HHS Public Access

Author manuscript

IFMBE Proc. Author manuscript; available in PMC 2016 August 14.

Published in final edited form as:

IFMBE Proc. 2014 ; 42: 280–283. doi:10.1007/978-3-319-03005-0_71.

A High-Resolution Tile-Based Approach for Classifying Biological Regions in Whole-Slide Histopathological Images

R.A. Hoffman[1], **S. Kothari**[2], **J.H. Phan**[1], and **M.D. Wang**[1,2,3]

[1]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA

[2]Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA

[3]Winship Cancer Institute, Parker H. Petit Institute of Bioengineering and Biosciences, Institute of People and Technology, Georgia Institute of Technology and Emory University, Atlanta, GA

Abstract

Computational analysis of histopathological whole slide images (WSIs) has emerged as a potential means for improving cancer diagnosis and prognosis. However, an open issue relating to the automated processing of WSIs is the identification of biological regions such as tumor, stroma, and necrotic tissue on the slide. We develop a method for classifying WSI portions (512x512-pixel tiles) into biological regions by (1) extracting a set of 461 image features from each WSI tile, (2) optimizing tile-level prediction models using nested cross-validation on a small (600 tile) manually annotated tile-level training set, and (3) validating the models against a much larger ($1.7x10^6$ tile) data set for which ground truth was available on the whole-slide level. We calculated the predicted prevalence of each tissue region and compared this prevalence to the ground truth prevalence for each image in an independent validation set. Results show significant correlation between the predicted (using automated system) and reported biological region prevalences with $p < 0.001$ for eight of nine cases considered.

I. Introduction

Histopathology plays a vital support role in oncology. Whole-slide histopathological images (WSIs) are digital images of sectioned and stained tissue samples that are scanned and digitally recorded at high resolutions. In traditional use, WSIs have enabled (1) more streamlined record-keeping, (2) training of laboratory technicians, and (3) offsite consultation for diagnoses and prognoses [1, 2]. WSIs have emerged as a growing area of interest in image processing and imaging informatics. WSIs have been shown to contain significant diagnostic and prognostic data, which can be extracted in reproducible and quantitative ways to support fast, accurate decisions [3, 4]. The emergence of large, multi-modal cancer data repositories, such as The Cancer Genome Atlas (TCGA), has also stimulated research in automated informatics methods for WSIs [6].

Correspondance: May D. Wang, PhD, Professor, Department of Biomedical Engineering, Georgia Institute of Technology, maywang@bme.gatech.edu, 313 Ferst Drive, Atlanta, GA 30332.

Significant computational, experimental, and biological challenges exist when attempting to use WSIs for diagnostic or prognostic applications. The size and resolution of the images pose computational challenges, with many images having dimensions on the order of gigapixels. This has limited many previous studies to relatively small data sets, e.g., on the order of dozens of WSIs. Experimentally, WSIs are subject to many of the same image artifacts as traditional slides, including tissue folds and non-tissue markings in the image [5]. WSIs are very heterogeneous in terms of the types of cells and tissues captured. Thus, a major biological challenge is region-of-interest (ROI) selection and classification, which is essential for accurate diagnosis. There is a pressing need for automated WSI informatics methods for artifact correction, feature extraction, ROI selection, and image classification [4].

This paper focuses on a specific biological challenge of ROI classification. Three distinct tissue types can appear in any cancer WSI: stroma, tumor tissue, and necrotic tissue. Tumor and necrotic tissue each contain distinct diagnostic and prognostic features [7], and it is important to exclude connective stroma from consideration with these regions. Here, we seek to develop and validate methods for the image-based classification of these tissue regions. We examine two types of carcinoma (ovarian serous cystad-enocarcinoma [OV] and renal clear cell carcinoma [KIRC]); optimize disease-specific and pooled classification models; and illustrate informative imaging markers for classifying biological regions in WSIs. Finally, we validate our approach by comparing prediction results to manually annotated ground truth data as well as to annotation data reported by TCGA.

