Utilizing Automated Breast Cancer Detection to Identify Spatial Distributions of Tumor Infiltrating Lymphocytes in Invasive Breast Cancer

Han Le[1,+], Rajarsi Gupta[2,3,*,+], Le Hou[1], Shahira Abousamra[1], Danielle Fassler[3], Tahsin Kurc[2], Dimitris Samaras[1], Rebecca Batiste[3], Tianhao Zhao[3], Alison L. Van Dyke[4], Ashish Sharma[6], Erich Bremer[2], Jonas S. Almeida[5], and Joel Saltz[2]

[1]Department of Computer Science, Stony Brook University, Stony Brook, NY 11974, USA
[2]Department of Biomedical Informatics, Stony Brook Medicine, Stony Brook, NY 11794, USA
[3]Department of Pathology, Stony Brook University Hospital, Stony Brook, NY 11794, USA
[4]Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
[5]National Cancer Institute Division of Cancer Epidemiology & Genetics, Bethesda, MD 20892, USA
[6]Department of Biomedical Informatics, Emory University, Atlanta, GA 30322, USA
[*]Rajarsi.Gupta@stonybrookmedicine.edu
[+]these authors contributed equally to this work

ABSTRACT

Quantitative assessment of Tumor-TIL spatial relationships is increasingly important in both basic science and clinical aspects of breast cancer research. We have developed and evaluated convolutional neural network (CNN) analysis pipelines to generate combined maps of cancer regions and tumor infiltrating lymphocytes (TILs) in routine diagnostic breast cancer whole slide tissue images (WSIs). We produce interactive whole slide maps that provide 1) insight about the structural patterns and spatial distribution of lymphocytic infiltrates and 2) facilitate improved quantification of TILs. We evaluated both tumor and TIL analyses using three CNN networks - Resnet-34, VGG16 and Inception v4, and demonstrated that the results compared favorably to those obtained by what believe are the best published methods. We have produced open-source tools and generated a public dataset consisting of tumor/TIL maps for 1,015 TCGA breast cancer images. We also present a customized web-based interface that enables easy visualization and interactive exploration of high-resolution combined Tumor-TIL maps for 1,015 TCGA invasive breast cancer cases that can be downloaded for further downstream analyses.

Introduction

Among women worldwide, invasive breast cancer is the most common cancer and the second most common cause of cancer-related death[1], despite decreasing mortality rates in recent years due to early diagnosis and current therapeutic options that significantly prolong survival. Invasive breast cancers are a heterogeneous category of disease phenotypes[2,3] that are histologically classified into subtypes based on growth patterns; the expression of estrogen (ER), progesterone (PR), human epidermal growth factor receptor 2 (HER2); and the Ki-67 proliferation index.

The role of tumor infiltrating lymphocytes (TILs) in invasive breast cancer has become increasingly important as a biomarker that can predict clinical outcomes, as well as to predict treatment response in the neoadjuvant and adjuvant setting[4–11]. TILs are a readily available biomarker and their evaluation is likely to expand with the emergence of immunotherapy. Elevated concentrations of TILs in HER2-positive[12] and triple-negative (ER-/PR-/HER2-)[13] breast cancers are associated with prolonged overall and disease-free survival; whereas elevated concentrations of TILs in luminal HER2-negative breast cancer have been associated with poor overall survival[4]. TILs can also serve as a predictive biomarker since a significant part of the cytotoxic effects of systemic chemotherapy and radiation-therapy are actually mediated by activating the immune system to kill cancer cells instead of directly targeting the tumor cells[14]. Targeted therapies against HER2 and vascular endothelial growth factor (VEGF) are mediated by both antibody-dependent and complement-mediated cytotoxicity in cancer cells through lymphocytes and other immune cells in the tumor microenvironment[15]. Recent studies suggest potential for synergistic effects between targeted and immune therapies in multiple disease sites[16,17].

Current practice routinely includes manual assessments of hematoxylin and eosin (H&E) stained tissue sections by surgical pathologists to identify and classify invasive breast cancer. Such diagnostic evaluation provides insight about clinical

management, treatment selection, survival, and recurrence. Since H&E tissue sections are readily available, there is a sustainable opportunity to provide potentially actionable data about TILs without the need for additional tissue samples (e.g., immunohistochemical (IHC) testing). H&E also permit the interpretation of the lymphocyte infiltrate within and proximal to the tumor in the context of histology to provide insight about the spatial relationships between tumor regions and TILs. The published guidelines for the histologic assessment of TILs in invasive breast cancer[18–20] require pathologists to select the region of tumor and to delineate stromal areas in order to assess the percentage of TILs (%TILs) in ***stromal*** regions as a continuous variable from 0-100% within the boundaries of the entire tumor that is used to classify the lymphocyte infiltrate as low, intermediate, and high, respectively.

However, this evaluation is intrinsically qualitative and often subject to inter-observer variability, so previous work as articulated these concerns[21] in an attempt to clearly state the need for automated methods to evaluate %TILs in H&E tissue sections of breast cancer. Computationally calculating %TILs intrinsically provides spatial information about how TILs are distributed in whole slide images (WSIs), where it is likely that the distinction between intratumoral and stromal TIL infiltrates is important. While there have been some relatively small studies examining intratumoral and stromal TILs[22], the predictive power of the spatial distribution of TILs within tumor and tumor-associated stroma needs to be better elucidated. Automated evaluation of TILs in H&E WSIs fundamentally requires tumor segmentation linked with the detection of lymphocyte infiltrates. Automation of H&E tumor-TIL analyses will make it possible to carry out large-scale correlative studies that quantitatively characterize TIL distributions in well-characterized clinical populations. Computer analysis of high-resolution images of whole slide tissue specimens can enable a data driven and quantitative characterization of TIL patterns.

With the recent success of deep learning[23] and the availability of public datasets[24–27], several research groups have proposed deep learning based algorithms to detect or segment cancer/tumor regions in breast cancer WSIs[28–31]. However, the overall performance of previous methods was constrained due to the insufficient capacity of customized convolutional neural networks[28,29] or the limited amount of training data[30,31]. Hence, we proposed to use standard state-of-the-art deep learning models with the usage of a large-scale dataset to detect invasive breast cancer regions in WSIs. Our experimental evaluation on a public test set showed better results than the models in the previous works. Our study also combines tumor detection with lymphocyte detection to identify TILs in a large number of WSIs which are publicly accessible for the breast cancer research community.

We utilized a previously published deep learning method to detect lymphocytes in WSIs[32]. In this report, we combine a breast tumor segmentation algorithm with lymphocyte detection to leverage the training data used in Saltz et al., 2018[32]. We developed a deep learning-based method to automate breast cancer detection at intermediate- to high-resolution in order to generate detailed probability-based heat maps of the tumor bed. Our approach achieves an F1-score of 0.82, a positive predictive value (PPV) of 79%, and negative predictive value (NPV) of 98% in terms of pixel-by-pixel evaluation in an unseen and independent test dataset consisting of 195 TCGA WSIs. We then combined the breast cancer detection with results from a published methods that utilizes deep learning to generate high-resolution TIL maps. The combined results represent regions of Tumor with intra- and peritumoral TILs in 1,015 WSIs from the publicaly available The Cancer Genome Atlas (TCGA) repository. We expect that the availability of high-resolution spatial Tumor-TIL maps will allow quantitative estimation and characterization of the relationship between tumor cells and TILs. The ability to quantify and visualize the spatial relationships between tumor and TILs can be a very practical and useful way to further elucidate intriguing observations in previous studies. It will also further our collective understanding of the biological behavior of invasive breast cancers within the context of cancer-immune interactions in the tumor microenvironment.

