Deep Learning to Estimate Human Epidermal Growth Factor Receptor 2 Status from Hematoxylin and Eosin-Stained Breast Tissue Images

Deepak Anand[1,*], Nikhil Cherian Kurian[1,*], Shubham Dhage[1], Neeraj Kumar[2,3], Swapnil Rane[4], Peter H. Gann[5], Amit Sethi[1,5]

[1]Department of Electrical Engineering, IIT Bombay, Mumbai, Maharashtra, India, [2]Department of Computing Science, University of Alberta, Edmonton, Canada, [3]Alberta Machine Intelligence Institute, Edmonton, Canada, [4]Department of Pathology, Tata Memorial Centre - ACTREC, HBNI, Navi Mumbai, Maharashtra, India, [5]Department of Pathology, University of Illinois, Chicago, USA

*Equal Contribution

Abstract

Context: Several therapeutically important mutations in cancers are economically detected using immunohistochemistry (IHC), which highlights the overexpression of specific antigens associated with the mutation. However, IHC panels can be imprecise and relatively expensive in low-income settings. On the other hand, although hematoxylin and eosin (H&E) staining used to visualize the general tissue morphology is a routine and low cost, it does not highlight any specific antigen or mutation. **Aims:** Using the human epidermal growth factor receptor 2 (HER2) mutation in breast cancer as an example, we strengthen the case for cost-effective detection and screening of overexpression of HER2 protein in H&E-stained tissue. **Settings and Design:** We use computational methods that reliably detect subtle morphological changes associated with the over-expression of mutation-specific proteins directly from H&E images. **Subjects and Methods:** We trained a classification pipeline to determine HER2 overexpression status of H&E stained whole slide images. Our training dataset was derived from a single hospital containing 26 (11 HER2+ and 15 HER2−) cases. We tested the classification pipeline on 26 (8 HER2+ and 18 HER2−) held-out cases from the same hospital and 45 independent cases (23 HER2+ and 22 HER2−) from the TCGA-BRCA cohort. The pipeline was composed of a stain separation module and three deep neural network modules in tandem for robustness and interpretability. **Statistical Analysis Used:** We evaluate our trained model through area under the curve (AUC)-receiver operating characteristic. **Results:** Our pipeline achieved an AUC of 0.82 (confidence interval [CI]: 0.65–0.98) on held-out cases and an AUC of 0.76 (CI: 0.61–0.89) on the independent dataset from TCGA. We also demonstrate the region-level correspondence of HER2 overexpression between a patient's IHC and H&E serial sections. **Conclusions:** Our work strengthens the case for automatically quantifying the overexpression of mutation-specific proteins in H&E-stained digital pathology, and it highlights the importance of multi-stage machine learning pipelines for added robustness and interpretability.

Keywords: Breast cancer, convolutional neural networks, histopathology, human epidermal growth factor receptor 2, immunohistochemistry, mutation detection, nucleus detection

Introduction

Breast carcinoma has the highest mortality rate among women.[1,2] Standard treatments differ by the well-recognized subtypes of breast cancer that are characterized by the different sets of mutations. Among the five major subtypes in breast cancer – luminal A, luminal B, human epidermal growth factor receptor 2 (HER2), basal and normal-type – the HER2 subtype was, overall, the most fatal until the last millennium.[3] The HER2 subtype is characterized by the over-expression of the HER2

gene, which is commonly accompanied by the mutations in other genes from the HER2 amplicon such as GRB7,[4,5] PGAP3,[4] and

How to cite this article: Anand D, Kurian NC, Dhage S, Kumar N, Rane S, Gann PH, *et al*. Deep learning to estimate human epidermal growth factor receptor 2 status from hematoxylin and eosin-stained breast tissue images. J Pathol Inform 2020;11:19.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2020/11/1/19/290573

Access this article online	
Quick Response Code:	**Website:** www.jpathinformatics.org
	DOI: 10.4103/jpi.jpi_10_20

to a certain extent TP53, which promotes tumor proliferation.[6] The advent of targeted anti-HER2 therapies (e.g. trastuzumab, lapatanib, and pertuzumab) has reduced mortality for HER2 positive breast cancer, but such therapy is expensive. Furthermore, for other subtypes of breast cancer, anti-HER2 therapies are useless, and in some cases, even harmful.[7] Therefore, it is important to determine HER2 overexpression in breast cancers. HER2 overexpression is sometimes identified using expensive but accurate precise fluorescence *in situ* hybridization (FISH) tests, or, more commonly, using inexpensive but less accurate HER2neu immunostaining.[6]

Immunohistochemistry (IHC) is the class of techniques used for visually tagging parts of tissue with high concentrations of specific antigens (proteins) for microscopic examination. IHC reagent is a combination of the tagging (dyeing or fluorescence) agent and an antibody that binds to the specific target antigen. It is widely used to diagnose the malignancy in tumors and determine genomic subtypes of cancers for precision therapy. By contrast, hematoxylin and eosin (H&E) staining is generic because it primarily increases the visual difference between the general basophilic nuclei and acidophilic stromal regions of the tissue to reveal the spatial structure like the shape of nuclei and glands. While costs and availability of various IHC panels vary widely depending on the targeted proteins, H&E staining is inexpensive and ubiquitous.