II. Methods

A. DATA

WSIs for both OV and KIRC were taken from the TCGA database. The high-resolution WSIs are first broken into 512x512-pixel tiles. This serves both to subdivide the computationally intensive work of extracting features and to limit those features to describing small neighborhoods of cells. Figure 1(b) shows the scope of a single WSI tile. We selected 300 high-resolution tiles from each cancer type for the ground truth data to be used in training. For each data set, we selected one hundred representative tiles for each tissue class considered: stroma, tumor tissue, and necrotic tissue.

B. QUALITY CONTROL

Before extracting features from the tiled WSIs, quality control is performed to isolate the tissue from the slide background and remove artifacts such as pen markings and tissue folds. Slide background and pen marks are defined by thresholding in the HSV color space [9]. Tissue folds are identified using a supervised, connectivity-based soft thresholding method called ConnSoftT [5].

Finally, tiles within the tissue region of interest are defined as those tiles with less than 10% tissue fold and less than 80% background and pen markings combined. We use these tiles for the subsequent feature-extraction step.

The full set of 461 features used in this analysis has previously been shown to capture useful information about the image. These features fall into one of nine classes: global color, global texture, eosinophilic texture, eosinophilic object shape, basophilic texture, basophilic object shape, no-stain object shape, nuclear shape, or nuclear topology [8]. The tile is first color-normalized against a fixed set of reference images to mitigate batch effects due to variation in staining and equipment. It is then divided into regions based on the presence of hematoxylin or eosin staining. Each of these regions is then processed separately, yielding its own shape and texture characteristics. Finally, the nuclear region mask is also processed for topological features describing the distribution of the nuclei in the tissue section [10].

C. FEATURE SELECTION AND CLASSIFICATION

We use the minimum redundancy, maximum relevance (mRMR) method to select features from among the 461 image features [11]. We apply the mRMR method in two ways: mutual information difference (MID) and mutual information quotient (MIQ). We use the linear support vector machine (SVM) [13] and various forms of discriminant analysis (DA) for classification. We consider four different DA classifiers: linear, diagonal linear, quadratic, and diagonal quadratic.

D. MODEL OPTIMIZATION AND VALIDATION

In order to compare the various models of interest and estimate classifier accuracy, we used a nested cross-validation scheme. The outer CV loop (10 iterations, 3 fold) randomly splits the annotated training set into a test set and a training set. Each training set was fit to a model using every permutation of each parameter. SVM cost parameters considered were 2^{-5}, 2^{-4}, … 2^{10}. Feature sizes tested were 5, 10, … 50 for both the MID and MIQ mRMR methods. Each of these models were themselves analyzed using a 3-fold, 10 iteration internal CV loop. This nested CV structure is necessary to produce unbiased estimates of model performance. By subdividing the training set before optimizing the model, a portion of unused training tiles remains to test the best model. (Fig. 4)

The best model was defined as the simplest model whose mean internal CV accuracy was within one standard deviation of the accuracy of the most accurate model. Smaller feature sets were preferred and small costs were preferred, in that order. In DA classification, linear classifiers were considered simpler than quadratic, and diagonal methods were considered simpler than non-diagonal.

As part of the metadata available for histopathology slides, TCGA provides expert annotation of the proportion of each tissue sample that falls into various biological regions. The OV data set consisted of 1087 slides with complete metadata. After applying tile-level quality control, 861,430 high-resolution tiles remained. For KIRC, 922 slides containing 802,833 tissue tiles were used. The “WSIs” block shown in Figure 4 refers to this data set.

Tiles for each slide were processed with the trained classifier and estimates of the prevalence of each biological region were generated for that slide. We compared these results to TCGA reported values and computed the Pearson correlation coefficient.

III. Results and Discussion

A. INFORMATIVE FEATURES

Using the mRMR method and number of features found to be most robust by the inner cross-validation loop, a Fisher's exact test was performed to determine if any class of features was significantly over-represented in the optimized feature spaces. Results are shown in Table 1.