Results

Datasets

We used high-resolution WSIs from the Virtual Tissue Repository (VTR) Pilot Program operated through the Surveillance, Epidemiology, and End Results (SEER: https://seer.cancer.gov/) cancer registry system and from The Cancer Genome Atlas (TCGA: https://portal.gdc.cancer.gov/) to train and evaluate the deep learning models and to generate cancer region maps. The training, validation, and test datasets consisted of image patches extracted from 102, 7, and 89 SEER WSIs, respectively. All of the images were scanned at 400x magnification and manually segmented by an expert pathologist into cancer and non-cancer regions using a web-based application[33]. Additionally, we evaluated the models using manual tumor region annotations in 195 TCGA WSIs (referred as here to T_{tcga}); this dataset had been generated by Cruz-Roa et al.[29] and was publicly accessible. The details of the training, validation and test datasets for tumor region segmentation are presented in Table 1. The trained models were applied to 1,015 diagnostic WSIs from TCGA invasive breast cancer cases. The same set of WSIs was also analyzed using the TIL classification models trained with data generated by Saltz et al.[32]. These data consisted of 86,154 and 653 image patches for training and validation, respectively. We created a test dataset of 327 patches extracted from TCGA invasive breast cancer WSIs to evaluate the trained models. The details of the training, validation, and test datasets for TIL classification are presented in Table 2.

Source	Purpose	ID	WSIs (N)	Patches (N)	Positive (N)	Negative (N)
SEER	Training	D_tr	102	333,604	99,889	233,715
	Validation	D_val	7	10,224	4,953	5,271
	Testing	T_seer	89	-	-	-
TCGA	Testing	T_tcga	195	-	-	-

Table 1. Data statisitcs of the Breast Cancer dataset.

Source	Purpose	Patches (N)	Positive (N)	Negative (N)
TCGA	Training	86,154	21,773	64,381
	Validation	653	295	357
	Testing	327	174	153

Table 2. Data statistics of the Lymphocytes dataset provided in Saltz et al.[32]

Experiments

We evaluated multiple state-of-the-art deep learning networks, namely VGG16[34], Resnet34[35], and Inception-v4[36], for the tasks of detecting and segmenting breast cancer regions and classifying TILs in WSIs. We used accuracy, F1-score, and area under the ROC (Receiver Operating Characteristic) curve (AUC) as performance evaluation metrics. Accuracy is the ratio of correct predictions to the total number of data elements in the ground truth dataset. Because a dataset is not always balanced between classes, we used the F1-score that considers both precision and recall to compute a score. Mathematically, F1-score = 2*(precision*recall)/(precision + recall). Lastly, we used AUC to evaluate the prediction capability of our model at different threshold settings. AUC shows the relationship between True Positive Rate (TPR) and False Positive Rate (FPR) of the model. It is a widely used metric to assess model performance for binary classification tasks.

The best models (Tables 3 and 4) were applied to WSIs from 1,015 TCGA invasive breast cancers to generate what we call *prediction probability maps* for cancer regions and TILs. A prediction probability map is constructed by partitioning a WSI into N×N image patches. The image patches were analyzed by the trained model and were assigned a label probability between 0.0 and 1.0. For cancer region segmentation, the label of a patch was either *positive* (i.e., the patch predicted to be within or intersect a cancer region) or *negative* (i.e., the patch is predicted to be outside the cancer regions in the WSI). For TIL classification, the label of a patch was either TIL-positive (i.e., the patch was predicted to contain lymphocytes) or TIL-negative. The model assigned a value between 0.0 and 1.0 to each patch to indicate the predicted probability of the corresponding label. We developed a Web-based application to visualize and interact with the prediction probability maps as heatmaps (please see "Methods" section).

Evaluation of Cancer Detection Models

We trained three cancer detection and segmentation models, C-VGG16, C-Resnet34 and C-Incepv4, by using VGG16, Resnet34 and Inception-v4, respectively. We compared the performances of the models to each other as well as to a model that was trained by a learning network called *ConvNet*, which was developed by Cruz-Roa et al.[28,29]. ConvNet was originally trained on a different training dataset, HUP and UHCMC/CWRU[28]. In order to use our training datasets, we implemented ConvNet using Pytorch[37] by precisely following the network description in the original paper[28]. We call our implementation *ConvNet-ours*.

We computed an average F1-score across all the test images by varying the threshold value from 0.0 to 1.0 in steps of 0.01. At each threshold value, prediction probability maps were computed for the 195 test images by the model under evaluation and the patch labels were assigned by thresholding the prediction probability maps by the threshold value. The thresholded label maps and the ground-truth masks[29] were then used to compute average F1-score, positive predictive value (PPV), negative predictive value (NPV), true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), and false negative rate (FNR). Table 3 shows the performance comparison between our models, the original ConvNet model[28,29], and our implementation of the ConvNet model (ConvNet-ours). We report the performance of ConvNet-ours both with and without applying our post-processing step (please see "Methods" section), because no post-processing step was applied to the original ConvNet model[28,29]. Table 3 shows that the post-processing step improves the average F1-score from 0.75 to 0.77 and PPV from 0.69 to 0.73. Moreover, ConvNet-ours slightly outperforms the original ConvNet model in all metrics. Figure 1 shows prediction probability maps produced for a set of representative WSIs in T_{tcga} using the C-Resnet34 model. The shades of red in the map images indicate how confident the model is that a patch is *positive (i.e., patch in cancer region)* or *negative (patch outside cancer region)*. A visual inspection of the maps and the respective WSIs showed that the model was able to detect and

Method	F1-score	PPV	NPV	TPR	TNR	FPR	FNR
ConvNet[29]	0.76 ± 0.20	0.72 ± 0.22	0.97 ± 0.05	0.87 ± 0.16	0.92 ± 0.08	0.08 ± 0.08	0.13 ± 0.16
ConvNet-ours	0.75 ± 0.18	0.69 ± 0.22	0.96 ± 0.09	0.87 ± 0.18	0.91 ± 0.09	0.09 ± 0.07	0.12 ± 0.16
ConvNet-ours*	0.77 ± 0.21	0.73 ± 0.23	0.97 ± 0.09	0.87 ± 0.23	0.92 ± 0.09	0.08 ± 0.09	0.13 ± 0.22
C-VGG16	0.80 ± 0.20	0.78 ± 0.20	0.97 ± 0.05	0.88 ± 0.21	0.94± 0.06	0.06±0.06	0.12±0.21
C-Resnet34	**0.82 ± 0.18**	**0.79 ± 0.20**	**0.98 ± 0.04**	**0.89 ± 0.18**	**0.95 ± 0.05**	**0.05 ± 0.05**	**0.11 ± 0.18**
C-Incepv4	0.81±0.19	0.79±0.20	0.97±0.05	0.88±0.19	0.94±0.06	0.06±0.06	0.12±0.19

Table 3. Performance comparison of the Cancer Detection task between the ConvNet[29] and our models. ConvNet-ours: Our implementation of the ConvNet[29] that was trained on the SEER dataset and results were reported without applying the post-processing method (please see "Methods" section for the post-processing method). ConvNet-ours*: Our implemented version of the ConvNet[29] that was trained on SEER dataset and results were reported after applying post-processing method. The last three rows are preformance of our CNNs. All models were trained on the SEER dataset (D_{tr}) and evaluated on 195 TCGA WSIs (T_{tcga}).

