Deep Learning Assessment of Tumor Proliferation in Breast Cancer Histological Images

Manan Shah  
Department of Computer Science  
Stanford University  
Stanford, USA  
manan@cs.stanford.edu

Dayong Wang  
Department of Pathology  
Beth Israel Deaconess Medical Center  
Boston, USA  
dwang5@bidmc.harvard.edu

Christopher Rubadue  
Department of Pathology  
Beth Israel Deaconess Medical Center  
Boston, USA  
crubadue@bidmc.harvard.edu

David Suster  
Department of Pathology  
Beth Israel Deaconess Medical Center  
Boston, USA  
dsuster@bidmc.harvard.edu

Andrew Beck  
Department of Pathology  
Beth Israel Deaconess Medical Center  
Boston, USA  
abeck2@bidmc.harvard.edu

Abstract—Current analysis of tumor proliferation, the most salient breast cancer prognostic biomarker, is limited to subjective mitosis counting by pathologists in localized regions of tissue images. This study presents the first data-driven integrative approach to characterize the severity of tumor growth and spread on a categorical and molecular level, utilizing multiple biologically salient deep learning classifiers to develop a comprehensive prognostic model. Our approach achieves pathologist-level performance on three-class categorical tumor severity prediction. It additionally pioneers prediction of molecular expression data from a tissue image, obtaining a Spearman’s rank correlation coefficient of 0.60 with \textit{ex vivo} mean calculated RNA expression. Furthermore, our framework is applied to identify over two hundred unprecedented biomarkers critical to the accurate assessment of tumor proliferation, validating our proposed integrative pipeline as the first to holistically and objectively analyze histopathological images.

I. INTRODUCTION

Breast cancer is the most common cancer in women worldwide, with over 1.7 million new cases diagnosed in 2012 [1]. Cancer assessment is influenced by environmental and clinical factors, but it is universally accepted that tumor proliferation speed (tumor growth) is an important biomarker representative of progression rate and patient outcomes [2], [3]. The assessment of this biomarker influences patient treatment plans, allowing for more aggressive tumors to be treated with the corresponding therapy [3].

In a clinical setting, tumor proliferation is manually assessed by pathologists under a regime of counting mitotic figures in hematoxylin & eosin (H&E) stained histological slide preparations that are examined under a high powered microscope. Although ubiquitous, this categorical score has been reported to suffer from reproducibility problems that reflect the underlying subjectivity of the process [2]. In addition, the simple methodology of pathologist mitosis counting and subsequent thresholding fails to account for a wide range of pathological features. The molecular score, which is a more objective proliferation measure defined as the mean expression of eleven prognostic RNA sequences, requires extensive \textit{ex vivo} molecular tests, relegating current practices to an inadequate pathologist diagnosis [3], [4]. To address this problem, we propose a data-driven approach to characterize tumor growth severity and spread on a categorical and molecular level. In particular, the main contributions of this paper are as follows.

- We developed a comprehensive deep learning pipeline constructing and unifying models across several associated tasks of tumor localization, mitotic figure identification, and high-level feature extraction to classify categorical tumor grades and predict RNA proliferation speed scores from histological whole slide images (WSIs).
- We achieved pathologist-level performance on three-class categorical tumor severity prediction.
- We identified salient biomarkers related to tumor diagnosis to serve as a basis for future studies. The data-driven integrative approach presented here is generalizable and will be useful to analyze other cancerous tumors.

II. METHODS

A. Dataset Description and Preprocessing

Three datasets were used in this study to train primary and auxiliary models [5]. Our primary evaluation dataset consisted of 500 WSIs with magnification levels from 10× to 40× that are annotated with a tumor score based on mitosis counting by pathologists and a molecular (RNA) proliferation score [5]. We additionally use 73 2mm² magnified images annotated with mitotic figures and 148 WSIs partially annotated with high cellularity regions for supplemental training.

As our WSIs originate from three international pathology centers, each exhibited different staining methods. To ensure that such variations in the color and intensity of H&E staining would not hamper the effectiveness of subsequent quantitative image analysis, we employed Bejnordi et al’s WSI Color Standardization (WSICS) procedure [6]. Normalizing each
WSI and images from auxiliary datasets ensured that our subsequent models exhibited stain invariance between tissue preparation methods. Tissue regions were next extracted from each stain standardized input WSI using Otsu’s method on a HSV representation of the original RGB image [7], [8]. We subsequently removed small artifacts and expanded remaining regions via binary dilation to obtain a holistic tissue mask.