Morphological correlates of specific mutations have been observed in H&E stained histology images (e.g. for KRAS mutation in lung cancer[8]). For other mutations, it is possible that other specific morphological correlates also exist to allow the screening of the mutation using computer vision on H&E images, even if these are too subtle to be reliably detected by manual inspection. To test this hypothesis, we specifically explore the detectability of HER2 overexpression in breast tumors from H&E-stained tissue images. We used a local HER2neu IHC response to prepare supervised training data for a computer-vision pipeline. The advantages of doing so include: (1) a reduction in the cost of tissue analysis by obviating IHC for screening, (2) the potential of predicting the presence of multiple antigens exclusively from H&E images by extending the technique to other mutations, and (3) the quantification of intra-tumor heterogeneity of genomic mutations in different regions of an individual's tumor. That is, while a single IHC panel tags only a single protein, which is usually a product of a single class of genomic mutations, morphological analysis of visual patterns revealed by H&E can find spatial foci of different mutation classes. To the best of our knowledge, such a study has not been done before for a major IHC panel such as HER2 in breast cancer.

Conventionally, image classification pipelines were based on hand-engineered features, which failed to generalize in the clinical settings. Deep neural networks that learn their own hierarchy of features have produced unprecedented image classification accuracy and have obviated the need for hand-engineered features in less than a decade for large datasets of labeled training examples.[9-11] With increasing the use of digital scanners, the volume of digitized whole slide images (WSIs) available for computation pathology using deep learning has grown multiple folds. This has enabled the emergence of computer vision pipelines to supplement and complement the pathologists.[12] For example, in the CAMELYON16 challenge for lymph node analysis, the trained deep-learning models reported a near human pathologist performance.[13] This is complemented by the recent surge in publicly available computational pathology datasets released in several international competitions.[14-16]

We have developed a method to classify between HER2 positive (HER2+) and HER2 negative (HER2–) breast tumors from their H&E-stained sections. We trained a multi-stage image classification pipeline composed of a stain separation stage followed by three convolutional neural networks (CNNs) in tandem to move toward explainable AI on carefully annotated data. The training data came from serial (adjacent) sections of tumor tissues, of which one is stained with H&E and the other one with HER2neu IHC. We tested the trained pipeline on held out cases from the same cohort as well as on an independent cohort. We achieved an area under the curve (AUC) of 0.82 (confidence interval [CI]: 0.65–0.98) on the held-out cases in the Warwick dataset and AUC of 0.76 (CI: 0.61–0.89) on the TCGA-BRCA cohort. We also visualized the regional correspondence of HER2 probability in H&E sections and their serial IHC sections in held-out cases.

The rest of the paper is organized as follows: background and data sources are presented in section 2, the proposed methods are described in section 3, and the results are presented in section 4. Finally, the concluding remarks appear in section 5.

SUBJECTS AND METHODS

Estimating human epidermal growth factor receptor 2 protein overexpression in breast cancers
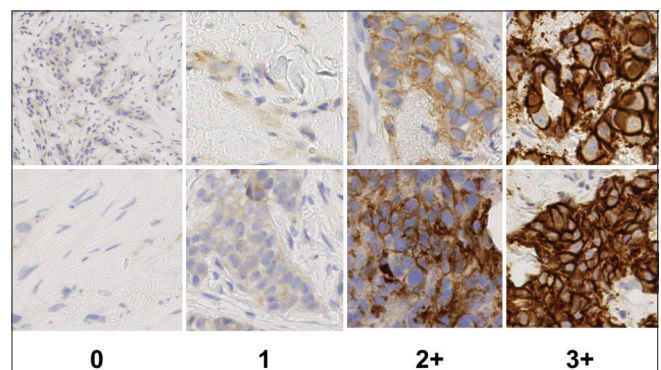
HER2 is a growth-promoting protein on the cell membrane. Assessment of HER2 IHC is done based on the guidelines given by the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP).[17] Pathologists observe formalin-fixed paraffin-embedded tissue stained using HER2 IHC under a microscope. Nuclei appear as faint blue due to counterstaining using hematoxylin. In HER2âcells, no or faint brown pattern of the HER2 IHC stain is seen around the cell membrane. In HER2+ cells, a brown boundary appears circumscribing the nucleus (chickenwire structure) marking the presence of HER2 protein. Pathologists score the IHC stained tissue slides by visual inspection under a microscope using the ASCO/CAP criteria is shown in Table 1 and illustrated in Figure 1. The interpretation of these guidelines by pathologists is subjective, especially, near the boundaries of two classes. For example, the percent of invasive cells with a particular staining pattern is subjectively estimated, and the difference between "faint" and "moderate" staining is also subjective as illustrated in Figure 1. If HER2 IHC results are equivocal (borderline 2+), then an expensive FISH test is ordered for disambiguation.

