ESPWA: a deep learning-enabled tool for precision-based use of endocrine therapy in resource-limited settings

Dagoberto Pulido-Arias[1]*, Rebecca Henderson[2]*, Christophe Millien[3], Joarly Lomil[3], Marie Djenane Jose[3], Gabriel Flambert[4], Jean Bontemps[4], Emmanuel Georges[4], Alekhya Gunturi[5], Palak Shah[5], Tiago Goncalves[6], Jayashree Kalpathy-Cramer[7], Elizabeth Gerstner[8], Seth Wander[8], S. Joe Sirintrapun[9], Dennis Sgroi[9], Jose Jeronimo[10], Philip E. Castle[10,11], Kenneth Landgraf[12], Ali Brown[12], Temidayo Fadelu[13], Lawrence N. Shulman[14], John Guttag[15], Dan Milner[12,16], Jane Brock[12]**, Christopher Bridge[1]**, Albert Kim[1,8]**

* - These authors contributed equally.
** - These authors jointly supervised this work.

Affiliations
[1]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital
[2] Department of Obstetrics & Gynecology, University of Alabama-Birmingham
[3]Hopital Universitaire de Mirebalais, Zanmi Lasante, Mirebalais, Haiti
[4]University of Florida College of Medicine
[5]Boston University School of Medicine
[6]Institute of Systems and Computer Engineering, Technology and Science, Faculdade de Engenharia, Universidade do Porto
[7]Department of Ophthalmology, University of Colorado-Anschutz
[8]Massachusetts General Hospital Cancer Center
[9]Department of Pathology, Massachusetts General Hospital
[10]Division of Cancer Prevention, National Cancer Institute, NIH
[11]Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute, NIH
[12]American Society for Clinical Pathology, Center for Global Health
[13]Center for Global Health Equity, Dana-Farber Cancer Institute
[14]Abramson Cancer Center, University of Pennsylvania
[15]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
[16]Department of Pathology, Olive View-UCLA Medical Center

Acknowledgments of Research Support
This work was supported by the intramural research program (ZIACP101237-01) of the US National Cancer Institute/NIH/DHHS. AK has received support by the William G. Kaelin, Jr., M.D., Physician-Scientist Award of the Damon Runyon Cancer Research Foundation (PST-36-21), American Association for Cancer Research Breast Cancer Research Fellowship (21-40-49-KIM), American Brain Tumor Association Basic Research Fellowship In Honor of Paul Fabbri, American Society of Clinical Oncology/Conquer Cancer Young Investigator Award, American Cancer Society Institutional Research Grant (IRG-21-130-10), a National Cancer Institute K12CA090354, and a National Institute of Biomedical Imaging and Bioengineering K08EB037077.

Corresponding authors
Albert Kim, M.D.

Simches Research Building
185 Cambridge Street, Office 3.828
Boston, MA 02114
akim46@mgh.harvard.edu

Christopher Bridge, D.Phil.
149 13th Street, Office 2303
Boston, MA 02129
cbridge@mgh.harvard.edu

Abstract: 302 words
Introduction: 677 words
Methods: 643 words
Results: 746 words
Discussion: 961 words
Total (without Abstract): 3027 words
Figures & Tables: 5
Supplemental Figures & Tables: 6

ABSTRACT

Background: Cancer morbidity disproportionately affects patients in low- and middle-income countries (LMICs), where timely and accurate tumor profiling is often nonexistent. Immunohistochemistry-based assessment of estrogen receptor (ER) status, a critical step to guide use of endocrine therapy (ET) in breast cancer, is often delayed or unavailable. As a result, ET is often prescribed empirically, leading to ineffective and toxic treatment for ER-negative patients. To address this unmet need, we developed **ESPWA** (**E**strogen Receptor **S**tatus **P**rediction for Haitian patients using deep learning-enabled histopathology **W**hole Slide Imaging **A**nalysis), a deep-learning (DL) model that predicts ER status directly from hematoxylin-and-eosin (H&E)-stained whole slide images (WSIs).

Methods: We curated two cohorts of H&E WSIs with tissue-matched ER status: The Cancer Genome Atlas (TCGA, n = 1085) and Zanmi Lasante (ZL, n = 3448) from Haiti. We trained two models using weakly supervised attention-based multiple instance learning: a "TCGA" model, trained on TCGA data, and ESPWA, trained on the ZL dataset. Model performance was evaluated using 10-fold cross validation.

Results: Performance of the "TCGA" model was sensitive to the domain shift between the TCGA and ZL datasets, with a performance of an area under receiver operating characteristic (AUROC) of 0.846 on the TCGA test sets and 0.671 on the ZL test sets. Compared to the "TCGA" model, ESPWA demonstrated improved performance on the ZL cohort (AUROC=0.790; p=0.005). Subgroup analyses revealed clinically relevant populations in which ESPWA demonstrated improved performance relative to the overall cohort. Finally, ESPWA outperformed an academic breast pathologist (accuracy: 0.726 vs 0.639 respectively; $p < 0.001$) in determining ER status from H&E WSIs.

