphocytic infiltration from high grade, HER2 breast cancer histology. Image Anal. Appl. Biol. (in Conjunction with MICCAI).

Basavanhally, A.N., Ganesan, S., Agner, S., Monaco, J.P., Feldman, M.D., Tomaszewski, J.E., Bhanot, G., Madabhushi, A., 2010. Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. IEEE Trans. Biomed. Eng. 57, 642–653.

Brook, A., El-Yaniv, R., Issler, E., Kimmel, R., Meir, R., Peleg, D., 2007. Breast cancer diagnosis from biopsy images using generic features and SVMs. In: Presented at the IEEE Transactions on Information Technology in Biomedicine.

Bueno, G., Fernández, M., Serrano, I., Vállez, N., Gutiérrez, J., Déniz, O., Salido, J., 2014. Statistical and frequential bag of features for automatic classification of gliomas. In: Proceedings of Digital Pathology Classification and Segmentation Challenge. Presented at the MICCAI. Boston, Massachusetts.

Cavenee, W.K., Louis, D.N., Ohgaki, H., Wiestler, 2007. WHO Classification of Tumours of the Central Nervous System. WHO Press, World Health Organization, Geneva, Switzerland.

Chang, H., Parvin, B., 2014. Segmentation of Necrosis based on Spatial Organization. In: Proceedings of Digital Pathology Classification and Segmentation Challenge. Presented at the MICCAI. Boston, Massachusetts.
Coons, S.W., Johnson, P.C., Scheithauer, B.W., Yates, A.J., Pearl, D.K., 1997. Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas. Cancer 79, 1381–1393.
Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D., Muller, H., 2014. Rotation covariant texture learning using steerable riesz wavelets. IEEE Trans. Image Process. 23, 898–908.
Divakar, R., 2009. Auto contrast. MATLAB Central File Exchange.
Doi, K., 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput. Med. Imaging Graph. 31, 198–211.
Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J., 2008. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In: Proceedings of 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008, pp. 496–499.
Doyle, S., Feldman, M.D., Shih, N., Tomaszewski, J., Madabhushi, A., 2012a. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. BMC Bioinf. 13, 282.
Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A., 2012b. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Trans. Biomed. Eng. 59, 1205–1218.
Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Feldman, M., Tomaszeweski, J., 2007. Automated grading of prostate cancer using architextural and textural image features,. In: Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2007, pp. 1284–1287.
Dundar, M.M., Badve, S., Bilgin, G., Raykar, V., Jain, R., Sertel, O., Gurcan, M.N., 2011. Computerized classification of intraductal breast lesions using histopathological Images. IEEE Trans. Biomed. Eng. 58, 1977–1984.
Dundar, M.M., Badve, S., Raykar, V.C., Jain, R.K., Sertel, O., Gurcan, M.N., 2010. A multiple instance learning approach toward optimal classification of pathology Slides. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR), pp. 2732–2735.
Esgiar, A.N., Naguib, R.N.G., Sharif, B.S., Bennett, M.K., Murray, A., 2002. Fractal analysis in the detection of colonic cancer images. IEEE Trans. Inf. Technol. Biomed. 6, 54–58.
Farjam, R., Soltanian-Zadeh, H., Jafari-Khouzani, K., Zoroofi, R.A., 2007. An image analysis approach for automatic malignancy determination of prostate pathological images. Cytometry B: Clin. Cytom. 72, 227–240.
Fatima, K., Arooj, A., Majeed, H., 2014. A new texture and shape based technique for improving meningioma classification. Microsc. Res. Tech 77, 862–873.
Friedman, J.H., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22.
Gil, J., Wu, H.-S., 2003. Applications of image analysis to anatomic pathology: realities and promises. Cancer Investig. 21, 950–959.
Glotsos, D., Kalatzis, I., Spyridonos, P., Kostopoulos, S., Daskalakis, A., Athanasiadis, E., Ravazoula, P., Nikiforidis, G., Cavouras, D., 2008. Improving accuracy in astrocytomas grading by integrating a robust least squares mapping driven support vector machine classifier into a two level grade classification scheme. Comput. Methods Prog. Biomed. 90, 251–261.
Gunduz, C., Yener, B., Gultekin, S.H., 2004. The cell graphs of cancer. Bioinformatics 20, i145–i151.