Identification of genetic mutations from histopathology images is gaining a lot of interest among the researchers. We clarify that we are not going to address the easier task of automatic scoring of IHC, such as that of HER2neu positivity.[18] Instead, we concentrate on detecting mutations in H&E-stained tissue images by estimating the over-expression of mutation-specific proteins. CNNs have been successfully used for detecting STK11, EGFR, FAT1, SETBP1, KRAS, and TP53 mutations in lung adenocarcinoma,[19] BRAF and NRAS mutations in melanoma,[20] and estrogen-receptor positivity in breast cancer using H&E images.[21-23] These methods use a single deep-learning stage and have not been used to show strong generalization from one to many hospitals. In contrast to previous works, our proposed method is able to give nucleus level correspondence with IHC-stained sections in H&E-stained tissue sections. Apart from this, the proposed method was able to achieve strong generalization from training data source from one hospital to a test data sourced from many hospitals.

Human epidermal growth factor receptor 2 immunohistochemistry and hematoxylin and eosin data sources

To increase the precision and accuracy of HER2 IHC assessment, a contest to train and test computational methods for HER2 scoring was organized by the Tissue Image Analytics (TIA) Lab at the University of Warwick in 2016.[18] WSIs of two serial sections of breast carcinoma tissue stained with HER2 IHC and H&E, respectively, were released for each patient. WSIs were scanned using a Hamamatsu^R scanner at × 40 magnification (with × 10 objective such that each pixel covered a square of side 0.25 μm).

We used a subset of the released dataset for our study. There were 26 cases in the training dataset whose HER2 scores were known, and 26 cases in the testing dataset whose scores were estimated by the trained neural network. Of the training cases, 15 were HER2– (scores 0 or 1+) and 11 were HER2+ (score 3+). We did not include slides with ambiguous HER2 IHC (score 2+) in training or testing. H&E sections were released to help the contestants to assess the percent of complete membrane staining, which was the secondary objective of the contest. These data gave us an excellent opportunity to study the spatial association between HER2 IHC and any corresponding morphological pattern in H&E. To test our trained model on an independent multi-hospital dataset, we took H&E-stained whole-slide images of 45 cases from the TCGA-BRCA cohort. Out of the 45 cases in TCGA-BRCA cohort, 23 were HER2+ and 22 HER2–. The details of both datasets are shown in Table 2.

Methods

Breast cancer WSIs have a lot of heterogeneous regions including uninformative (or nontumorous) ones such as background whitespace, adipose tissue (fat), blood vessels, nerves, benign epithelium, stroma, inflammation, etc., as shown in Figure 2. While deep learning is often applied in one go on the entire WSI, such an approach does not generalize well outside the training cohort and is also computationally inefficient. Therefore, we broke the problem of analyzing a WSI using a cascade of steps, including stain separation, followed by three neural networks that funnel down to the classes of interest. This was accomplished using the following steps, which are also illustrated in Figure 3. For testing, we start with a patient's WSI and extract patches by rejecting the background regions. All nuclei are detected in the extracted patch to further mine smaller nuclei-centric patches. These small patches are passed through tumor versus non-tumor classifiers to remove non-tumorous patches from further processing. The tumorous patches from the previous classifier are further tested for HER2+ versus HER2– classification. Finally, we use a threshold on the percentage of HER2+ nuclei for ascertaining a patient's HER2 status. This process depended on careful data preparation, which we describe first.

Table 1: Human epidermal growth factor receptor 2 scoring guidelines by the American Society of Clinical Oncology/College of American Pathologist