Values marked with a (*) are significant with $p < 0.05$. Color, texture, and shape features were each found to be significantly over-represented in at least one classifier. This validates the use of a large, varied set of image features over the use of texture-only or color-only feature spaces, as are common in prior literature [14, 15].

B. CLASSIFICATION CROSS-VALIDATION

Table 2 shows the outer CV accuracy of SVM classification for each disease and classifier model considered. The most likely sources of bias in these results are selection bias in the training data set and the possibility of multiple tiles from one WSI spanning the training and test sets. While Table 2 may not be an unbiased estimate of external validation accuracy, the relative accuracies show which models consistently outperform the others.

Two-tailed t-tests were used to compare models across disease cases and classifier models. SVM models were found to perform slightly better for KIRC and OV, however this difference was not significant. Pooled DA significantly outperformed the corresponding SVM. Pooled models significantly underperformed compared to disease-specific models in all cases except OV-DA.

C. EXTERNAL VALIDATION

SVM based predictions of the relative size of biological regions in TCGA WSIs had higher correlations than DA counterparts in 8 of 9 cases, as shown in Table 3. Correlation was higher for pooled samples than disease-specific classifiers in 5 of 6 cases, and statistically significant for all classes. Neither the OV nor KIRC classifiers yielded significant correlation in all cases shown in Table 3. Figure 5 shows larger and more intense regions of correlation between predicted and reported regions sizes in Pooled results.

The selected SVM classification model for each disease case was trained using the entire 300 or 600 tile ground truth data set, and then used to predict the class of the corresponding large TCGA data set.

In all disease cases, when validating on ground-truth tiles, pooled models underperform as compared to disease-specific models. This difference was significant in three of four cases but the performance of pooled models was still high. Moreover, pooled models tended to perform better on larger datasets in external validation. Hence, pooled models are more generalized. This indicates the possibility of developing a generalized tissue-classification model for different cancer, subject to further investigation.

IV. Conclusion

In this paper, we developed and validated an automated system for classifying biological regions in WSIs. The proposed system uses a tile-based approach, where each tile is represented by a comprehensive set of features and classified into one of the regions using pre-optimized classifiers. The overall trend of the external validation shows significant correlation between the classifier-predicted values and expert-reported values suggesting successful classification of biological regions. In future work, these methods could be expanded to include information about a tile's neighborhood to arrive at more accurate and better predictions of biological regions. These methods may also be applicable to other cancer types and biological regions.

Acknowledgments

The authors thank the NCI, the NHGRI, and the patients who contributed to the TCGA data set. We thank Michael Glover for his work in assembling the tile-level training set. This research has been supported by grants from NIH (U54CA119338, 1RC2CA148265, and R01CA163256).

References

1. Isaacs MM, Lennerz JKJ, Yates SS, et al. Implementation of whole slide imaging in surgical pathology: A value added approach. J Pathol Inform. 2011; 2:39–39. [PubMed: 21969920]
2. Pantanowitz L, Evans A, Pfeifer J, et al. Review of the current state of whole slide imaging in pathology. J Pathol Inform. 2011; 2:36. [PubMed: 21886892]
3. Kothari S, Phan JH, Young AN, Wang MD. Histological Image Feature Mining Reveals Emergent Diagnostic Properties for Renal Cancer. BIBM. 2011:422–425.
4. Kothari S, Phan JH, Stokes TH, Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. Journal of the American Medical Informatics Association. 2013
5. Kothari S, et al. Eliminating Tissue-Fold Artifacts in Histopathological Whole-Slide Images for Improved Image-based Prediction in Cancer Grading. 2013 In Press.
6. The Cancer Genome Atlas. NCI and NHGRI.
7. Costa J, Wesley RA, Glatstein E, Rosenberg SA. The grading of soft tissue sarcomas. Results of a clinicohistopathologic correlation in a series of 163 cases. Cancer. 1984; 53:530–541. [PubMed: 6692258]
8. Kothari S, Phan JH, Osunkoya AO, Wang MD. Biological interpretation of morphological patterns in histopathological whole-slide images. 2012
9. Smith AR, Smith AR. Color gamut transform pairs. ACM SIGGRAPH Computer Graphics. 1978; 12:12–19.
10. Kothari S, Chaudry Q, Wang MD. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. Proc IEEE Int Symp Biomed Imaging. 2009:795–798.
11. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005; 27:1226–1238. [PubMed: 16119262]
12. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol. 2005; 3:185–205. [PubMed: 15852500]
13. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans on Intelligent Sys and Tech (TIST). 2011; 2:27.
14. Karaçali B, Tözeren A. Automated detection of regions of interest for tissue microarray experiments: an image texture analysis. BMC Med Imaging. 2007; 7:2. [PubMed: 17349041]