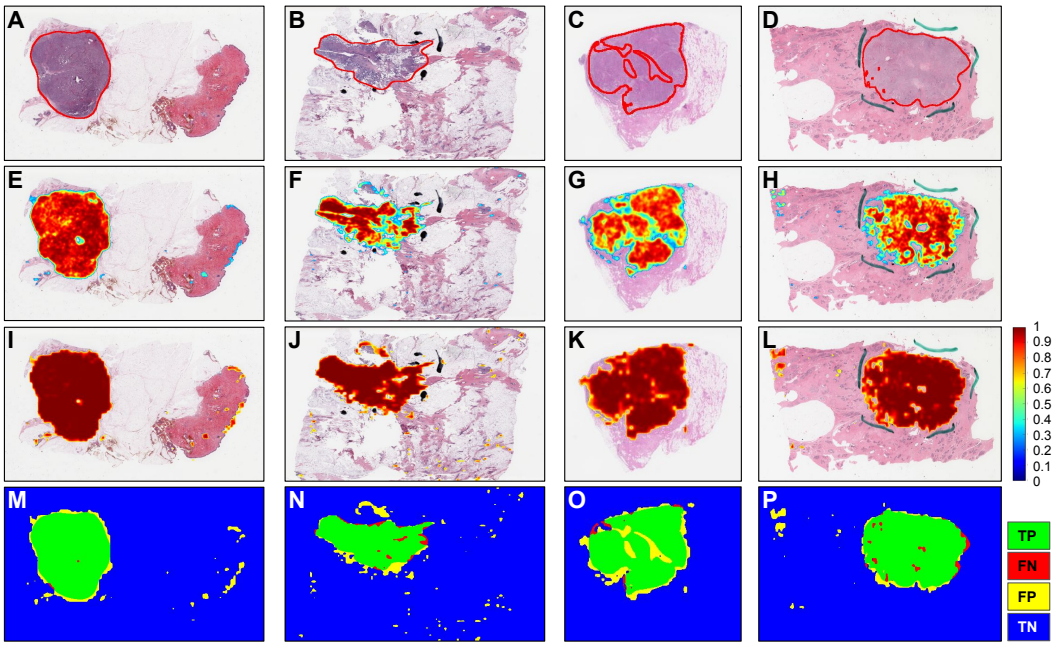

Figure 1. Prediction map of representative slides from D_{tcga}. Figures A-D are WSIs with ground truth generated by our expert pathologists. Figures E-H show the corresponding prediction heatmap generated by our Cancer detection algorithm, C-Resnet34, prior to applying any aggregation methods. Figures I-L show the corresponding prediction map after applying Max aggregation function with window size of 4 then applying threshold of 0.6 to exclude prediction scores that are less than 0.6. Figures M-P show results of our algorithm in terms of TP (green), FN (red), FP (yellow), and TN (blue) regions.

segment cancer regions faily well.

Evaluation of Lymphocyte Classification Models

We trained three lymphocyte detection models: L-VGG16, L-Resnet34, and L-Incepv4, using VGG16, Resnet34, and Inception-v4, respectively. The training dataset[32] contained only 2,912 image patches (out of 86,154) from invasive breast cancer WSIs. We found that it resulted in more accurate classification models when a network was trained with all the training patches than with only the invasive breast cancer patches. We tested the trained models with a set of image patches extracted from TCGA invasive breast cancer WSIs. Table 4 shows the performance comparison between our models with the model developed in the previous work[32]. Our new models consistently outperformed the model in the previous work in all of the performance metrics.

Our experimental evaluation showed that the cancer region segmentation and lymphocyte classification models achieved very good performance with respect to the F1-score, accuracy, and AUC metrics and performed better than the previous

Method	F1-score	Accuracy	AUC
Saltz et al.[32]	0.770	74.9%	0.808
L-VGG16	0.891	88.4%	0.943
L-Resnet34	**0.893**	**89.0%**	**0.950**
L-Incepv4	0.879	87.5%	0.938

Table 4. Performance comparison of the Lymphocytes detection task between Saltz et al.[32] and our models.

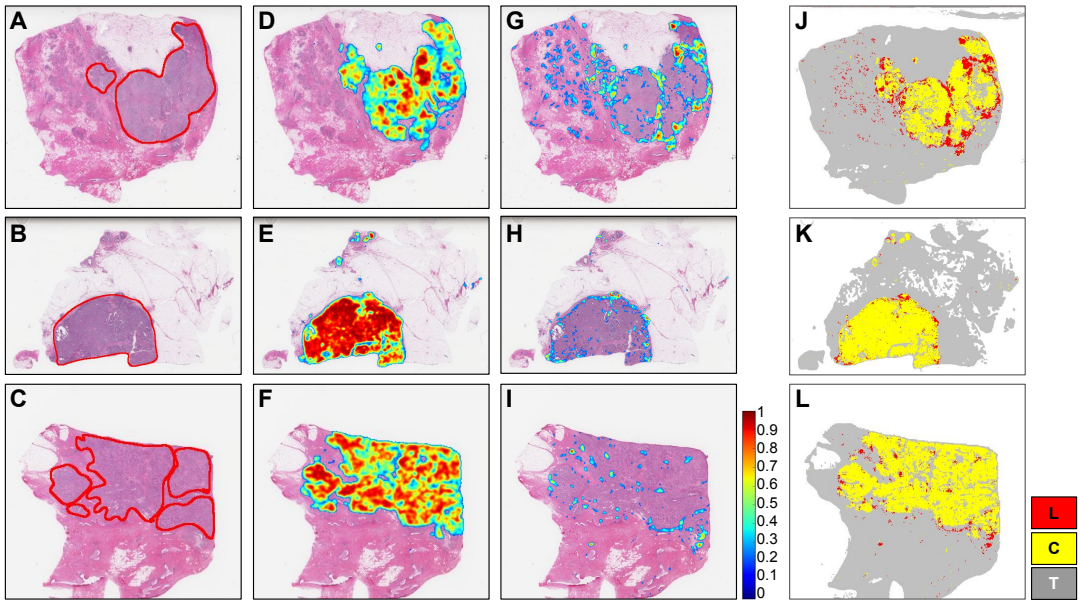

Figure 2. Cancer and lymphocyte probability maps along with map of cancer and lymphocyte labels generated through analysis of representative slides from D_{tcga} (A: TCGA-A2-A0CL-01Z-00-DX1, B: TCGA-A2-A04X-01Z-00-DX1, C: TCGA-A2-A0CW-01Z-00-DX1). Figures in a given row are results generated from the WSI depicted in the first column. Figures A-C depict WSIs with ground truth generated by our expert pathologist. Figures D-F depict the corresponding cancer probability maps generated by our cancer detection models, C-Resnet34. Figures G-I depict the corresponding lymphocyte probability maps generated by the Lymphocyte classification models, L-Resnet34. Figures J-L depict a combined heatmap of cancer and lymphocytes. Invasive breast cancer detection denoted in yellow with superimposed lymphocyte detection denoted in red. The legends of figures J-L are L, C, and T which refer to lymphocyte, cancer, and tissue region, respectively.

models. We applied the best of these models to 1,015 TCGA invasive breast cancer WSIs and generated Tumor, TIL, and combined Tumor-TIL maps. We will make these maps publicly available (please see "Data Availability" section). Figure 2 shows example Tumor-TIL combined maps overlaid on WSIs and as heatmaps. The figure visualizes the spatial relationships between lymphocytes and tumor regions. The lymphocyte patches in these examples WSIs show TILs and tumor-associated lymphocytes (TALs) that surround the cancer regions. These visual representations of TILs, TALs, and cancer regions provide valuable information for further analyses.