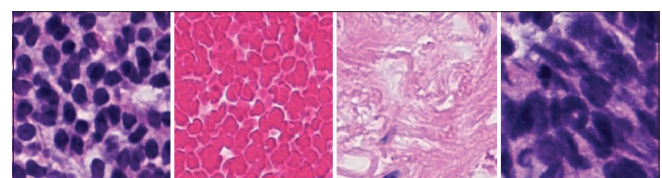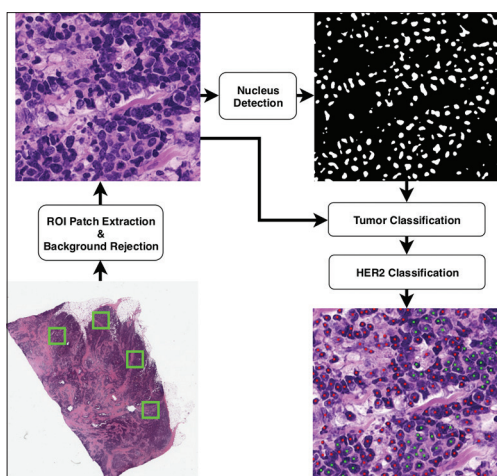| Score	Pattern	Assessment
0	No observable staining, or membrane staining that is incomplete and is faint/barely perceptible in <10% of tumor cells	Negative
1+	Incomplete membrane staining that is faint/barely perceptible in >10% of invasive tumor cells	Negative
2+	Circumferential membrane staining that is incomplete and/or weak/moderate in >10% of invasive tumor cells, or complete and circumferential intense membrane staining in ≤10% of invasive tumor cells	Equivocal
3+	Homogeneous, dark, circumferential (chicken wire) pattern in >10% of invasive tumor cells	Positive

Figure 1: Examples of HER2neu immunohistochemistry staining that shows patches from slides with different HER2 score varying with the staining intensity. HER2: Human epidermal growth factor receptor 2

Figure 2: Examples of patches without tumor from the Warwick training set

Figure 3: Block diagram of the proposed method

Figure 4: A sample annotation of H&E image (right) using the serial immunohistochemistry image (left) included in the training dataset

Training data preparation

Given the limited set of cases in the dataset, our training and testing methods were guided by the need to avoid the contamination by equivocal cases, low quality samples, and histology that can be attributed to rare subtypes of breast cancer. Thus, we excluded the cases in which H&E sections were smudged or covered with slide markers or represented a rare morphology. We also excluded the cases with inadequate tumor regions. WSIs have inter-region variations in HER2 IHC because of various tissue structures as well as intra-tumor heterogeneity (e.g. some regions appear 2+ and some 3+ in HER2+ cases). Therefore, using the Aperio Imagescope[R], we annotated regions that were clearly HER2– or HER2+ on H&E images by identifying their corresponding regions in the IHC-stained serial section. A sample annotated image is shown in Figure 4, where the region marked in green is noncancerous and regions marked in cyan are cancerous. One can notice the care taken to exclude any ambiguous areas.

An exact image registration of the IHC and H&E stained images was not applicable due to the use of two very different stains and different, although serial or adjacent, tissue sections. Hence, we first annotated a few clearly tumorous and some clearly nontumorous regions in H&E WSIs for each case. On an average, there were six matched regions-of-interest of variable sizes in each slide that were annotated for training. The focus of the annotations was to cover the majority of the tumor regions with regional correspondence in the IHC serial section. We then identified the strong HER2+ and HER2– tumor regions within the IHC, and annotated the corresponding regions in H&E by visually matching gland shapes.

Test data preparation

Similar to the selection of training cases, 26 cases from the Warwick dataset were selected for testing, out of which eight were HER2+, and the rest were HER2–. To reduce computational time required for testing gigapixel WSIs we
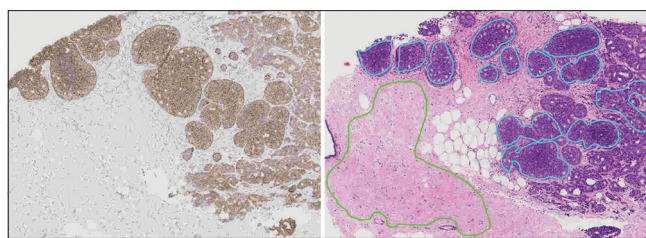
divided these into patches of size 2000 × 2000 and discarded those patches that had <70% tissue area (with grayscale intensity < 220). Thus, we made our method fully automated for test images. Similarly, for the TCGA-BRCA dataset, we selected sections of tumorous regions on the WSIs and extracted 2000 × 2000 patches from the annotated regions of H&E WSIs.

Stain separation and nucleus detection

A lot of confounding classes and intra-class variability encountered by the downstream classification modules can be reduced by merely detecting all nuclei with subsequent sampling of nucleus-centric patches for further processing. This idea also incorporates the insight that nuclear morphology and inter-nuclear relations are important in determining the disease states.[24,25] The first deep learning step of our pipeline determines the locations of nuclei in WSIs. As detecting nuclei requires examining each pixel location (after discarding obvious noncandidates such as background whitespace using a gray-scale intensity threshold), we used a pretrained UNet-based nucleus detector.[26] The UNet was trained on a large dataset of annotated nuclei[15] using stain-separated hematoxylin stain from H&E images.[27,28] The trained U-Net gave satisfactory qualitative results on the WSIs from Warwick and TCGA datasets, as shown in Figure 5.