Gupta, N., Sarkar, C., Singh, R., Karak, A.K., 2001. Evaluation of diagnostic efficiency of computerized image analysis based quantitative nuclear parameters in papillary and follicular thyroid tumors using paraffin-embedded tissue sections. Pathol. Oncol. Res. 7, 46–55.
Gurcan, M.N., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B., 2009. Histopathological image analysis: a review. IEEE Rev. Biomed. Eng. 2, 147–171.
Gurcan, M.N., Pan, T., Shimada, H., Saltz, J., 2006. Image analysis for neuroblastoma classification: segmentation of cell nuclei. In: Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1, pp. 4844–4847.
Hall, B.H., Ianosi-Irimie, M., Javidian, P., Chen, W., Ganesan, S., Foran, D.J., 2008. Computer-assisted assessment of the human epidermal growth factor receptor 2 immunohistochemical assay in imaged histologic sections using a membrane isolation algorithm and quantitative analysis of positive controls. BMC Med. Imaging 8, 11.
Hamilton, P.W., Bankhead, P., Wang, Y., Hutchinson, R., Kieran, D., McArt, D.G., James, J., Salto-Tellez, M., 2014. Digital pathology and image analysis in tissue biomarker research. Methods: Adv. Bound. Mol. Cell. Pathol. 70, 59–73.
Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEE Trans. Syst. Man Cybern. SMC-3, 610–621.
Hou, H.S., Andrews, H., 1978. Cubic splines for image interpolation and digital filtering. IEEE Trans. Acoust. Speech Signal Process. 26, 508–517.
Huang, P.-W., Lai, Y.-H., 2010. Effective segmentation and classification for HCC biopsy images. Pattern Recognit 43, 1550–1563.
Huang, P.-W., Lee, C.-H., 2009. Automatic classification for pathological prostate images based on fractal analysis. IEEE Trans. Med. Imaging 28, 1037–1050.
Jafari-Khouzani, K., Soltanian-Zadeh, H., 2003. Multiwavelet grading of pathological images of prostate. IEEE Trans. Biomed. Eng 50, 697–704.
Kamentsky, L., Jones, T.R., Fraser, A., Bray, M.-A., Logan, D.J., Madden, K.L., Ljosa, V., Rueden, C., Eliceiri, K.W., Carpenter, A.E., 2011. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinfor. Oxf. Engl. 27, 1179–1180.
Khasraw, M., Lassman, A.B., 2010. Advances in the treatment of malignant gliomas. Curr. Oncol. Rep. 12, 26–33. doi:10.1007/s11912-009-0077-4.
Kong, J., Cooper, L.A.D., Wang, F., Gao, J., Teodoro, G., Scarpace, L., Mikkelsen, T., Schniederjan, M.J., Moreno, C.S., Saltz, J.H., Brat, D.J., 2013. Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates. PLoS One 8, e81049.
Kong, J., Sertel, O., Shimada, H., Boyer, K.L., Saltz, J.H., Gurcan, M.N., 2009. Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation. Pattern Recognit. 42, 1080–1092.
Kong, J., Shimada, H., Boyer, K., Saltz, J., Gurcan, M., 2007. Image analysis for automated assessment of grade of neuroblastic differentiation. In: Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2007, pp. 61–64.
Kwak, J.T., Hewitt, S.M., Sinha, S., Bhargava, R., 2011. Multimodal microscopy for automated histologic analysis of prostate cancer. BMC Cancer 11, 62.
Land, W.H., McKee, D.W., Zhukov, T., Song, D., Qian, Wei, 2008. A kernelised fuzzy-support vector machine CAD system for the diagnosis of lung cancer from tissue images. Int. J. Funct. Inform. Pers. Med. 1, 26–52.
Lessmann, B., Nattkemper, T.W., Hans, V.H., Degenhard, A., 2007. A method for linking computed image features to histological semantics in neuropathology. J. Biomed. Inform. 40, 631–641 (Intelligent Data Analysis in Biomedicine).
Meng, T., Lin, L., Shyu, M.-L., Chen, S.-C., 2010. Histology image classification using supervised classification and multimodal fusion. In: Proceedings of 2010 IEEE International Symposium on Multimedia (ISM), pp. 145–152.
Meyer, F., 1979. Iterative image transformations for an automatic screening of cervical smears. J. Histochem. Cytochem. 27, 128–135.