B. Network Construction

Our pipeline (Fig. 1) next performed three tasks: metastatic tumor localization, mitotic figure identification, and WSI general feature extraction. We used a magnification level of 10× for the tumor localization process to identify high-level patterns, and we conducted mitosis identification and feature extraction on the 40× level for detailed analysis.

1) Metastatic Tumor Localization: Having extracted and normalized tissue regions from each input WSI, it is important to identify candidate regions for mitotic activity indicative of tumor proliferation. Such regions are high-cellularity areas with proliferative activity represented at tissue abnormality edges. Fig. 2 depicts the four employed network architectures, the first three of which are recognized as state-of-the-art convolutional neural networks (CNNs) for object recognition [9]–[11]. Each CNN generated output tumor probability heatmaps with a “sliding-window” approach, classifying overlapping tissue patches for tumor probability and assigning the resulting value to the center pixel of each patch. The fourth network, which we named LocNet, reduced the number of free parameters and operated on a fully convolutional paradigm [12]. LocNet allowed for arbitrarily sized inputs (we use 1k × 1k patches) and produced downsampled corresponding heatmaps for each patch. We upsampled and stitched these probability heatmaps over each WSI to rapidly generate a comprehensive heatmap.

Network training was framed as an active learning problem [13] as each annotated image contained a non-exhaustive list of tumor regions. We, therefore, separated the process into two components, with the first stage defining annotated patches as positive and identifying a random sample of remaining tissue patches as negative. Heatmaps were subsequently produced using each model and additional regions predicted as positive with over 95% confidence were appended to the initial positive training set. All models were retrained with the refined data; subsequent second stage results better eliminated regions misclassified in the first stage. The trained models were used both to identify tumor regions within which to perform mitotic figure identification and to provide informative features regarding tumor shape, density, area, extent, and location.

2) Mitotic Figure Identification: We next constructed mitotic figure detectors to identify biologically salient features within tumor areas. Due to aberrant tumor chromosomal makeup, mitotic figure appearances may vary from typical hyperchromatic objects with an absence of a clear nuclear membrane and hairy protrusions around edges.

The employed networks are depicted in Fig. 2. DenseNet, requiring the most parameters, constructed repeated connections between network layers to develop a robust approach. Although the GoogLeNet architecture and its modified fully convolutional counterpart performed reasonably well, the additional complexity encoded within the network architecture excessively distilled the already small mitotic figures. To remedy this issue, we applied the LocNet model and developed a specialized architecture called MitosNet. LocNet and MitosNet exhibited more fine-tuned layers, capturing the variance between mitoses without degrading effective inference.

3) WSI General Feature Extraction: In addition to developing methods for the identification of anatomical structures in tumor severity analysis, we created end-to-end networks that predict the output categorical severity grade of the whole slide image from individual patches. These developed networks (Fig. 2) utilize outputs of tumor and mitosis networks to extract detailed computational features. Patches are extracted from original WSIs and input to the first network (with static weights)
which computes coarse features corresponding to either mitosis identification or tumor localization. Convolutions from the first network’s feature volume are next mapped to the input of a second network (with dynamic weights). The second network performs categorical predictions and extracts WSI features. These 1,024 features, and 3 predicted class probabilities, were incorporated in the final predictive model.

C. Tumor-Specific Feature Extraction

We applied our mitosis detection and tumor localization methods to identify biologically salient features in WSIs on both a patch-based and a whole slide level. Specifically, we preferentially selected fifty patches from the fringes of localized tumor regions with the largest area. Each patch, a 1k×1k tissue sample at magnification level 40×, was input to our mitosis detectors which produced heatmaps of corresponding size identifying mitotic figure probability in the input image. We additionally represented each WSI with a comprehensive heatmap depicting mitotic figures in all tumor regions. Both individual patches and the WSI heatmap were used to compute biological and data-driven mitosis features.

Biological Features. From each selected patch, we extracted fifty morphometric and intensity based features to characterize biologically salient structural mitosis components. These features describe compositional and formational patterns that pathologists might observe. In addition, we characterized the distribution of mitoses throughout the entire WSI with sixty architectural features. Particularly, we analyzed the sparsity of mitosis distribution and second-order attributes including kurtosis, entropy, and skewness, providing a high-level interpretation of proliferative activity.