Tumor versus nontumor classifier

Once all nuclei are detected, we classify each nucleus into a tumor or non-tumor. Most of the non-tumorous nuclei were stromal, and hardly any were from the benign epithelium. We trained a custom CNN with architecture shown in Table 3, that classified fixed sized patches, centered at the locations of the detected nuclei, into tumor and nontumor. The training dataset for tumor versus nontumor classification was prepared using the tumor and nontumor annotations, as shown in Figure 4. The tumor regions contained both HER2+ and HER2– regions. This classifier was key to our pipeline as it screens all irrelevant parts of the tissues.

Human epidermal growth factor receptor 2+ versus human epidermal growth factor receptor 2– classifier for tumorous cells

A third neural network classified the patches centered at tumor nuclei (extracted by the previous stage) into HER2+ and HER2–. Doing so in a cascade of neural networks helped us
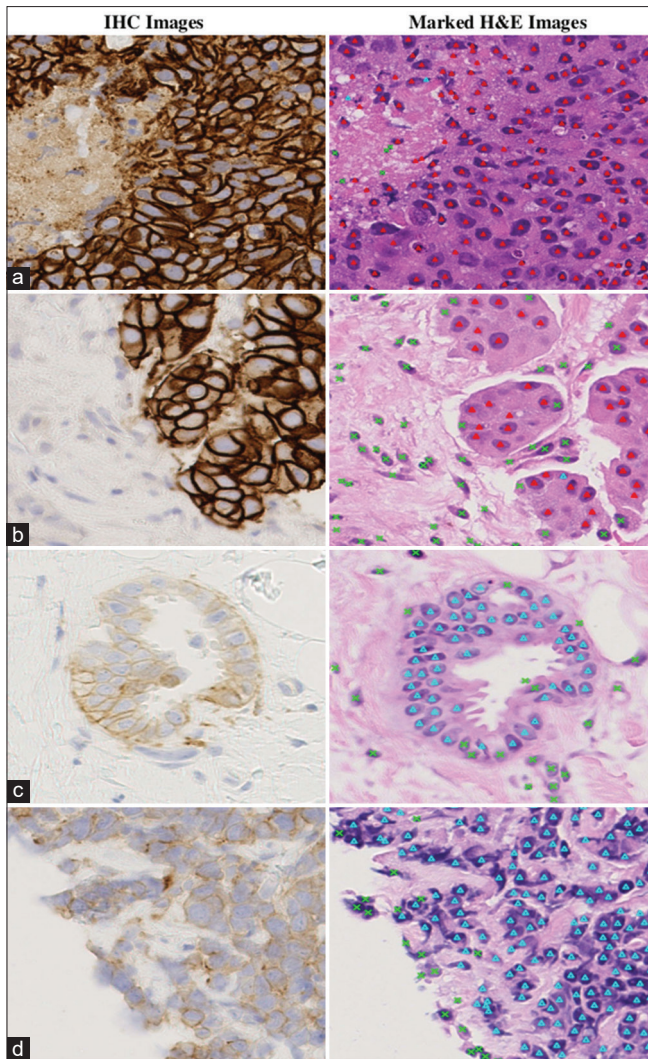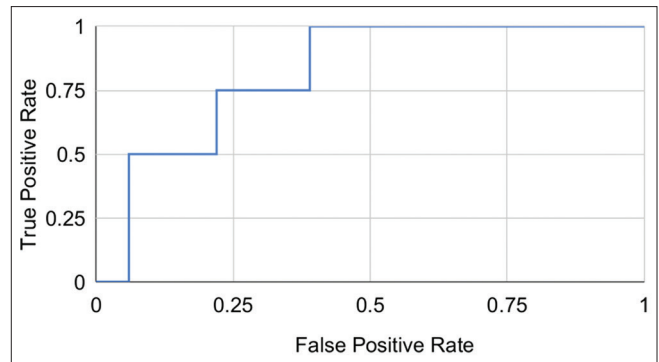
Figure 5: Sample visual results showing spatial correspondence with immunohistochemistry: Δ: HER2+, Δ: HER2− and ×: Noncancerous. (a and b) HER2+ image and its corresponding H&E marked images. (c and d) HER2− images and its corresponding H&E marked images. HER2: Human epidermal growth factor receptor 2

Figure 6: Receiver operating characteristic curve for held-out patients in the Warwick dataset for HER2 + versus HER2− task. HER2: Human epidermal growth factor receptor 2