Discussion

Studies have shown tumor-infiltrating lymphocytes (TILs) can be used as a biomarker to predict clinical outcomes, including treatment response, in invasive breast cancer patients [9–11]. With the emergence of immunotherapy in breast cancer, the evaluation of the concentration of TILs as a readily available biomarker. As shown in Figure 3, the cancer detection algorithm shows that the cancer region occupies approximately 50-60% of the total tissue area in the WSI. The lymphocyte detection algorithm shows high probability areas with TILs. The tumor-TIL method provides insight about scattered TILs that occupy approximately 20-30% of the cancer region, consistent with a low TIL% categorization with additional spatial information that shows a sparse multi-focal distribution. Combined breast cancer tumor-TIL maps like the one shown in this example have been generated for 1,015 TCGA breast cancer WSIs and will be made publicly available in our custom web-based application.

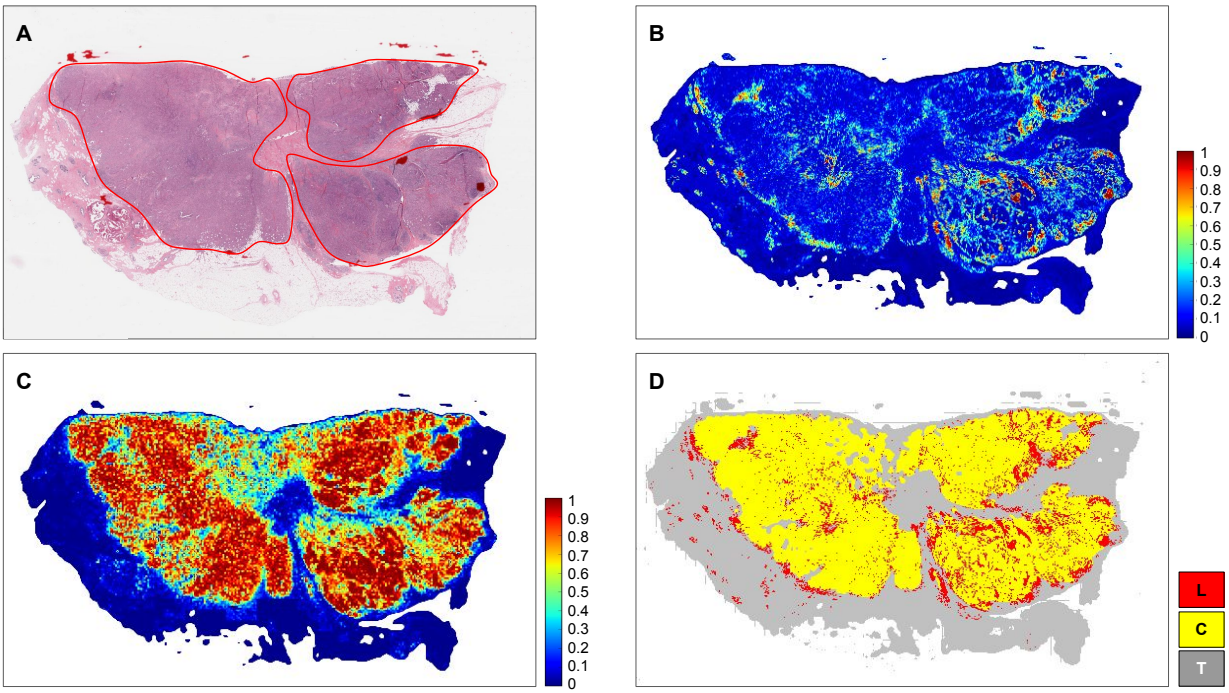

Figure 3. Enlarged example of a cancer and lymphocyte probability map and cancer along with map of cancer and lymphocyte labels for TCGA WSI (case ID: TCGA-E9-A248-01Z-00-DX1) generated by our algorithms, C-Resnet34 and L-Resnet34. A: WSI of an invasive breast cancer H&E tissue section. The viable tumor region is annotated by a pathologist with a red line. B and C are lymphocyte probability map and cancer probability map predicted by our algorithm, respectively. The probabilities are in range from 0 to 1. D: Invasive breast cancer detection denoted in yellow with superimposed lymphocyte detection denoted in red. Grey areas outside of the yellow tumor region denote non-tumor connective and adipose tissues.

The evaluation of TILs in invasive breast cancer is likely to expand due to the accumulating evidence showing how TILs can be used to predict treatment response in the settings of neoadjuvant and adjuvant chemotherapy. However, the routine evaluation of TILs has not achieved widespread adoption even though the established methodology by the International Immuno-Oncology Biomarker Working Group[18] is relatively straightforward, uncomplicated, and based on the examination of TILs on standard H&E-stained tissue sections. Figure 3 readily identifies TILs and a focal area with peritumoral TALs as a surrogate computational biomarker that is similar to how IHC is routinely utilized by pathologists to highlight cells and structures. However, IHC is not routinely performed to identify and classify subsets of TILs in breast cancer due to the time constraints of pathologists, desire to preserve diagnostic tissue, and additional costs, whereas this kind of insight can be made readily available in a low-cost and scalable manner to achieve the goals of the International Immuno-Oncology Biomarker Working Group. With emerging methods like our breast cancer tumor-TIL detection tool, pathologists will be able to add the evaluation of TILs to the standard IHC panel to determine ER, PR, HER2 expression status.

In previous work, several research groups carried out image analyses focused on detection of metastatic breast cancer[38–40] and mitosis[41–43] using highly curated but relatively small datasets from algorithm evaluation challenges[24–27]. Cruz-Roa et al. 2017 and 2018[28,29] used deep learning approaches for detecting invasive breast cancer in WSIs. The deep learning models were trained using WSIs from the Hospital of the University of Pennsylvania (HUP) and from University Hospitals Case Medical Center/Case Western Reserve University (UHCMC/CWRU) and evaluated with 195 WSIs from TCGA. Kwok[30] and Dong et al.[31] proposed methods to classify breast cancer regions in WSIs using datasets provided by the ICIAR2018 Grand Challenge on Breast Cancer Histology Images[27]. The ICIAR2018 dataset contains 2 subsets of training data: Part A consists of 400 images of 2048×1536 pixels at $0.42\,\mu m \times 0.42\,\mu m$ resolution and Part B is made up 10 WSIs with manual annotations from pathologists. Kwok[30] implemented a 2-stage training approach where a basic CNN network is trained in the first stage to mine *hard examples* on data from part B. These examples were then used to train a deep learning model in the second stage. Dong et al.[31] employed deep reinforcement learning to decide whether regions of interest should be processed for segmentation at high or low image resolutions. Most recently, Amgad, M. et al., 2019[44] proposed a fully convolutional framework for semantic segmentation of histology images via structured crowdsourcing. This was the first work using crowdsourcing in pathology task which involved a total of 25 participants at different expertise levels from medical students to expert pathologists to

generate training data for a deep learning algorithm. The authors solely focused on segmenting triple-negative breast cancer (TNBC), an aggressive genomic subtype that comprises ∼15% of breast cancer cases, into five distinct classes: Tumor, Stroma, Inflammatory Infiltration, Necrosis and Other. Using a training dataset of 151 representative region of interests (ROI, mean ROI size of $1.18mm^2$) selected from 151 H&E TCGA WSIs with detailed curated annotations, a fully convolutional VGG16-FCN-8 network was able to achieve an AUC of 0.941 for Tumor region.