Conclusion: ESPWA ("Hope" in Haitian Creole) offers an accessible framework to identify individualized therapeutic insights from H&E WSIs in LMICs. We have initiated clinical trials, using ESPWA, in ZL and sub-Saharan African countries to inform precision-based use of ET for prospective patients.

INTRODUCTION

Breast cancer is the most common cancer type and leading cause of cancer mortality for females worldwide[1]. This mortality disproportionately affects patients in LMICs, rather than those in high-income countries (HICs). Despite recent advances in hormonal and targeted therapy that render early-stage breast cancer increasingly curable[2], patients in LMICs continue to experience delayed diagnoses and suboptimal tumor-directed therapy[3]. In Haiti, one of the most resource-limited settings in the world, the challenge is especially dire. As only five pathologists serve the entire country of nearly 12 million people, timely and complete pathologic assessment of tumor tissue is nearly impossible. A delayed diagnosis of cancer leads to a higher likelihood of the patient presenting in the metastatic stage, which in turn decreases life expectancy. The mortality-to-incidence ratio for Haitian breast cancer patients is over 60%[4], compared to just 2.3% in the United States[5]. This stark disparity underscores an urgent need for innovative methods that enable delivery of high-quality cancer care in low-resource settings.

In clinical workflows, pathologists perform a manual analysis of H&E-stained slides, which are widely available in HICs and LMICs. However, as a typical H&E slide contains hundreds of thousands of cells, it is impractical for a human to ingest, analyze, and report this vast amount of data expeditiously. To this end, DL applications in digital pathology have emerged as powerful tools to enhance the efficiency and yield of pathology workflows. Furthermore, recent studies have demonstrated DL's potential to identify subtle spatial patterns of cell populations[6,7] and individualized therapeutic insights, thus highlighting its potential for precision oncology.

Given efficacy observed with ETs for ER-positive breast cancer[8], a timely assessment of ER status is arguably the most critical element of the diagnostic-therapeutic pathway for breast

cancer. However, immunohistochemistry (IHC), the gold standard for measuring ER expression, is often unavailable in LMICs and dependent on donated services from academic medical centers and medical equipment companies in HICs, thus resulting in delays on the order of months. Therefore, breast cancer patients in Haiti and other LMICs are often empirically treated with ET[9], which subjects ER-negative patients to ineffective therapy and unnecessary toxicity[10].

To address this clinical need, we developed **ESPWA** (**E**strogen Receptor **S**tatus **P**rediction for Haitian patients using deep learning-enabled H&E **W**hole Slide Imaging **A**nalysis), a DL model that infers quantitative ER expression from H&E-stained WSIs. Although ER status is not routinely assessed from H&E slides in clinical practice, prior DL models have demonstrated the feasibility of inferring ER expression from H&E WSIs[11–13]. However, these models have been trained on publicly available datasets of H&E WSIs primarily sourced from HICs. Given intrinsic differences between HIC and LMIC datasets, such models may not generalize well to the LMIC population. For example, there are notable differences in tissue quality and fixation protocols between HICs and LMICs[9,14]. In addition, publicly available datasets such as The Cancer Genome Atlas (TCGA) are largely derived from patients of Caucasian descent[15], which differs from the predominantly Black population in Haiti and other LMICs. These differences are relevant for model generalizability, given a higher incidence of aggressive histological features and poorly differentiated tumors in the Black population compared to Caucasians[16–18].

In this proof-of-concept study, we demonstrate the utility of DL to enhance cancer care in resource-limited settings and to identify individualized therapeutic insights using widely available H&E WSIs. To develop ESPWA ("Hope" in Haitian Creole), we curated a dataset of H&E WSIs and tissue-matched ER status from breast cancer patients treated at Zanmi Lasante

(ZL), the largest provider of cancer care in Haiti. All tissue samples were fixed and embedded locally, ensuring that our dataset reflects the variability in tissue quality observed in LMICs. Here, we demonstrate that ESPWA achieves robust performance in assessing ER status across tissue types, tumor grades, and histological subtypes. To build upon this work, we have initiated prospective clinical trials in ZL and partner sites across Africa to evaluate the utility of ESPWA in informing precision-based use of ET in resource-limited settings. More broadly, this work serves as a scalable framework for employing DL to interrogate tumor biology across large populations.

METHODS

Dataset curation

We curated two different cohorts of H&E whole slide images (WSIs) and tissue-matched ER status. One cohort was from The Cancer Genome Atlas – Breast Invasive Carcinoma (TCGA), which comprised of 1085 H&E WSIs from 1073 patients. Each WSI had tissue-matched ER status derived from IHC.