Table 2: Composition of training and testing datasets

Dataset	Warwick	TCGA-BRCA
Training (number of cases)		
HER2+	11	-
HER2-	15	
Total	26	
Testing (number of cases)		
HER2+	8	23
HER2-	18	22
Total	26	45

TCGA-BRCA: The Cancer Genome Atlas Breast Invasive Carcinoma, HER2: Human epidermal growth factor receptor 2

Table 3: Convolutional neural network architecture for tumor versus nontumor classification

Layer	Filter size	Input layer	Input size	Output size
Input	-	-	100×100×3	-
Conv1a + BN	1×1	Input	100×100×3	100×100×4
Conv1b + BN	3×3	Input	100×100×3	100×100×4
Concat	-	Conv1a, Conv1b	100×100×4	100×100×8
Conv2 + BN + D	3×3	Concat	100×100×8	50×50×16
Conv3 + BN + D	5×5	Conv2	50×50×16	25×25×32
FC1 + BN	1024	Conv3	20,000	1024
FC2	64	FC1	1024	64
FC3	2	FC2	64	2

Conv: Convolution, BN: Batch normalization, ReLU: Rectified linear unit, FC: Fully connected

experiment with different architectures to accomplish each of the sub-goals (nucleus detection, tumor classification, and HER2 classification) in the cascade and derive visual insights into the regions marked by each of the networks. The architecture for the HER2+ versus HER2− classifier is shown in Table 4. The training dataset for this CNN comprised 28,021 HER2+ and 29,296 HER2− patches, each of size 100 × 100 pixels obtained after excluding the nontumorous patches.

RESULTS

We now share the results of the tumor detection and HER2 classification stages.

For training the tumor-detection stage, from the training images in the Warwick dataset, 25,187 patches of size 100 × 100 pixels that were centered at the detected nuclei were used for training the second CNN, of which 12,549 were tumorous nuclei. On 5,000 patches in the test set, of which 2,500 were tumorous nuclei, the CNN achieved 98% classification accuracy.
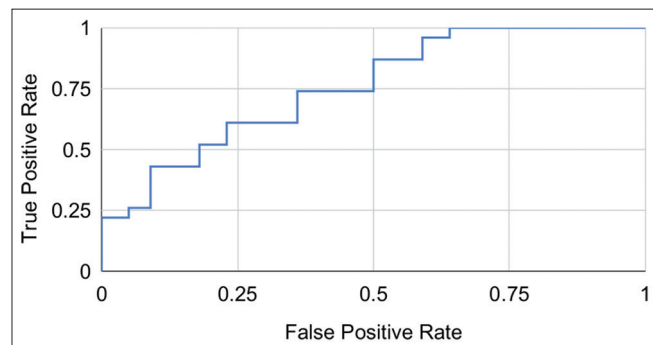
The training of the third CNN to classify tumorous nuclei into HER2+ versus HER2− was done using 28,021 HER2+ and 29,296 HER2− patches mined from the Warwick training dataset. We tested the entire pipeline on two test datasets – the Warwick test dataset comprising 26 held-out cases and on an independent 45 patient TCGA-BRCA dataset.

Human epidermal growth factor receptor 2 overexpression estimation on Warwick dataset

We tested the proposed approach on the WSI by applying all three stages sequentially, i.e. detect nuclei, identify tumorous nuclei, and classify them as HER2+ or HER2–. The total number of tumorous nuclei and pecent of HER2+ nuclei detected for each patient was calculated. To obtain a patient-level HER2+ decision, merely, a threshold for the proportion of HER2+ nuclei was used as there was a large gap between the maximum value for HER2– cases and minimum value for the HER2+ test cases. We achieved a patient-level AUC of 0.82 (CI: 0.65–0.98) on the Warwick dataset, as shown in Figure 6. We achieve sensitivity of 0.75 with specificity of 0.78. We also show the qualitative results on the Warwick test dataset, where IHC serial sections were available, in Figure 5. The Δ symbol denotes HER2+ nuclei, Δ denotes HER2– and × denotes non-tumorous nuclei. The first two rows highlight the efficacy of the algorithm in HER2+ nuclei identification. The bottom two rows show the efficacy to highlight HER2– nuclei. The subtle green marks in these figures identify non-tumorous nuclei in both HER2+ and HER2– cases. Figure 5 shows a direct correspondence between the predicted HER2 overexpression and IHC staining.