15. Lesack K, Naugler C. Performance of a simple chromatin-rich segmentation algorithm in quantifying basal cell carcinoma from histology images. BMC Research Notes. 2012; 5:35. [PubMed: 22251818]

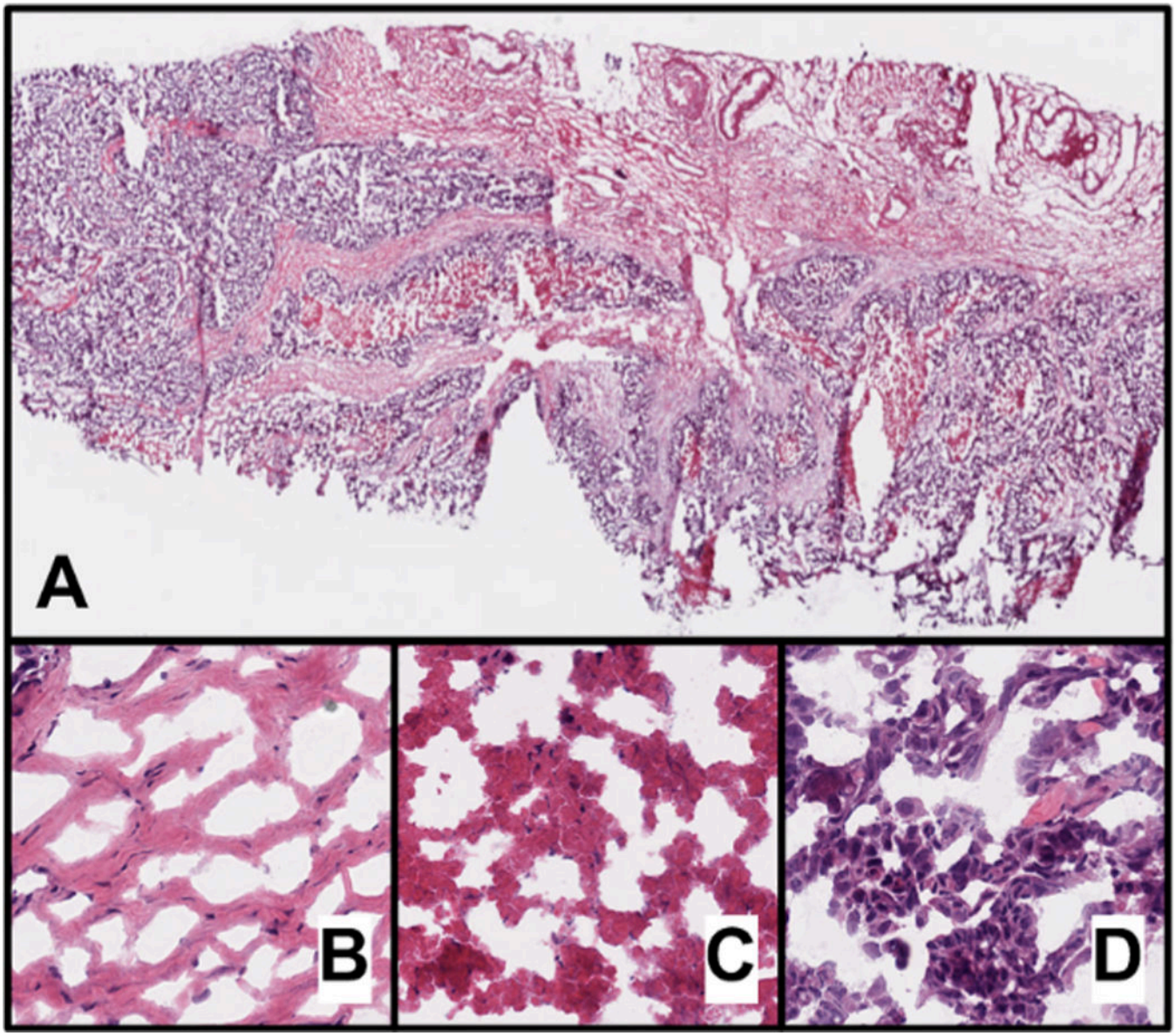

Fig. 1.