Our second cohort included 3448 H&E WSIs from 1808 breast cancer patients from Hôpital Universitaire de Mirebalais, the teaching hospital of Zanmi Lasante (ZL) in Mirebalais, Haiti. Tissue blocks were created in ZL. Sectioning and staining with H&E were performed at the Brigham & Women's Hospital (BWH). ER status for all ZL cases was determined with IHC at BWH. This IHC-derived ER status served as the ground truth label for the ZL dataset. All H&Es were digitized using the KFBIO KF-PRO-005 scanner at 40x magnification.

WSI Processing

For both the TCGA and ZL datasets, quality control and tissue segmentation were performed as per prior literature[19,20]. In addition, we performed label-preserving data augmentation to extracted image patches (**Figure 1A**). Further details are provided in the *Supplementary Methods.*

Model training

We trained two models to predict ER status from H&E WSIs: one using the TCGA dataset for training, and the other using ZL for training (**Figure 1B**). Both models used a random data split of 70% for training, 15% for validation, and 15% for testing, ensuring that each patient's data was assigned to only one partition. We assessed the impact of domain shift across

the TCGA and ZL datasets by evaluating the TCGA model on the ZL dataset. For both models, we performed 10-fold cross-validation.

Both the TCGA model and ESPWA shared an architecture consisting of (1) a frozen, pre-trained image encoder using the publicly available weights from the CONtrastive learning from Captions for Histopathology (CONCH) model[20] that converts high-resolution image patches into a 768-dimensional feature vector; (2) a trainable multiple instance learning (MIL) attention aggregator that integrates all encoded patches from a slide into a global slide-level feature representation following the clustering constrained attention multiple instance learning (CLAM) architecture[19]; (3) a trainable fully-connected classifier head with a sigmoid activation that outputs the probability of a patient's ER status based on the slide-level embedding. The TCGA model and ESPWA quantified ER expression using binary classification: ER-negative or ER-positive. As the ZL cohort also had quantitative ER expression scores, we also trained a separate regression model to assess a quantitative measure of ER expression from the H&E WSI.

Further details about model architecture, hyperparameters, and hardware used for model training are included in *Supplementary Methods*.

Analysis of model performance

For binary classification, we evaluated model performance by calculating accuracy, sensitivity (also called recall), positive predictive value (also called precision), area under the receiver operating characteristic curve (AUROC), and the area under precision-recall curve (AUPRC). For regression, we evaluated model performance with the Pearson's correlation coefficient (PCC) and intra-class correlation coefficient (ICC). All performance metrics were calculated using 10-fold cross validation. To estimate variability, 95% confidence intervals were

computed via bootstrapping with 1000 re-samples across the test folds. Performances across different models and subgroups were compared using the DeLong test.

We also conducted a decision curve analysis following a previously published framework[21–23] to evaluate the benefit of the model to support clinical decision making under a range of different preferences. Since the current standard practice in Haiti is to treat all patients with ET, we use the "interventions avoided" variant that measures net benefit in terms of a balance between the net decrease in false positives and net increase in false negatives compared to the "treat all" strategy that considers all patients as ER-positive.

Interpretation of model predictions via attention heatmaps

To interpret the relative importance of different regions in a WSI to the final WSI-level prediction of the model, we created heatmaps with colors corresponding to attention weights of each patch token used to calculate the slide-level embedding[19].

RESULTS

Substantial domain shift between TCGA and ZL datasets

Consistent with previous studies[11,13], our TCGA-trained model demonstrated strong performance in predicting ER status using H&E WSIs from a predominantly Caucasian population. Our TCGA model achieved an accuracy of 0.850 (95% CI: 0.827-0.879) and an AUROC of 0.846 (95% CI: 0.831-0.879) on the held-aside TCGA test sets (**Figure 2**). Next, to evaluate the ability of our TCGA model to generalize to Haitian patients, we ran inference using the ZL dataset. We observed a marked decrease in performance, with an accuracy of 0.600 (95% CI: 0.590-0.611) and an AUROC of 0.671 (95% CI: 0.662-0.675, **Figure 2**).

ESPWA achieves robust performance in ER-status prediction

For binary classification, ESPWA (accuracy: 0.726; 95% CI: 0.723-0.743, AUROC: 0.790; 95% CI: 0.788-0.807) demonstrated improved performance (accuracy: p=0.03; AUROC: p=0.005) compared to the TCGA model in assessing ER status from the ZL cohort (**Figure 2**). ESPWA's regression model also achieved reasonable performance in predicting quantitative expression of ER, with a PCC of 0.706 (95% CI: 0.660-0.724) and an ICC of 0.524 (95% CI: 0.471-0.551, **Supplemental Figure 1**). To identify subregions of high value within a WSI that ESPWA relied on for classification, we generated exemplar attention heatmaps for both well differentiated and poorly differentiated tumors (**Figure 3A-F**). We observed a high degree of concordance between the high-attention areas highlighted by ESPWA and the tumor regions of interest by an expert pathologist. Furthermore, our decision curve analysis illustrates that ESPWA shows higher net benefit than the current "treat all" paradigm and an alternative "treat

none" approach across a wide range of clinician preferences, demonstrating its potential for use in diverse settings (**Figure 3G; Supplemental Figure 2, Supplemental Table 1**).