The current methods for assessing TILs in individual patients are still subjective, laborious, and may be difficult to quantify. More rigorous, objective, and efficient methods are needed. This is especially true for precision medicine applications since the tumor microenvironment in breast cancer is heterogeneous and composed of malignant cells, premalignant lesions, adjacent normal tissue, stroma, immune cell infiltrates, vessels, nerves, and fat. Therefore, to help further our understanding of breast cancer biology for research and clinical applications, we developed a tumor-TIL spatial mapping tool to automatically detect breast cancer in H&E stained WSIs to quantitatively estimate and characterize the relationship between tumor cells and TILs.

In the current state, the breast tumor-TIL maps can be used to identify spatial patterns of distributions of TILs within intra- and peritumoral regions of invasive cancer, as well as lymphocyte infiltrates in adjacent tissues beyond the borders of the tumor. This tool can also be adapted for practical uses that include improving the reproducibility and precision in reporting tumor size and features of the tumor boundary for radiologic-pathologic correlation. As a potential clinical application to quantify TILs and identify spatial patterns of distribution of TILs, this tool can help guide management and select treatment in conjunction with existing molecular subtyping platforms to predict survival and recurrence since TILs are being shown to be reliable prognostic and predictive biomarkers in invasive breast cancer. Another potential application of this tool is to screen candidates who may benefit from immunotherapy in primary, refractory, and recurrent disease since such treatments are not expected to be useful if a significant amount and distribution of TILs are not present.

Most existing software algorithms for TILs assessments are proprietary, expensive, and cannot be customized by the user. Therefore, we are making our invasive breast cancer TCGA tumor-TIL dataset publicly available with an interface to visually interact with the data. The interface permits quantification of TILs in tumor areas and the ability to rapidly spot check and evaluate true-positive and false-positive predictions by the deep learning models. The invasive breast cancer TCGA-TIL maps are displayed side-by-side with an interactive H&E slide viewer to permit a high level of exploration within the entire data set. We also intend to further combine this tumor-TIL method to characterize tumor immune heterogeneity and spatially characterize local patterns of the lymphocytic infiltrate in different parts of the tumor, e.g. center of the tumor, invasive margins, and metastases. The tumor-TIL heatmaps can also be combined with other types of digital pathology-based image analyses that extract object-level of information, such as size, shape, color, texture, etc. (collectively known as Pathomics), to generate an unprecedented quantitative examination of invasive breast cancer. Such analytic data can complement traditional histopathologic evaluation that can be correlated with clinical information, radiologic imaging, molecular studies, survival, and treatment response. We believe that the availability of Tumor-TIL maps along with software that allows interactive viewing of the computational analysis will improve reproducibility and precision in reporting tumor size, tumor boundary features, TILs assessment, and extraction of relevant nuclear and cellular features. These improvements will in turn enhance clinical and pathology decision support in guiding management, treatment selection, and predicting survival and recurrence, in conjunction with existing molecular subtyping platforms.

The need to quantify spatial inter-relationships between tumor regions and infiltrating lymphocytes is becoming increasingly important in invasive breast cancer. Tumor-TIL maps generated from H&E images can be employed to carry out a wide range of correlative studies in the context of clinical trials, epidemiological investigations, and surveillance studies. Our methods leverage open source convolutional neural networks; the programs we have developed are also being made public and freely available. The results of this effort is a reliable and robust methodology, datasets, and programs that can be employed to carry out tumor-TIL analyses for 1,015 TCGA breast cancer WSIs. In future studies, we will further refine this tool to differentiate between invasive and *in situ* premalignant lesions and explore methods that can facilitate faster predictions for practical real-time clinical applications.

Methods

Ethics Statement

The WSI from SEER came from a larger pilot program examining the feasibility of and best practices for the virtual tissue repository (VTR), which included WSI (VTR Pilot). As all data in the VTR Pilot, including the whole slide images, were deidentified prior to receipt by the NCI SRP, the NIH Office of Human Subjects Research Protection determined that the study was excluded from NIH IRB review. The SEER registries supplying the deidentified WSIs each obtained IRB approval from their respective institution(s). The Stony Brook IRB has classified the dataset as being a non-human subjects research dataset.

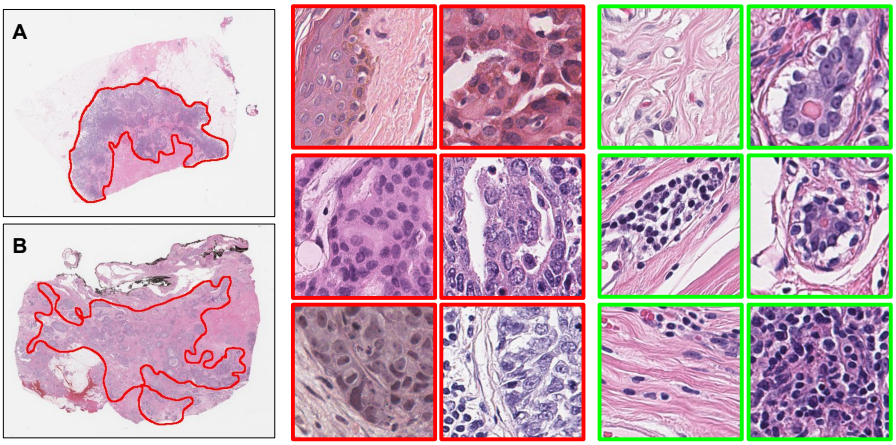

Figure 4. Annotation example from pathologist (Figures A and B) and image patches extracted from WSIs. Red lines in figures A and B were manually drawn by expert pathologist to enclose the cancer regions. Regions outside the annotated regions are non-cancer regions. Patches surrounded in red boxes are positive samples which contain invasive cancer cells. Patches surrounded in green boxes are negative samples which do not contain invasive cancer cells.

Patch Extraction for the Breast Cancer Dataset

We extracted image patches at the highest image resolution within and outside the segmented cancer regions using an open source library called OpenSlide[45]. Patches with a size of 350×350 pixels at 400x magnification (equivalent to 88 µm×88 µm) resulted in the best classification performance and were used to create the training datasets. Each patch was labeled positive (i.e., it intersected or was in a cancer/tumor region) or negative (i.e., it was outside cancer/tumor regions). Figure 4 shows an example of the pathologist's annotations. The region inside the red line represents the cancer region. The figure also shows the sample patches extracted from the cancer and non-cancer regions. We trained the cancer detection models with patches from the SEER data. In order to compare our results, we evaluated the models on a publicly dataset consisted of 195 TCGA WSIs provided in Cruz-Roa et al.[29].