(A) A typical whole-slide image. (B) A 512x512 high-resolution WSI tile showing stroma. (C) A typical necrotic tissue tile. (D) A tile containing typical tumor tissue.

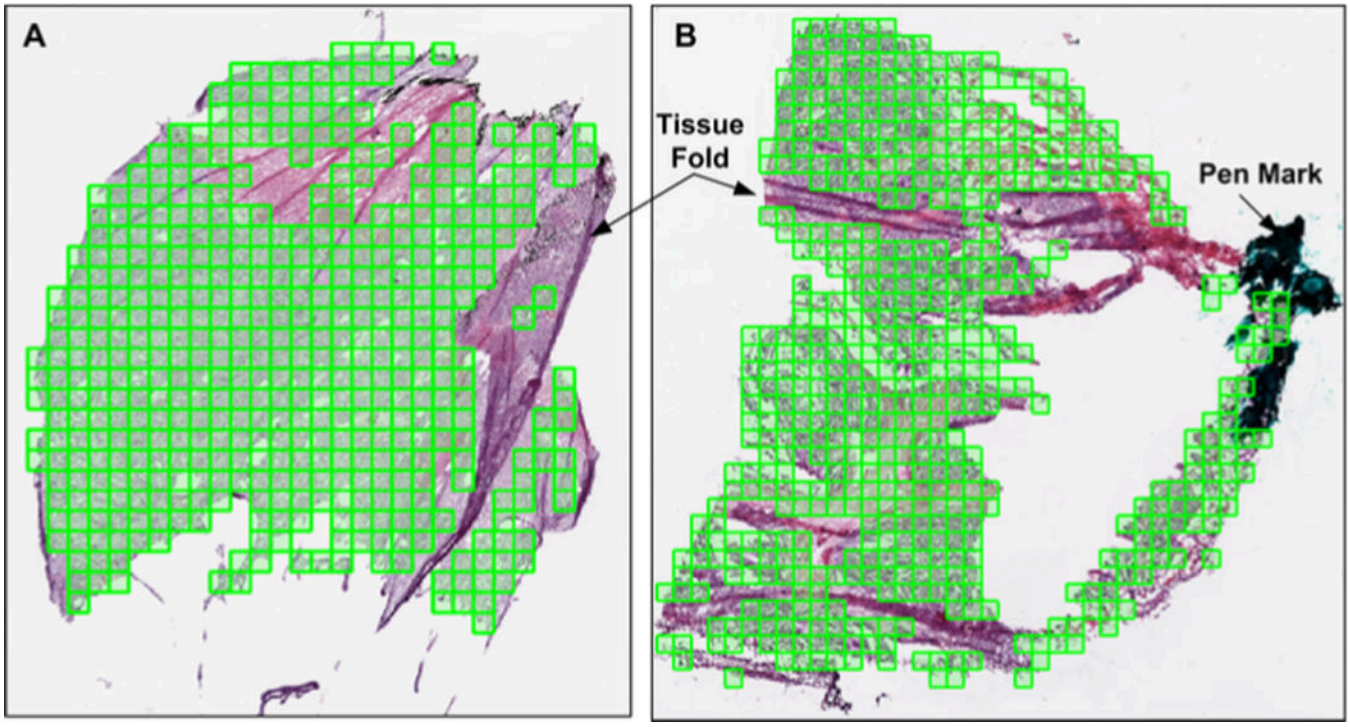

Fig. 2.
Two typical WSIs with the ROI tiles after quality control outlined in green.