ESPWA's performance is noteworthy, given the high prevalence of poorly differentiated tumors in the Black population[16]. Accurate ER assessment based on H&E-stained slides of poorly differentiated tumors, due to loss of typical indicators of ER-positive status (e.g., tubule architecture, scant mitoses, low nuclear grade), is a difficult task for pathologists. In the ZL cohort, 65.3% (2254/3448) of tumors were poorly differentiated (**Table 1**), compared to a historical prevalence of 29.7% in the White population. This disparity highlights the increased difficulty of ER classification in the ZL population relative to the TCGA cohort.

Subgroup analyses

Next, we conducted an analysis across clinically relevant subgroups within our cohort (**Table 1**). There was no difference in ESPWA's ability to assess ER status between the ER-positive (0.723, 95% CI: 0.708-0.732) and ER-negative subgroups (0.729, 95% CI: 0.721-0.746, p=0.705). Within the ER-positive cohort, there was a marked difference in ESPWA accuracy between the ER-low positive (1-10% by IHC; 0.396, 95% CI: 0.303-0.467), and ER-high positive cohorts (>10%; 0.768, 95% CI: 0.751-0.776, p<0.001). This discrepancy accurately reflects tumor biology, as ER-low positive tumors behave more similarly to ER-negative tumors, which are resistant to ET.

Next, we compared ESPWA's performance between well-differentiated (grade 1-2) and poorly differentiated (grade 3) tumors. While the AUROCs were comparable between these two groups, the AUPRC was higher for grade 1-2 tumors (0.816, 95% CI: 0.785-0.846) than for grade 3 tumors (0.703, 95% CI: 0.675-0.738, **Supplemental Figure 3**). This reflects the higher

prevalence of ER-positive tumors in the grade 1-2 cohort (749/1194 = 62.7%) compared to the grade 3 cohort (935/2254 = 41.4%), leading to a higher baseline precision value. ESPWA's observed AUPRC for grade 3 tumors reflects reasonable performance above random chance in a subgroup that is challenging for human experts.

Next, ESPWA's performance was higher on primary tumor samples (AUROC = 0.802; 95% CI: 0.793-0.822), compared to lymph node metastases (AUROC = 0.707, 95% CI: 0.659-0.733, p=0.005). This phenomenon may be due to the presence of dense lymphocytic infiltrates in lymph nodes, which are associated with ER-negative status in primary tumors but not in lymph node metastases. Finally, we did not observe substantial variation in performance across different histological subtypes. Confusion matrices for all subgroup analyses (**Supplemental Figure 4**) show no major imbalance between false positives and false negatives across any subgroup.

ESPWA outperforms a human expert in ER-status prediction

To benchmark model performance against human expertise, we compared ESPWA's classification accuracy to that of an academic breast pathologist with 22 years of experience (J.E.B.). The human expert evaluated ER status based on tumor morphology on H&E WSIs from the ZL cohort and obtained an accuracy of 0.639 (95% CI: 0.619-0.652). By comparison, ESPWA achieved a superior accuracy of 0.726 (95% CI: 0.723-0.743, p<0.001) across the entire cohort (**Figure 4, Supplemental Table 2**) and across most subgroups.

DISCUSSION

DL, through its ability to execute tasks involving high-dimensional data, is poised to transform healthcare delivery in resource-limited settings. In recent years, DL has enabled the development of tools that analyze clinical data to provide a timely diagnosis across diverse diseases[24–26] in LMICs. However, to date, few of these tools have extended beyond diagnostic applications to inform *therapeutic* decision making. While diagnostic tools are an essential part of clinical care, tools that interrogate an individual patient's disease biology to inform selection of the *optimal* treatment are equally as critical to improve cancer outcomes.

As a proof-of-concept, we developed ESPWA, a DL model that uses widely available H&E WSIs to guide precision-based use of ET in LMIC breast cancer patients. By recognizing spatial patterns of therapeutically relevant cell populations within the tumor microenvironment, ESPWA achieved strong performance in ER status prediction across diverse clinical and histologic subgroups. Our model performed well in poorly differentiated tumors – a clinically noteworthy finding given the high prevalence of such tumors in LMICs with predominantly Black populations[9]. Moreover, ESPWA outperformed an academic breast pathologist in assessing ER status directly from H&E WSIs, suggesting that DL may uncover biologically meaningful patterns on H&E WSIs that are not readily discernable by pathologists. Therefore, in settings where IHC is inaccessible and empiric ET is the standard of care, ESPWA offers a precision-based approach that could substantially improve cancer outcomes in resource-constrained settings.