Previous work has shown that it is beneficial to have more negative samples than positive samples in a training dataset for image classification in digital pathology[29,46–48]. A good ratio of negative to positive samples (i.e., patches in our case) will increase the generalization of a CNN model and decrease false positive rate. We experimented with a range of ratios of negative patches to positive patches with the same validation dataset. The details of the training, validation, and test datasets are presented in Table 1.

Convolutional Neural Networks

We investigated multiple state-of-the-art deep learning architectures for the task of classifying cancer regions and lymphocyte regions in WSIs. Considering both the computational complexity and the capacity of different neural networks, we used three CNN architectures: the VGG 16-layer[34], the Resnet 34-layer[35], and the Inception-v4 network[36]. These models are state-of-the-art CNN architectures which are widely used in a range of application domains. Earlier work[49,50] showed that refining a CNN pre-trained on the ImageNet dataset[51] is a good approach to boosting image classification performance in digital pathology. In our work, we refined the pre-trained CNN models with our training data. We implemented the CNN networks using pyTorch 0.4[37].

The VGG16 and Resnet34 were designed to process 224×224-pixel patches whereas the Inception-v4 accepts 299×299-pixel image patches. Since our dataset consists of 350×350-pixel patches at 400x magnification, input patches were resized to the desired input size for each network. In addition, for Resnet34 and Inception-v4, we changed the dimension of the output layer from 1,000 classes to two classes for our binary classification problem. For VGG16, we reduced the size of the intermediate features of the classification layer from 4,096 to 1,024 and only kept the first four layers in the classification layer. This modification reduced the number of trainable parameters of this network from 138 Millions to 41 Millions. Our modifications to the classification layers of the CNN architectures are presented in Table 5.

The Lymphocyte training datasets contain 100×100-pixel patches at 200x magnification. We resized the input 100×100-pixel patches to 200×200-pixel patches for the Lymphocyte classification CNNs.

We used the same training procedure for all of the networks. At the beginning of the training, the weights of the networks were initially fixed except for the classification layer. The networks were trained in this state for N epochs (N is three for the Cancer models and N is five for the Lymphocyte models) with a batch size of B (B is 256 for the Cancer models and B is 128

VGG16		Resnet34		Inception-v4	
Original	Modified	Original	Modified	Original	Modified
Linear(25088,4096) ReLU → Dropout Linear(4096,4096) ReLU → Dropout Linear(4096,1000)	Linear(25088,1024) ReLU → Dropout Linear(1024,2)	Linear(512,1000)	Linear(512,2)	Linear(1536,1000)	Linear(1536,2)

Table 5. Modifications to the classification layers of the CNNs

for the Lymphocyte models) , an initial learning rate of 0.01, a momentum of 0.9 and a weight decay of 0.0001. After N epochs, the training process enabled updates to the initially fixed weights. The network was then trained for total of 20 epochs, updating all the weights. The training process used a stochastic gradient descent method[52] in order to minimize a cross entropy loss function.

The color profiles of WSIs may vary from image to image because of variations in staining and image acquisition[53–55]. We used data augmentation to reduce the effects of color/intensity variability and data acquisition artifacts. The data augmentation operations included random rotation between 0 and 22.5 degrees, random vertical and horizontal flipping, perturbations in patch brightness, contrast, saturation, and normalization of R,G, and B channels to a mean of 0.0 and standard deviation of 1.0.

In the prediction (test) phase, no data augmentation was applied except for the normalization of the color channels. Each patch was assigned a predicted classification probability value between 0.0 and 1.0 by the trained model, indicating the probability of a patch being positive. This value is thresholded to generate the final class label of each patch.

Post-processing Step for Cancer Heatmaps

Most patch-based classification algorithms[56,57] predict the label of a patch independent of other patches in an image. They do not take into account the characteristics and labels of neighbor patches. Invasive cancer regions in breast cancer tend to be close to each other. In other words, the probability of a patch to be positive is correlated to its surrounding patches. To incorporate this information in our analysis pipeline, we employed a simple, yet effective, aggregation approach as a post-processing step. This approach takes per-patch classification probability values, converts them into a probability map, called **H**, and produces an aggregated probability map, called **A**. The classification probability value of a patch in **A** is computed by an aggregation operation over neighbor patches within a specific distance of the patch in **H**. The relationship between the aggregated probability map **A** and **H** can be formulated as follows:

$$\mathbf{A}(i,j) = f\left(\left\{\mathbf{H}(m,n) \middle| m \in \left[\left\lfloor \frac{i}{w} \right\rfloor w, \left(\left\lfloor \frac{i}{w} \right\rfloor + 1\right)w\right], n \in \left[\left\lfloor \frac{j}{w} \right\rfloor w, \left(\left\lfloor \frac{j}{w} \right\rfloor + 1\right)w\right]\right\}\right) \tag{1}$$

Here, $\mathbf{H}(m,n)$ is the probability values of a patch at location (m,n) of the original probability map. Similarly, $\mathbf{A}(i,j)$ is the probability value of the aggregated patch at location (i,j) computed by the aggregation operation. f is the aggregation function over a set of patches in a window of $\left[\left\lfloor \frac{i}{w} \right\rfloor w, \left(\left\lfloor \frac{i}{w} \right\rfloor + 1\right)w\right] \times \left[\left\lfloor \frac{j}{w} \right\rfloor w, \left(\left\lfloor \frac{j}{w} \right\rfloor + 1\right)w\right]$, and w is the window size. *In our aggregation approach, all patches within the window will have the same prediction score after the aggregation operation.* $\lfloor x \rfloor$ is the *Floor* operation which takes x as an input and returns the largest integer that is less than or equal to x. We explored different aggregation function such as Average, Median, and Max. The experiments were carried out using T_{seer}. The best aggregation method from these experiments was used to generate aggregated probability maps for T_{tcga}. Empirically, we have found that the Max function and a window of 4x4 resulted in the best performance on the T_{seer}. Hence, we applied these settings for the post-processing step on the T_{tcga} WSIs.

Combined Tumor-TIL Maps

We merged each pair of cancer and lymphocyte heatmaps into a single heatmap as an RGB image. The R channel stores the lymphocyte probabilities quantized to 0-255; the G channel stores the cancer probabilities quantized to 0-255; and the B channel stores if a patch is tissue or glass background.