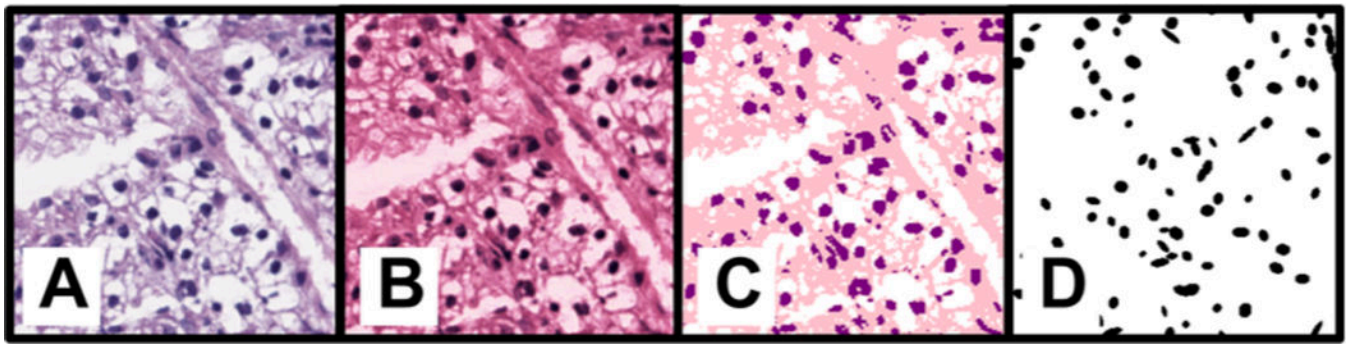

Fig. 3.

(A) A portion of a high-resolution WSI tile. (B) The tile is color-normalized. (C) Thresholding is used to segment stain regions (light pink = eosinophilic, dark purple = basophilic). (D) A nuclear mask of the tile.