Given the limited infrastructure in LMICs, deployed tools must not only be clinically impactful but also be cost-effective and technologically appropriate. The American Society for Clinical Pathology (ASCP) has led initiatives in twenty LMIC countries across Africa, Asia, and the Caribbean to improve laboratory quality for H&E-staining, which includes capacity for slide

digitization and secure file transfer. At ASCP-supported sites equipped with digital scanning, ESPWA could be deployed using a cloud server with minimal additional resources. Based on this framework, we have initiated prospective feasibility clinical trials to deploy ESPWA to ZL and other ASCP-supported LMICs. These trials will generate additional data to validate and refine ESPWA. Furthermore, we will expand our model's capabilities to predict progesterone receptor (PR) status and HER2 status from H&E WSIs. These molecular markers are therapeutically relevant, as ER-positive, HER2-positive patients typically respond less favorably to ET compared to ER-positive, HER2-negative patients. Incorporating PR status may also enhance model performance through regularization, given the frequent co-expression of ER and PR.

A distinctive strength of our study is the use of H&E WSIs derived from tissue blocks prepared in Haiti. Here, we illustrate that models trained using publicly available H&E WSIs do not generalize to the ZL dataset. This phenomenon is consistent with known differences in breast cancer biology[16,17] and tissue processing quality between HICs and LMICs[14,27], thus underscoring the need for domain-specific training datasets. Further investigation into exact causes of this domain shift is warranted and would inform curation of datasets that enable training of generalizable models across diverse contexts. Nonetheless, ESPWA and our dataset of H&E WSIs from ZL are a valuable foundation for future efforts aimed at building deployable tools that extract therapeutic insights from H&E-stained slides in the LMIC setting.

Our study raises several interesting follow-up questions. First, while tissue blocks from the ZL dataset were prepared in Haiti, sectioning of these blocks and H&E-staining were performed at BWH. In our upcoming prospective trials, we will evaluate whether ESPWA generalizes to H&E slides stained locally in Haiti. Second, the impact of different pathology

slide scanners on model generalizability remains unclear. Future work will evaluate whether digitizing the same slide with different slide scanners would alter ESPWA's predictions. Another interesting area of study is whether image augmentation or domain generalization during model training may overcome biological or technical causes of domain shift. These studies could have broad implications for developing generalizable models in digital pathology for other therapeutic applications.

We also acknowledge that ESPWA demonstrated reduced performance in certain subgroups, particularly the ER-low positive cohort. This observation was not unexpected, given the inherent difficulty in accurately distinguishing ER-low positive from ER-negative tumors. However, since the primary objective of ESPWA was to identify patients most likely to benefit from ET, misclassifying ER-low positive tumors as ER-negative tumors may not be detrimental. Clinically, ER-low positive tumors are less likely to derive a sustained benefit to ETs[28,29]. Additionally, we observed that ER classification of lymph node metastases was less accurate compared to that of the overall dataset. Notably, the inclusion of lymph node metastases in our training set is a strength, as most prior DL models in breast cancer have been trained exclusively on WSIs of primary tumors. Given the distinct histological differences between lymph node metastases (e.g., dense lymphocyte-rich background) and primary tumors (e.g., epithelial or glandular tissue infiltrating fibrous or fatty stroma), the morphological features predictive of ER status likely vary between these tissue types. Additional training data for these subgroups may improve model performance.

In summary, we present ESPWA, a DL model specifically designed for LMIC contexts to enable precision-based administration of ETs in breast cancer patients. ESPWA achieved strong predictive performance across clinically relevant subgroups and outperformed an expert

pathologist in assessing ER status directly from H&E WSIs. By leveraging existing infrastructure in LMICs to enable *individualized* treatment decisions, ESPWA is a technologically appropriate and scalable solution that markedly improves upon the current standard of empiric ET. Most importantly, our findings, combined with prior work[6,14,30,31], illustrate the promise of DL to both advance precision oncology in resource-constrained environments and reveal fundamental insights in cancer biology. Using ESPWA's framework, we are launching prospective trials to validate and further refine ESPWA. Finally, we hope our work will inspire further use of DL-enabled digital pathology analysis to identify under-appreciated spatial patterns of cell populations that represent actionable therapeutic insights for precision oncology.

Figure 1: Experimental design. A) The model input was an H&E WSI, and the ground truth was defined by tissue-matched IHC-derived ER expression. Following tissue segmentation on the H&E WSI, image patches are extracted from the tissue regions of the WSI. Label-preserving data augmentation was applied to these image patches. Using a pre-trained vision-language foundation model trained on H&E WSIs and image captions, these image patches were encoded into feature vectors. We then trained an attention network to aggregate patch-level representations into slide-level representations, which was used to make the final prediction for ER status.