Software Tools for Data Management and Visualization

Engaging multivariate analysis of digital pathology data in a manner that is transparent to the domain expert has always been challenging due to the size of whole slide images. A typical digitized whole slide image (WSI), at 400x magnification, is 1-4 GBs in size. A modest cohort of a hundred patients (approximately two slides per patient) can easily result in 0.5-1 TBs of

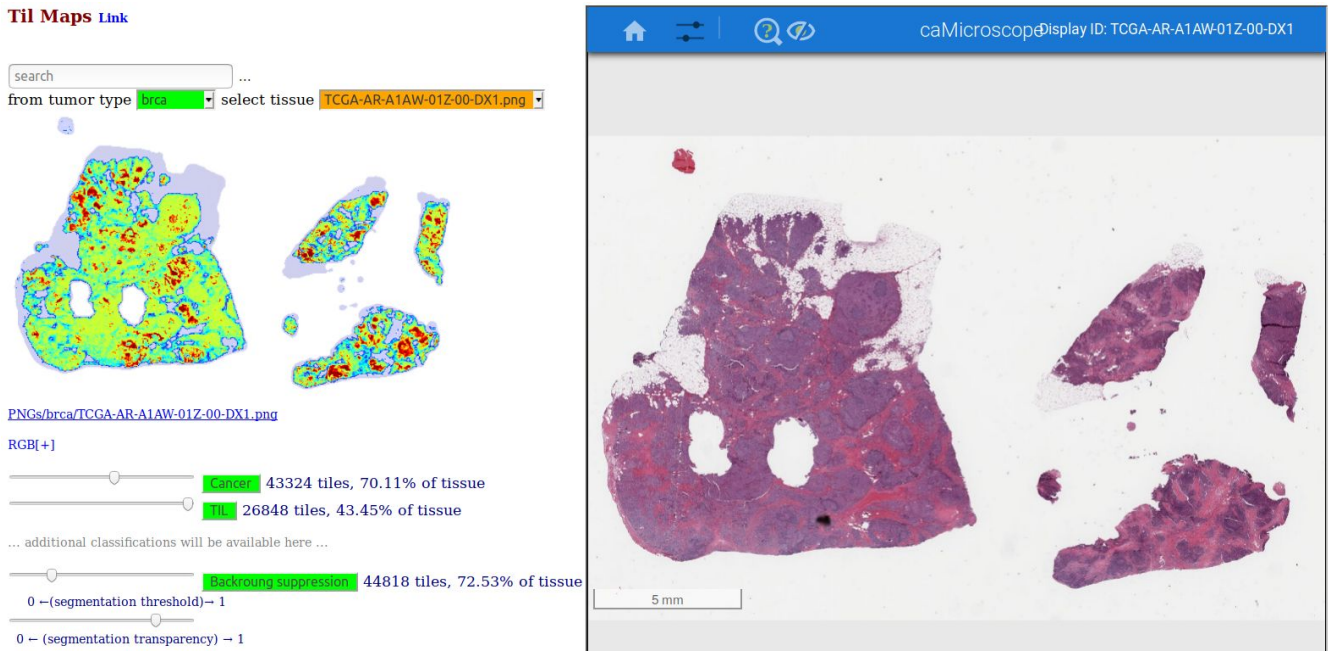

Figure 5. User interface of our web-based application to study the spatial relationship between cancer regions and lymphocyte regions. Figure on the left shows the TILs heatmap where invasive breast cancer detection denoted in yellow with superimposed lymphocyte detection denoted in red. Figure on the right is the caMicroscope[33] that displays the regions of the WSI. Users can click on the TILs map to zoom in the corresponding regions on the caMicroscope.

image data. Processing, analyzing, and visualizing data of this size is a non-trivial task. Various image processing and analysis methods, including deep learning based classifiers, are used to generate high-level representations (eg., areas of an image with a high probability of being tumor) and low-level representations (eg., segmented nuclei). The examination of these results and their representations requires their interactive interrogation and visual analytic tools that link these representations with the underlying images as well as associated clinical, demographic, and molecular data. We have developed a portable FeatureMap web-based application that combines cached calling to a data-intensive backend (PathDB - https://github.com/SBU-BMI/pathdb) with a browser-based multivariate visualization library that is sufficiently lightweight to run on a mobile device. We employed the FeatureMap application to host and visualize TILs maps of 1,015 TCGA invasive breast cancer cases. PathDB is integrated with our pathology viewer, caMicroscope[33], that allows users to click on parts of a TIL Map, zoom-in on the corresponding region on the WSI image, and interactively examine interesting areas of the TIL Map. Users can also freely move to any region in the WSI by dragging the cursor and zoom in or zoom out by scrolling the mouse. Readers are welcome to try our application at https://mathbiol.github.io/tcgatil/. Figure 5 shows a screenshot of our web-based application. Technical details of FeatureMap will be provided in a separate paper.

An algorithm for processing and analysis of WSI has significant computational needs, often requiring multiple CPU cores, large amounts of RAM, and one or more GPUs. Given the size of an individual image and the volume of data in a cohort, the processing and analysis of WSI on a large scale are significant challenges. Cloud computing today affords researchers with an unlimited pool of computing that is elastic, scalable, and provisioned on-demand. Our team has leveraged the cost and computational advantages of cloud computing by deploying our various deep learning algorithms for computational identification of TILs, similar algorithms for nuclear segmentation, tumor region segmentation, and characterizing and quantifying other Pathomics features. The deep learning algorithms and associated methods to characterize features are containerized using Docker and then deployed as workflows on the Google Cloud. The workflows are represented using the Workflow Description Language[58] and orchestrated on the cloud behind APIs. This pathway also allows us to disseminate algorithms and workflows with other researchers without requiring extensive installation or software maintenance, while users only pay for the computing they use to run the algorithms. A detailed description of our cloud-based algorithm management and dissemination system will be provided in a separate paper.

Data Availability

The SEER images used in the training dataset were gathered in a work carried out with the SEER consortium and will be made publicly available in the future as part of a separate pilot project. The invasive breast cancer images are publicly available and provided by TCGA (http://cancergenome.nih.gov/ and the Genomic Data Commons (GDC) Data Portal in https://portal.gdc.cancer.gov/).

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA: a cancer journal for clinicians* **69**, 7–34 (2019).
2. Anderson, W. F. & Matsuno, R. Breast cancer heterogeneity: a mixture of at least two main types? (2006).
3. Heselmeyer-Haddad, K. *et al.* Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of myc during progression. *The Am. journal pathology* **181**, 1807–1822 (2012).
4. Denkert, C. *et al.* Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *The lancet oncology* **19**, 40–50 (2018).
5. Loi, S. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase iii randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: Big 02-98. *J Clin Oncol* **31**, 860–867 (2013).
6. Mao, Y. *et al.* The value of tumor infiltrating lymphocytes (tils) for predicting response to neoadjuvant chemotherapy in breast cancer: a systematic review and meta-analysis. *PloS one* **9**, e115103 (2014).
7. Pruneri, G. *et al.* Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. *Annals oncology* **27**, 249–256 (2015).
8. Denkert, C. *et al.* Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* **28**, 105–113 (2010).
9. Denkert, C. *et al.* Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Mod. Pathol.* **29**, 1155 (2016).
10. Loi, S. *et al.* Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the finher trial. *Annals oncology* **25**, 1544–1550 (2014).
11. Heppner, B. I. *et al.* Tumor-infiltrating lymphocytes: a predictive and prognostic biomarker in neoadjuvant-treated her2-positive breast cancer. *Clin. Cancer Res.* **22**, 5747–5754 (2016).
12. Salgado, R. *et al.* Tumor-infiltrating lymphocytes and associations with pathological complete response and event-free survival in her2-positive early-stage breast cancer treated with lapatinib and trastuzumab: a secondary analysis of the neoaltto trial. *JAMA oncology* **1**, 448–455 (2015).
13. Adams, S. *et al.* Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase iii randomized adjuvant breast cancer trials: Ecog 2197 and ecog 1199. *J. clinical oncology* **32**, 2959 (2014).
14. West, N. *et al.* Tumour-infiltrating foxp3+ lymphocytes are associated with cytotoxic immune responses and good clinical outcome in oestrogen receptor-negative breast cancer. *Br. journal cancer* **108**, 155 (2013).
15. Strome, S. E., Sausville, E. A. & Mann, D. A mechanistic perspective of monoclonal antibodies in cancer therapy beyond target-related effects. *The oncologist* **12**, 1084–1095 (2007).
16. Sharma, P. & Allison, J. P. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* **161**, 205–214 (2015).
17. Gotwals, P. *et al.* Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nat. Rev. Cancer* **17**, 286 (2017).
18. Dieci, M. V. *et al.* Update on tumor-infiltrating lymphocytes (tils) in breast cancer, including recommendations to assess tils in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the international immuno-oncology biomarker working group on breast cancer. In *Seminars in cancer biology*, vol. 52, 16–25 (Elsevier, 2018).
19. Salgado, R. *et al.* The evaluation of tumor-infiltrating lymphocytes (tils) in breast cancer: recommendations by an international tils working group 2014. *Annals oncology* **26**, 259–271 (2014).
20. Denkert, C. *et al.* Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Mod. Pathol.* **29**, 1155 (2016).