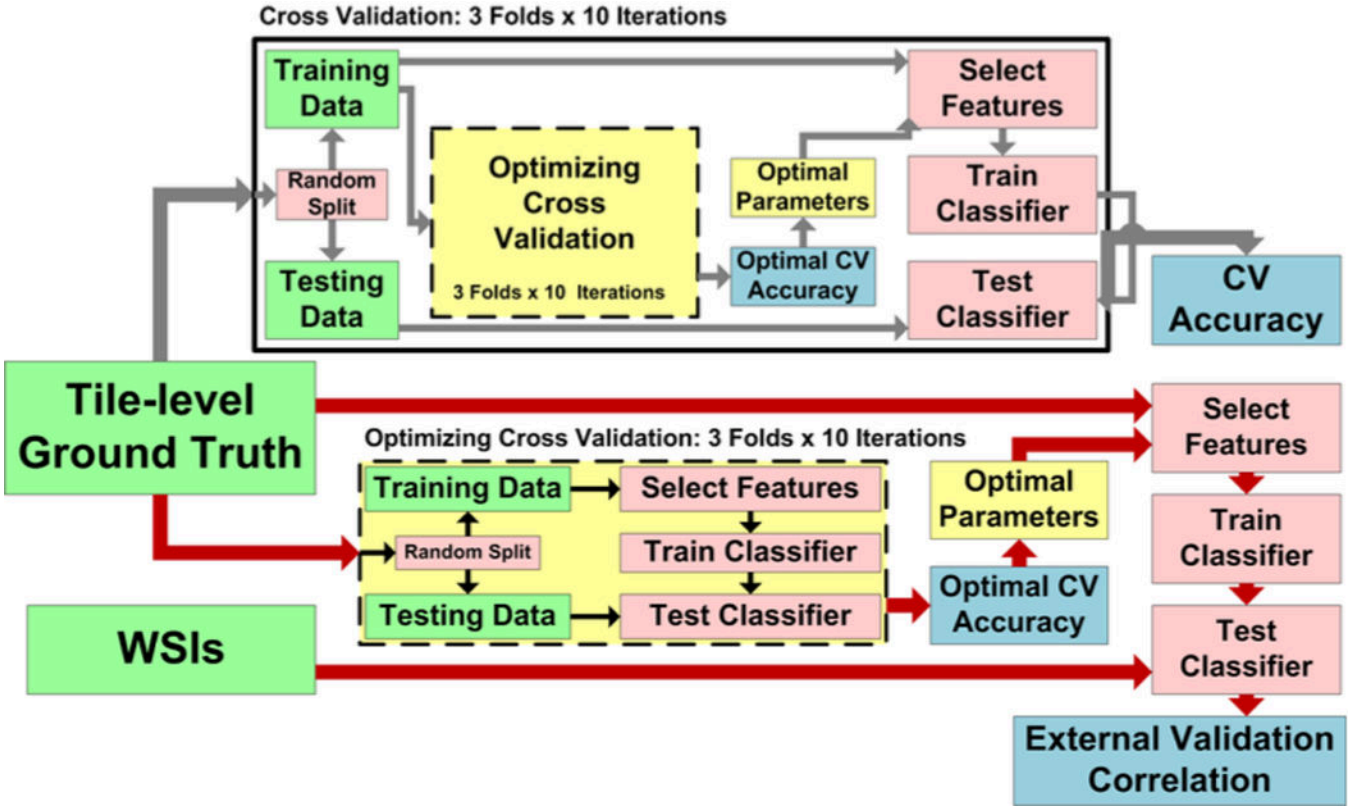

Fig. 4.

The nested cross-validation flow chart. The “tile-level ground truth” data set consisted of 600 expert-annotated high-resolution tile. The “WSIs” data set consisted of 1.7X 106 tiles. These tiles were grouped into their constituent WSIs and analyzed together for normalized region sizes, as slide-level metadata is what is available from TCGA.