B) To assess for domain shift between the TCGA and ZL datasets, we first trained a model using the TCGA dataset and evaluated this model using 10-fold cross-validation on the TCGA dataset. This TCGA-trained model was then evaluated on the ZL dataset. Finally, ESPWA was trained on the ZL dataset, and performance was measured using 10-fold cross-validation. The TCGA-trained model used binary classification for ER status prediction. ESPWA used both binary classification and regression for ER status prediction.

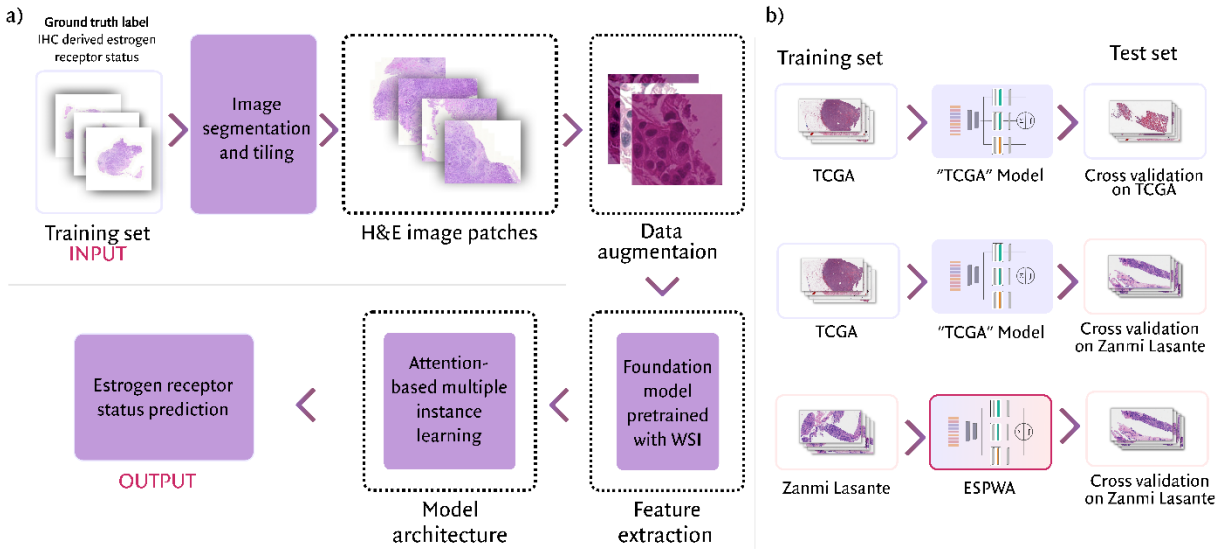
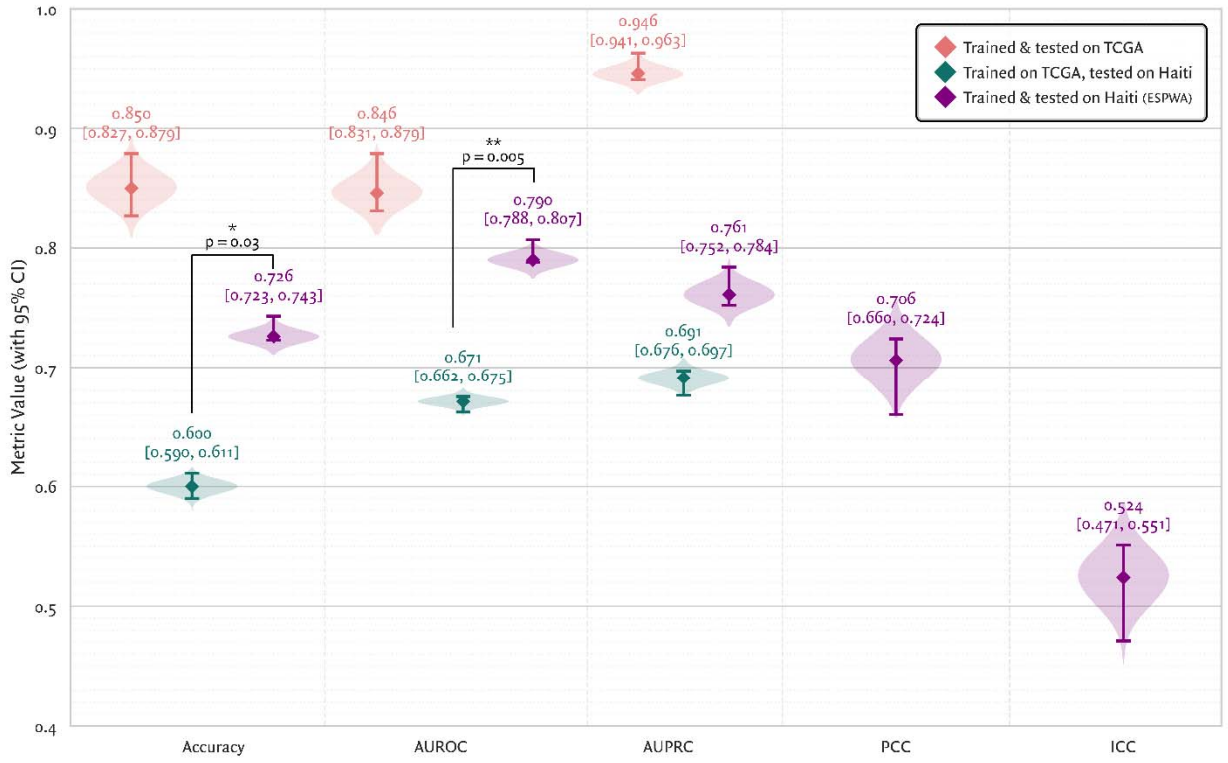