Testing human epidermal growth factor receptor 2 overexpression determination on an independent dataset

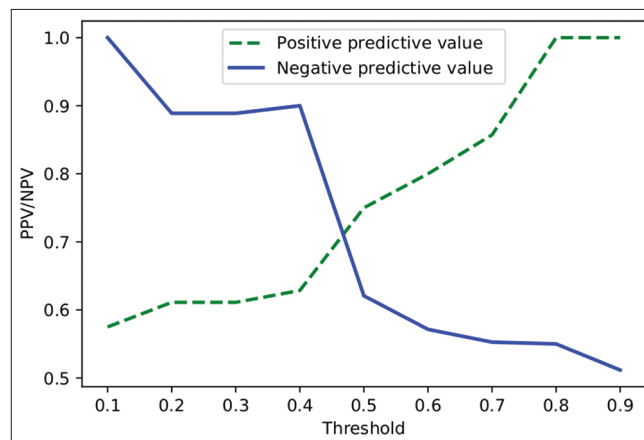We tested our trained model on an independent dataset to examine whether its generalization performance was strong enough to go from a single-center to a multi-center dataset. There were 45 cases (23 HER2+ and 22 HER2–) in the independent test set obtained from the TCGA-BRCA cohort. We achieved a patient AUC of 0.76 (CI: 0.61–0.89) on this dataset, as shown in Figure 7, without using any part of it in any way for training. Positive predictive values and negative predictive values (NPVs) for this dataset are shown in Figure 8. It is evident that the trained model could be used as a screening algorithm for HER2 overexpression detection, as the NPV holds around 0.90 for a threshold value of up to 0.40. The testing achieves sensitivity of 0.87 with specificity of 0.60.

Testing estimation of human epidermal growth factor receptor 2 overexpression on human epidermal growth factor receptor 2+ grade cases

For completeness of the analysis, a total of 25 cases were selected at random having an IHC score of 2+ and available FISH status from TCGA-BRCA cohort. Out of the 25 cases, 11 were FISH positive and 14 were FISH negative. The external validation on this dataset achieved an AUC score of 0.73 (CI: 0.53–0.93) as shown in Figure 9. The wide CI was expected as the model was never trained to predict the FISH status.
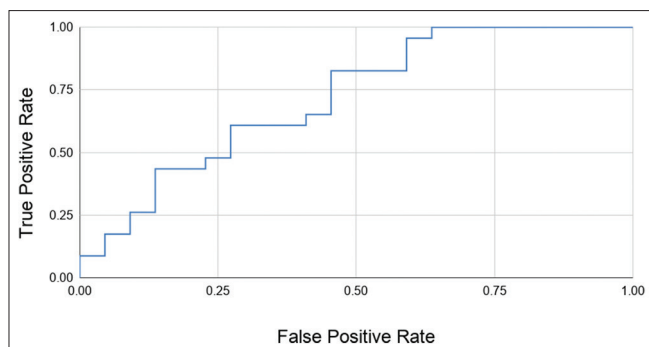

Figure 7: Area under the curve-receiver operating characteristic curve for independent testing dataset TCGA-BRCA

Figure 8: Positive predictive value and negative predictive value curves for independent testing on the TCGA-BRCA dataset

Table 4: Convolutional neural network architecture for human epidermal growth factor receptor 2+ versus human epidermal growth factor receptor 2– classification

Layer	Filter size	Activation	Input layer	Input size	Output size
Input	-	-	-	100×100×3	-
Conv1a + BN	1×1	ReLU	Input	100×100×3	100×100×4
Conv1b + BN	3×3	ReLU	Input	100×100×3	100×100×4
Concat	-	-	Conv1a, Conv1b	100×100×4	100×100×8
Conv2 + BN + Dropout	3×3	ReLU	Concat	100×100×8	50×50×16
Conv3 + BN + Dropout	5×5	ReLU	Conv2	50×50×16	25×25×32
FC 1 + BN	64	ReLU	Conv3	20,000	64
FC 2 + BN	64	ReLU	FC 1	64	64
FC 3	2	Softmax	FC 2	64	2

Conv: Convolution, BN: Batch normalization, ReLU: Rectified linear unit, FC: Fully connected

Figure 9: Area under the curve-receiver operating characteristic curve for testing on human epidermal growth factor receptor 2 2+ cases from TCGA-BRCA cohort