Monaco, J.P., Tomaszewski, J.E., Feldman, M.D., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A., 2008. Detection of prostate cancer from whole-mount histology images using Markov random fields. In: Workshop on Microscopic Image Analysis with Applications in Biology. Presented at the MICCAI, New York, NY.
Monaco, J.P., Tomaszewski, J.E., Feldman, M.D., Hagemann, I., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A., 2010. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Med. Image Anal. 14, 617–629.
Mousavi, H.S., Monga, V., Rao, G., Rao, A.U.K., 2015. Automated discrimination of lower and higher grade gliomas based on histopathological image analysis. J. Pathol. Inform. 6, 15.
Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., Tomaszewski, J., 2008. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008, pp. 284–287.
Naik, S., Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A., 2007. Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. In: Presented at the MIAAB Workshop.
Onega, T., Weaver, D., Geller, B., Oster, N., Tosteson, A.N.A., Carney, P.A., Nelson, H., Allison, K.H., O'Malley, F.P., Schnitt, S.J., Elmore, J.G., 2014. Digitized whole slides for breast pathology interpretation: current practices and perceptions. J. Digit. Imaging 27, 642–648.
Ostrom, Q.T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., Wolinsky, Y., Kruchko, C., Barnholtz-Sloan, J., 2014. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. Neuro-Oncol. 16, iv1–iv63.
Ozolek, J.A., Tosun, A.B., Wang, W., Chen, C., Kolouri, S., Basu, S., Huang, H., Rohde, G.K., 2014. Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning. Med. Image Anal. 18, 772–780.
Pencina, M.J., D' Agostino, R.B., D' Agostino, R.B., Vasan, R.S., 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat. Med. 27, 157–172.
Petushi, S., Garcia, F.U., Haber, M.M., Katsinis, C., Tozeren, A., 2006. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. BMC Med. Imaging 6, 14.
Preusser, M., de Ribaupierre, S., Wöhrer, A., Erridge, S.C., Hegi, M., Weller, M., Stupp, R., 2011. Current concepts and management of glioblastoma. Ann. Neurol. 70, 9–21. doi:10.1002/ana.22425.
Qureshi, H., Rajpoot, N., Nattkemper, T., Hans, V., 2007. Comparative analysis of discriminant wavelet packet features and raw image features for classification of meningioma subtypes. In: Presented at the Medical Image Understanding and Analysis, Aberystwyth, Wales.
Qureshi, H., Sertel, O., Rajpoot, N., Wilson, R., Gurcan, M., 2008. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 196–204.
Ruifrok, A.C., Johnston, D.A., 2001. Quantification of histochemical staining by color deconvolution. Anal. Quant. Cytol. Histol.: Int. Acad. Cytol. Am. Soc. Cytol. 23, 291–299.
Ruiz, A., Sertel, O., Ujaldon, M., Catalyurek, U., Saltz, J., Gurcan, M., 2007. Pathological image analysis using the GPU: stroma classification for neuroblastoma. In: Proceedings of IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2007., pp. 78–88.
Sboner, A., Eccher, C., Blanzieri, E., Bauer, P., Cristofolini, M., Zumiani, G., Forti, S., 2003. A multiple classifier system for early melanoma diagnosis. Artif. Intell. Med. 27, 29–44.
Schnorrenberg, F., Pattichis, C.S., Kyriacou, K.C., Schizas, C.N., 1997. Computer-aided detection of breast cancer nuclei. IEEE Trans. Inf. Technol. Biomed. 1, 128–140.

Sertel, O., Kong, J., Shimada, H., Catalyurek, U.V., Saltz, J.H., Gurcan, M.N., 2009. Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development. Pattern Recognit. 42, 1093–1103.
Sertel, O., Lozanski, G., Shana'ah, A., Gurcan, M.N., 2010. Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood based cell segmentation. IEEE Trans. Biomed. Eng. 57, 2613–2616.
Sparks, R., Madabhushi, A., 2013. Explicit shape descriptors: novel morphologic features for histopathology classification. Med. Image Anal. 17, 997–1009.
Sudbø, J., Marcelpoil, R., Reith, A., 2000. New algorithms based on the Voronoi diagram applied in a pilot study on normal mucosa and carcinomas. Anal. Cell. Pathol. 21, 71–86.