21. Klauschen, F. *et al.* Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning. In *Seminars in cancer biology* (Elsevier, 2018).

22. Catacchio, I. *et al.* Intratumoral, rather than stromal, cd8+ t cells could be a potential negative prognostic marker in invasive breast cancer patients. *Transl. oncology* **12**, 585–595 (2019).

23. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press, 2016).

24. Bejnordi, B. E. *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama* **318**, 2199–2210 (2017).

25. Bándi, P. *et al.* From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE Transactions on Med. Imaging* (2018).

26. Litjens, G. *et al.* 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *GigaScience* **7**, giy065 (2018).

27. Aresta, G. *et al.* Bach: Grand challenge on breast cancer histology images. *arXiv preprint arXiv:1808.04277* (2018).

28. Cruz-Roa, A. *et al.* Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Sci. reports* **7**, 46450 (2017).

29. Cruz-Roa, A. *et al.* High-throughput adaptive sampling for whole-slide histopathology image analysis (hashi) via convolutional neural networks: Application to invasive breast cancer detection. *PloS one* **13**, e0196828 (2018).

30. Kwok, S. Multiclass classification of breast cancer in whole-slide images. In *International Conference Image Analysis and Recognition*, 931–940 (Springer, 2018).

31. Dong, N. *et al.* Reinforced auto-zoom net: Towards accurate and fast breast cancer segmentation in whole-slide images. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 317–325 (Springer, 2018).

32. Saltz, J. *et al.* Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports* **23**, 181 (2018).

33. Saltz, J. *et al.* A containerized software system for generation, management, and exploration of features from whole slide tissue images. *Cancer research* **77**, e79–e82 (2017).

34. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *ICLR* (2015).

35. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

36. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence* (2017).

37. Paszke, A., Gross, S., Chintala, S. & Chanan, G. Pytorch (2017).

38. Wang, D., Khosla, A., Gargeya, R., Irshad, H. & Beck, A. H. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718* (2016).

39. Nazeri, K., Aminpour, A. & Ebrahimi, M. Two-stage convolutional neural network for breast cancer histology image classification. In *International Conference Image Analysis and Recognition*, 717–726 (Springer, 2018).

40. Golatkar, A., Anand, D. & Sethi, A. Classification of breast cancer histology using deep learning. In *International Conference Image Analysis and Recognition*, 837–844 (Springer, 2018).

41. Albarqouni, S. *et al.* Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE transactions on medical imaging* **35**, 1313–1321 (2016).

42. Veta, M. *et al.* Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. image analysis* **20**, 237–248 (2015).

43. Rao, S. Mitos-rcnn: A novel approach to mitotic figure detection in breast cancer histopathology images using region based convolutional neural networks. *arXiv preprint arXiv:1807.01788* (2018).

44. Amgad, M. *et al.* Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics* (2019).

45. Goode, A., Gilbert, B., Harkes, J., Jukic, D. & Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Informatics* **4**, 27, DOI: 10.4103/2153-3539.119005 (2013). openslide.org.

46. Hagos, Y. B., Mérida, A. G. & Teuwen, J. Improving breast cancer detection using symmetry information with deep learning. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, 90–97 (Springer, 2018).

47. Liu, Y. *et al.* Artificial intelligence–based breast cancer nodal metastasis detection: Insights into the black box for pathologists. *Arch. pathology & laboratory medicine* (2018).

48. Lee, B. & Paeng, K. A robust and effective approach towards accurate metastasis detection and pn-stage classification in breast cancer. In *MICCAI* (2018).

49. Xu, Y. *et al.* Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 947–951 (IEEE, 2015).

50. Hou, L. *et al.* Automatic histopathology image analysis with cnns. *Proc. New York Sci. Data Summit (NYSDS)* (2016).

51. Russakovsky, O. *et al.* Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).

52. Bengio, Y., Courville, A. & Vincent, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis machine intelligence* **35**, 1798–1828 (2013).

53. Macenko, M. *et al.* A method for normalizing histology slides for quantitative analysis. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, 1107–1110 (IEEE, 2009).

54. Reinhard, E., Adhikhmin, M., Gooch, B. & Shirley, P. Color transfer between images. *IEEE Comput. graphics applications* **21**, 34–41 (2001).

55. Vahadane, A. *et al.* Structure-preserved color normalization for histological images. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, 1012–1015 (IEEE, 2015).

56. Hou, L. *et al.* Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2424–2433 (2016).

57. Zhu, X., Yao, J., Zhu, F. & Huang, J. Wsisa: Making survival prediction from whole slide histopathological images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 7234–7242 (2017).

58. Birger, C. *et al.* Firecloud, a scalable cloud-based platform for collaborative genome analysis: Strategies for reducing and controlling costs. *bioRxiv* DOI: 10.1101/209494 (2017).

Acknowledgements

This work was supported in part by 1U24CA180924-01A1, 3U24CA215109-02, and 1UG3CA225021-01 from the National Cancer Institute, R01LM011119-01 and R01LM009239 from the U.S. National Library of Medicine. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC). NCI Surveillance Research Program overseeing the Virtual Tissue Repository (VTR) Pilot Program, from which participating SEER cancer registries (Greater California, Connecticut, Hawaii, Iowa, Kentucky, and Louisiana) supplied the whole slide images utilized in algorithm development and testing. The SEER VTR Pilot Program is supported by the Division of Cancer Control and Population Sciences at the National Cancer Institute of the National Institutes of Health.

Author Contributions

Conceptualization, J.S, T.K, R.G, H.L, A.L.V.D, D.S, D.F, T.Z, R.B; Methodology, J.S, T.K, H.L, L.H, S.A, D.S; Data Curation, R.G; Running Experiments, H.L, S.A; Writing – Original Draft, H.L, R.G, T.K, J.S.A, A.S; Writing – Review & Editing, H.L, R.G, T.K, J.S, A.L.V.D; Formal Analysis, J.S, R.G, A.L.V.D; Training Convolutional Neural Networks R.G, H.L; Supervision, J.S, T.K, D.S; Visualization, H.L, J.S.A, E.B; Software, E.B, J.S.A, A.S.

Additional information

Competing interests: The author(s) declare no competing interests.