DISCUSSION AND CONCLUSION

Using judiciously prepared training data from serial IHC and H&E sections, and a multi-stage deep learning pipeline, we showed that the estimation of HER2 IHC score is possible from H&E images by quantifying the over-expression of HER2 protein. The trained model not only gives good performance on the held-out cases from the discovery cohort but also works for an independent and multi-center dataset, which is a stronger challenge in the medical community as compared to training on a multi-center dataset and testing on an independent dataset from a single center. We attribute these results to the multi-stage pipeline as opposed to training a single neural network to implicitly perform all tasks. The evaluation of the proposed approach as a potential screening method before HER2 IHC testing was also satisfactory, retaining 0.90 NPV. In future, we plan to expand the dataset to include more representatives of the rare morphologies that were excluded in the training and testing datasets used in this article. This is a successful pilot study displaying the potential of deep-learning methods to present affordable and quality medical care. We will also extend the study to other IHC panels for computational multiplexing. Our work suggests that deep-learning pipelines can be trained on H&E morphology using supervised data from serial IHC sections or even genomic tests. This opens up the possibility of detecting the presence of multiple antigens or sub-clonal populations at different spatial foci to study tumor subtypes and intra-tumor heterogeneity.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Cancer of the Breast (Female) – Cancer Stat Facts. Available from: https://seer.cancer.gov/statfacts/html/breast.html. [Last accessed on 2019 Oct 02].
2. American Cancer Society. Cancer Facts and Statistics. Available from: http://cancerstatisticscent er.cancer.org/. [Last accessed on 2019 Oct 02].
3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, *et al*. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci United States Am 2003;100:8418-23.
4. Dai X, Chen A, Bai Z. Integrative investigation on breast cancer in ER, PR and HER2-defined subgroups using mRNA and miRNA expression profiling. Sci Rep 2014;4:6566.
5. Perou CM, SÃžrlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, *et al*. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
6. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, *et al*. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res 2015;5:2929-43.
7. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr., Davidson NE, *et al*. Trastuzumab plus adjuvant chemotherapy for operable her2-positive breast cancer. N Engl J Med 2005;353:1673-84.
8. Finberg KE, Sequist LV, Joshi VA, Muzikansky A, Miller JM, Han M, *et al*. Mucinous differentiation correlates with absence of EGFR mutation and presence of KRAS mutation in lung adenocarcinomas with bronchioloalveolar features. J Mol Diagn 2007;9:320-6.
9. 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; 2016. p. 770-8.
10. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2009. p. 248-55.
11. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, *et al*. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 1-9.
12. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva WK, Busam V, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Med 2019;25:1301-9.
13. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep Learning for Identifying Metastatic Breast Cancer. arXiv preprint arXiv: 1606.05718; 2016.
14. Grand Challenge. Available from: https://grand-challenge.org/. [Last accessed on 2019 Oct 02].
15. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Transact Med Imaging 2017;36:1550-60.
16. Kumar N, Verma R, Anand D, Zhou Y, Onder OF, Tsougenis E, *et al*, A Multi-Organ Nucleus Segmentation Challenge. IEEE Transactions on Medical Imaging 2020;39:1380-91.
17. Wolff AC, Hammond ME, Allison KH, Harvey BE, Mangu PB, Bartlett JM, *et al*. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists Clinical practice guideline focused update. J Clin Oncol 2018;36:2105-22.
18. Qaiser T, Mukherjee A, Reddy PC, Munugoti SD, Tallam V, Pitkäaho T, *et al*. Her 2 challenge contest: A detailed assessment of automated her 2 scoring algorithms in whole slide images of breast cancer tissues. Histopathology 2018;72:227-38.
19. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, *et al*. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Med 2018;24:1559.
20. Kim RH, Nomikou S, Dawood Z, Coudray N, Weber JS, Shapiro RL, *et al.* Using deep learning algorithms on histopathology images for the prediction of BRAF and NRAS mutations in invasive melanoma. Journal of Clinical Oncology 2018;36e21561-e21561.
21. Kim RH, Nomikou S, Dawood Z, Jour G, Donnelly D, Moran U, *et al*. A Deep Learning Approach for Rapid Mutational Screening in Melanoma. bioRxiv 610311; 2019.
22. Andersen L, Brügmann N, Lelkaitis A, Nielsen G, Friis Lippert S, Vyberg MM. Virtual double staining: A digital approach to immunohistochemical quantification of estrogen receptor protein in breast carcinoma specimens. Applied Immunohistochem Mol Morphol 2018;26:620-6.
23. Rawat RR, Ruderman D, Macklin P, Rimm DL, Agus DB. Correlating nuclear morphometric patterns with estrogen receptor status in breast cancer pathologic specimens. NPJ Breast Cancer 2018;4:32.

24. Verma R, Kumar N, Sethi A, Gann PH. Detecting multiple subtypes of breast cancer in a single patient. In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE; 2016. p. 2648-52.
25. Kumar N, Verma R, Arora A, Kumar A, Gupta S, Sethi A, *et al*. Convolutional neural networks for prostate cancer recurrence prediction. In: Medical Imaging 2017: Digital Pathology. Vol. 10140. International Society for Optics and Photonics; 2017. p. 101400H.
26. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2015. p. 234-41.
27. Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, *et al.* Structure-preserving color normalization and sparse stain separation for histological images. IEEE Transact Med Imaging 2016;35:1962-71.
28. Anand D, Ramakrishnan G, Sethi A. Fast GPU-enabled color normalization for digital pathology. In: 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE; 2019. p. 219-24.