Tabesh, A., Kumar, V.P., Pang, H.-Y., Verbel, D., Kotsianti, A., Teverovskiy, M., Saidi, O., 2005. Automated prostate cancer diagnosis and Gleason grading of tissue microarrays. pp. 58–70.
Tabesh, A., Teverovskiy, M., 2006. Tumor classification in histological images of prostate using color texture. In: Proceedings of the Fortieth Asilomar Conference on Signals, Systems and Computers, ACSSC'06, pp. 841–845.
Tabesh, A., Teverovskiy, M., Pang, H.-Y., Kumar, V.P., Verbel, D., Kotsianti, A., Saidi, O., 2007. Multifeature prostate cancer diagnosis and gleason grading of histological images. IEEE Trans. Med. Imaging 26, 1366–1378.
Tahir, M.A., Bouridane, A., 2006. Novel round-robin tabu search algorithm for prostate cancer classification and diagnosis using multispectral imagery. IEEE Trans. Inf. Technol. Biomed. 10, 782–793.
Tasoulis, D.K., Spyridonos, P., Pavlidis, N.G., Cavouras, D., Ravazoula, P., Nikiforidis, G., Vrahatis, M.N., 2003. Urinary bladder tumor grade diagnosis using on-line trained neural networks. In: Palade, V., Howlett, R.J., Jain, L. (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 199–206.
Teverovskiy, M., Kumar, V., Ma, J., Kotsianti, A., Verbel, D., Tabesh, A., Pang, H.-Y., Vengrenyuk, Y., Fogarasi, S., Saidi, O., 2004. Improved prediction of prostate cancer recurrence based on an automated tissue image analysis system. In: Proceedings of IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004, 1, pp. 257–260.
Thiran, J.P., Macq, B., 1996. Morphological feature extraction for the classification of digital images of cancerous tissues. IEEE Trans. Biomed. Eng. 43, 1011–1020.
Tsai, H.-W., Tsai, H.-H., Kuo, F.-Y., Chang, K.-C., 2009. Computerized analyses of morphology and proliferative activity differentiate hepatoblastoma from paediatric hepatocellular carcinoma. Histopathology 54, 328–336.
Tuzel, O., Yang, L., Meer, P., Foran, D.J., 2007. Classification of hematologic malignancies using texton signatures. Pattern Anal. Appl. 10, 277–290.
Wang, W., Ozolek, J.A., Rohde, G.K., 2010. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images. Cytometry A 77, 485–494.
Wetzel, A.W., Crowley, R., Kim, S., Dawson, R., Zheng, L., Joo, Y.M., Yagi, Y., Gilbertson, J., Gadd, C., Deerfield, D.W., Becich, M.J., 1999. Evaluation of prostate tumor grades by content-based image retrieval. In: 27th AIPR Workshop: Advances in Computer-Assisted Recognition. Presented at the AIPR, Washington, DC, pp. 244–252. doi:10.1117/12.339826.
Weyn, B., van de Wouwer, G., van Daele, A., Scheunders, P., van Dyck, D., van Marck, E., Jacob, W., 1998. Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. Cytometry 33, 32–40.
Xu, Y., Jia, Z., Zhang, F., Ai, Y., Lai, M., Chang, E.I.-C., 2014. Deep convolutional activation features for large brain tumor histopathology image classification, in: digital pathology classification and segmentation challenge. In: Presented at the MICCAI. Boston, Massachusetts.
Xu, Y., Zhu, J.-Y., Chang, E.I.-C., Lai, M., Tu, Z., 2014. Weakly supervised histopathology cancer image segmentation and classification. Med. Image Anal. 18, 591–604.
Yang, L., Chen, W., Meer, P., Salaru, G., Goodell, L.A., Berstis, V., Foran, D.J., 2009. Virtual microscopy and grid-enabled decision support for large-scale analysis of imaged pathology specimens. IEEE Trans. Inf. Technol. Biomed. 13, 636–644.
Zhao, D., Daut, D.G., 1991. Shape recognition using morphological transformations. In: Proceedings of 1991 International Conference on Acoustics, Speech, and Signal Processing, ICASSP-91, 4, p. 2565—2568.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the Elastic Net. J. R. Stat. Soc. Ser. B 67, 301–320.