Figure 2: Domain shift between TCGA and ZL datasets. We present 10-fold cross validation prediction results with 95% confidence intervals for the experiments noted in Figure 1B. Performance metrics for the TCGA-trained model on the TCGA dataset are presented in red. Metrics for the TCGA-trained model on the ZL dataset are presented in green. Metrics for ESPWA on the ZL dataset are presented in purple. ESPWA demonstrates improved performance over the TCGA-trained model (accuracy: p=0.03; AUROC: p=0.005) when applied to the ZL dataset, suggesting that there are intrinsic differences between the TCGA and ZL datasets. As the ZL dataset also had tissue-matched quantitative expression of ER status, we trained a regression model with the ZL dataset.



Subgroup	N. samples	ER+	ER-	Accuracy	AUROC	AUPRC	PCC	ICC
Overall								
All Samples	3,448	1684	1764	0.726 (0.723-0.743)	0.790 (0.788-0.807)	0.761 (0.752-0.784)	0.706 (0.660-0.724)	0.524 (0.471-0.551)
ER Status								
ER-positive	1,684	1684	0	0.723 (0.708-0.732)	—	—	—	—
ER-negative	1,764	0	1764	0.729 (0.721-0.746)	—	—	—	—
ER quantitative expression								
ER-Low expression (0-10%)	144	144	0	0.396 (0.303-0.467)	—	—	—	—
ER-High expression (>10%)	1540	1540	0	0.768 (0.751-0.776)	—	—	—	—
Tumor grade								
Grade 1-2 (Well-differentiated)	1,194	749	445	0.747 (0.719-0.765)	0.775 (0.744-0.801)	0.816 (0.785-0.846)	0.608 (0.533-0.625)	0.423 (0.342-0.444)
Grade 3 (Poorly - differentiated)	2,254	935	1319	0.715 (0.697-0.737)	0.776 (0.762-0.802)	0.703 (0.675-0.738)	0.728 (0.714-0.761)	0.546 (0.520-0.588)
Tissue type								
Lymph node	515	265	250	0.676 (0.645-0.694)	0.707 (0.659-0.733)	0.654 (0.625-0.697)	0.619 (0.581-0.730)	0.437 (0.388-0.563)
Primary tumor	2,933	1419	1514	0.735 (0.726-0.753)	0.802 (0.793-0.822)	0.779 (0.764-0.803)	0.721 (0.692-0.738)	0.539 (0.504-0.564)
Histological subtype								
Ductal	2,102	1295	807	0.710 (0.695-0.732)	0.770 (0.750-0.785)	0.831 (0.816-0.844)	0.714 (0.666-0.743)	0.530 (0.474-0.564)
Lobular	138	101	37	0.681 (0.639-0.725)	0.712 (0.666-0.769)	0.860 (0.825-0.913)	0.670 (0.540-0.804)	0.485 (0.367-0.648)
Other	1,208	288	920	0.760 (0.742-0.773)	0.806 (0.784-0.820)	0.519 (0.484-0.556)	0.685 (0.601-0.708)	0.511 (0.416-0.535)

Table 1. ESPWA Performance Metrics Across Clinical and Histologic Subgroups Across the ZL dataset

Figure 3: Model introspection. A-F) Exemplar attention heatmaps for H&E WSIs. Red, orange and yellow signify high-attention areas of the H&E image, green, blue and purple signify low-attention areas. G) Decision curve analysis as per Vickers et al[21] ("interventions avoided" variant)

A) This is a well-differentiated ER-positive tumor case (excision specimen), which ESPWA correctly predicted. The upper panel shows gland formation, but this area to a human eye looks like poorly preserved benign glands with nuclear atypia, and part of a fibroadenomatoid background. The lower panel shows distinct nests of infiltrating tumor cells with low-grade nuclei, consistent with a low-grade, well-differentiated tumor (tubule score 3, nuclear score 1, mitoses score 1 in the Bloom-Richardson grading system).