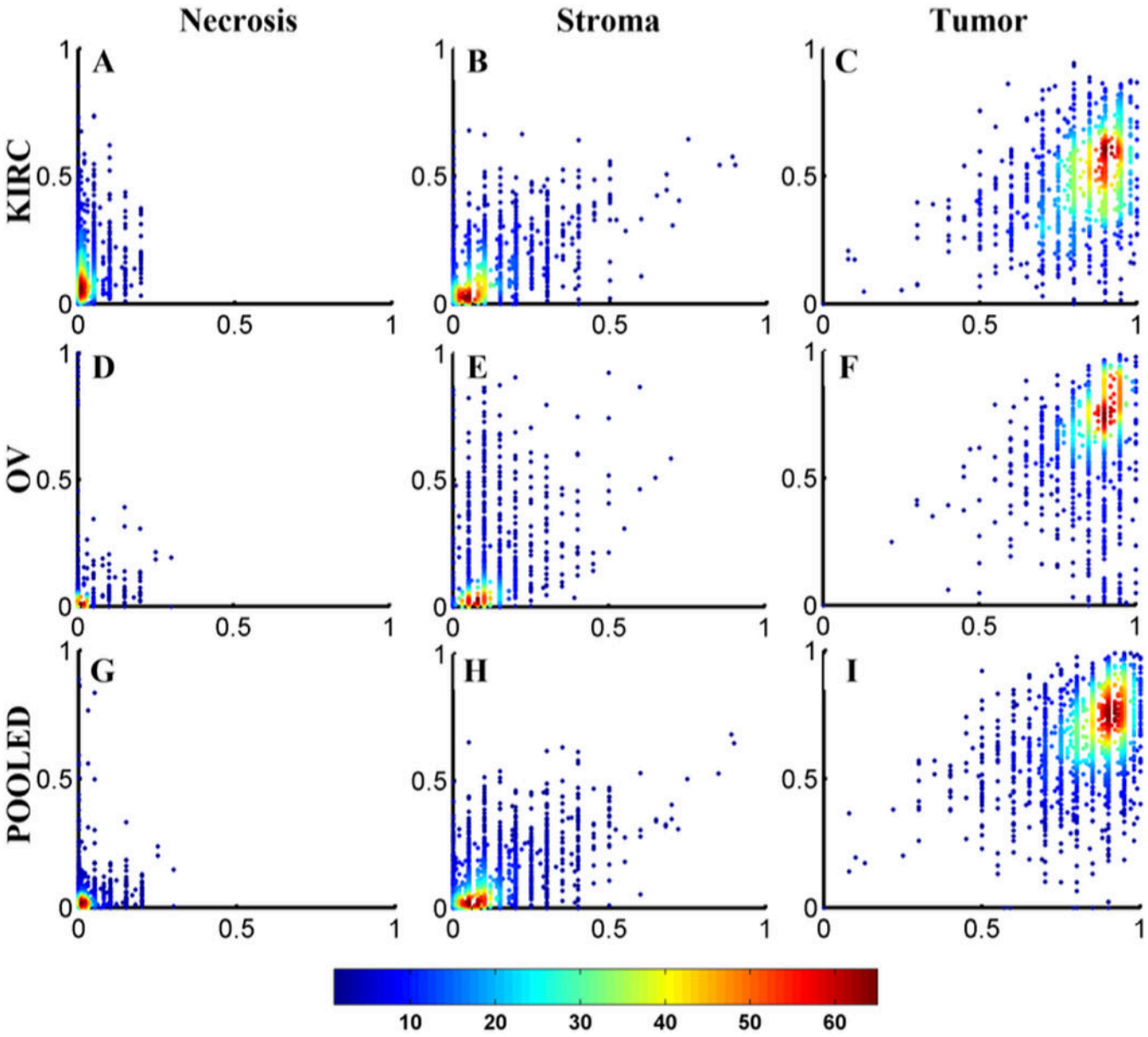

Fig. 5.

External Validation. Each scatterplot shows the classifier predicted biological region size versus the TCGA reported biological region size.

Table 1

Contribution of each feature class to each classifier model.

Feature Class	OV SVM	OV DA	KIRC SVM	KIRC DA	Pooled SVM	Pooled DA
Color	0.2	0.22*	0.23*	0.23*	0.21	0.22*
Global texture	0.22	0.37*	0.32	0.34	0.27	0.33
Eosinophilic shape	0.08	0.08	0.04	0.05	0.09	0.1
Eosinophilic texture	0.06	0.04	0.09*	0.12*	0.03	0.07
No-stain shape	0.08	0.04	0.04	0	0.07	0.04
Basophilic shape	0.20*	0.16	0.18*	0.17*	0.18*	0.16
Basophilic texture	0.04	0	0.04	0.03	0.04	0.02
Nuclear shape	0.08	0.04	0.07	0.05	0.08	0.05
Nuclear topology	0.04	0.04	0	0	0.02	0

Table 2

Cross-validation accuracy of each classification model

	SVM	DA
OV	97.6±1.59 *	97.5±1.46
KIRC	98.7±0.76 †	98.4±1.22 ‡
Pooled	95.9±1.14 *†○	96.8±1.29 ‡○

Models marked with a symbol are significantly different from models marked with the same symbol by a two-tailed Student's t-test with $p<0.05$.

Values are given as mean ± one standard deviation.

Table 3

Correlation coefficients for slide-level classification models

	SVM			DA		
Class	**Necrosis**	**Stroma**	**Tumor**	**Necrosis**	**Stroma**	**Tumor**
OV	0.031	0.425 *	0.371 *	−0.002	0.439 *	0.253 *
KIRC	0.130 *	0.299 *	0.351 *	0.075	0.346 *	0.276 *
Pooled	0.098 *	0.491 *	0.479 *	0.109 *	0.469 *	0.389 *

Values marked with a (*) have significant correlation with $p<0.001$