Data-Driven Features. Within each magnified patch, we additionally computed abstract deep learning-based features that represent attributes from learned filters. We segmented a 63 × 63 tissue patch around each identified mitosis for input to our mitosis detection networks. Each mitotic figure was subsequently characterized by 4,096 attributes to describe mitosis-specific structural minutiae. We conducted post-processing k-means clustering on all individual mitosis feature vectors (of length 4,096) from every WSI patch in a 200-dimensional vector space. Each vector was associated with a cluster label ∈ [1, 200] identifying its most similar sub-space. Finally, each WSI was distilled into a 200-bin histogram with frequencies corresponding to the cluster labels of each mitotic region.

III. EXPERIMENTS AND RESULTS

A. Performance Evaluation

Categorical Tumor Severity. A receiver operating characteristic (ROC) curve detailing the ratio of true positives and false positives at varying thresholds is depicted in Fig. 3. Each class was predicted in a one vs all manner with mean values determined in five-fold cross-validation. The resulting micro-average AUROC of 0.78 validates our overall f-measure of 0.62, establishing the model’s powerful discriminative potential among the three classes. Our method additionally achieved an accuracy of 0.72 (95% CI: 0.67, 0.76) when compared to pathologist severity gradation, indicating marginal deviation of our predictions from the inter-pathologist agreement of 0.79 (95% CI: 0.70, 0.85) [2].

Molecular RNA Expression. Our best-performing regression model achieved a mean squared error of 0.119. Fig. 4 depicts the correlation between our regression predictions and the calculated mean expression of eleven prognostic RNA strands. Our model, the first ever to predict gene expression data from histopathological image slides, achieves a Pearson’s correlation coefficient value r = 0.58 and a Spearman’s rank correlation of 0.58.
These results indicate the ability of our pipeline to provide meaningful insights into causative biological pathways that contribute to the onset and extent of tumor growth. With additional clinical verification, we anticipate that this new integrative approach relying solely on tissue samples will enable the rapid classification and specific biomarker identification of other cancerous diseases.

### References

[1] Jiemin Ma and Ahmedin Jemal, “Breast cancer statistics,” in *Breast Cancer Metastasis and Drug Resistance*, pp. 1–18. Springer, 2013.

[2] Mitko Veta et al., “Mitosis counting in breast cancer: Object-level interobserver agreement and comparison to an automatic method,” *PloS one*, vol. 11, no. 8, pp. e0161286, 2016.

[3] Maggie CU Cheang et al., “Immunohistochemical detection using the new rabbit monoclonal antibody sp1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody 1d5 in predicting survival,” *Journal of clinical oncology*, vol. 24, no. 36, pp. 5637–5644, 2006.

[4] Torsten O Nielsen et al., “A comparison of pam50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor–positive breast cancer,” *Clinical Cancer Research*, vol. 16, no. 21, pp. 5222–5232, 2010.

[5] Medical Image Analysis Group Eindhoven (IMAG/e), “Reading whole-slide images,” 2016.

[6] Babak Ehteshami Bejnordi et al., “Stain specific standardization of whole-slide histopathological images,” *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 404–415, 2016.

[7] Shamik Sural, Gang Qian, and Sakti Pramanik, “Segmentation and histogram generation using the hsv color space for image retrieval,” in *Image Processing, 2002. Proceedings. 2002 International Conference on*. IEEE, 2002, vol. 2, pp. II–589.

[8] Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[9] Christian Szegedy et al., “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” *arXiv preprint arXiv:1512.03385*, 2015.

[11] Karen Simonyan, Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[12] Jonathan Long et al., “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[13] David A Cohn et al., “Active learning with statistical models,” *Journal of artificial intelligence research*, 1996.

[14] Tao Huang, Zhiqiang Liu, and Kilian Q Weinberger, “Densely connected convolutional networks,” *arXiv preprint arXiv:1608.06993*, 2016.

[15] Samad Muhammadnejad et al., “Correlation of microvessel density with nuclear pleomorphism, mitotic count and vascular invasion in breast and prostate cancers at preclinical and clinical levels,” *Asian Pacific Journal of Cancer Prevention*, vol. 14, no. 1, pp. 63–68, 2013.

[16] R Rappaport, “Cytokinesis: the effect of initial distance between mitotic apparatus and surface on the rate of subsequent cleavage furrow progress,” *Journal of Experimental Zoology*, vol. 221, no. 3, pp. 399–403, 1982.