B) This is a well-differentiated ER-negative tumor case (biopsy specimen), which ESPWA correctly predicted. In the upper panel, areas of high-attention correlate with tubules (red) in a dense stroma of inflammatory cells (orange and yellow). It is not obvious to the human eye that this is infiltrative tumor. By contrast the lower panel shows a distinct infiltrating tumor with lobular features. There are no tubules, the architecture is single file spread, nuclei are small, and mitoses are absent (tubule score 3, nuclear score 1, mitoses score 1 in the Bloom-Richardson grading system). Normally this low-grade histomorphology would suggest ER-positive status, but in the setting of ER-negative status, this is best classified as a histiocytoid invasive lobular carcinoma.

C) This is a poorly differentiated ER-positive tumor case (biopsy specimen), which ESPWA correctly predicted. The upper panel high-attention areas illustrate a solid sheet of tumor cells showing pleomorphic nuclei and frequent mitoses (tubule score 3, nuclear score 3, mitoses score

3 in the Bloom-Richardson grading system). The lower panel of high attention shows necrotic tumor, suggesting high cell turnover.

D) This is a poorly differentiated ER-negative tumor case (biopsy specimen), which ESPWA correctly predicted. High-attention areas in the upper and lower panels depict infiltrating solid nests of tumor cells showing pleomorphic nuclei and frequent mitoses (tubule score 3, nuclear score 3, mitoses score 3 in the Bloom-Richardson grading system).

E) This is an ER-positive lymph node metastasis (excision specimen), which ESPWA correctly predicted. High-attention areas show tumor cells with uniform, low-grade nuclei and scant mitoses, consistent with low-grade tumor. The lymphocytic background is mostly low attention (green).

F) This is an ER-negative lymph node metastasis (excision specimen), which ESPWA correctly predicted. High-attention areas illustrate tumor cells; the upper panel shows pleomorphic nuclei and numerous mitoses, consistent with high-grade tumor. The lower panel highlights pleomorphic tumor cells with necrotic tumor.

G) The x-axis shows the model threshold used, derived from the relative preference of avoiding unnecessary treatments for those who will not benefit compared to incorrectly withholding treatment from those who would benefit: on the left, a low threshold is applied to the model output because correctly treating all those who can benefit is considered more important, on the right a high threshold is used because avoiding unnecessary treatment is considered more important. The y-axis shows the net benefit relative to the “treat all” strategy, defined as the number of unnecessary treatments correctly avoided minus the number of treatments incorrectly withheld, weighted by the implied preference ratio. ESPWA show higher net benefit than either

the “treat all” or “treat none” strategies across a wide range of preferences, demonstrating its potential for impact in wide-ranging clinical scenarios.

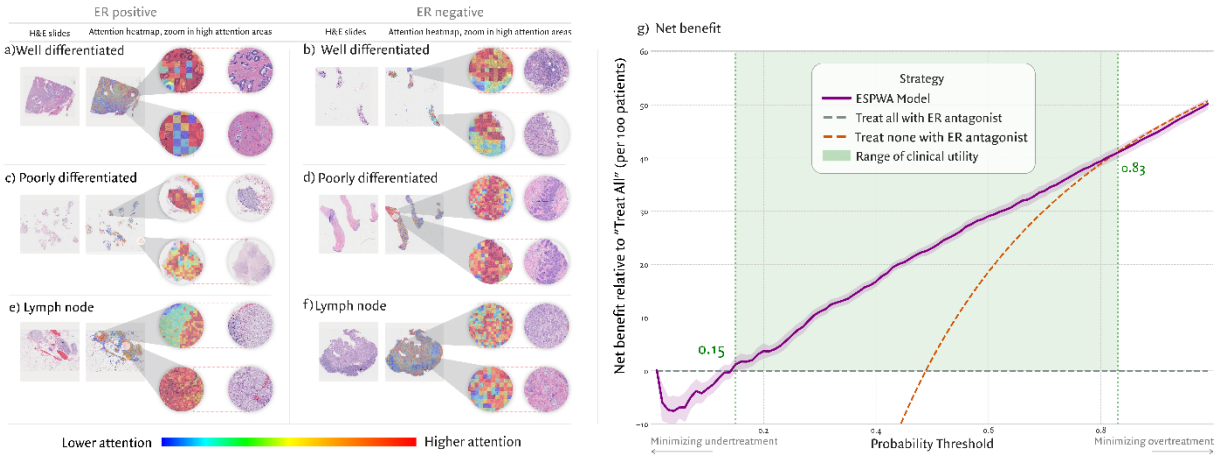
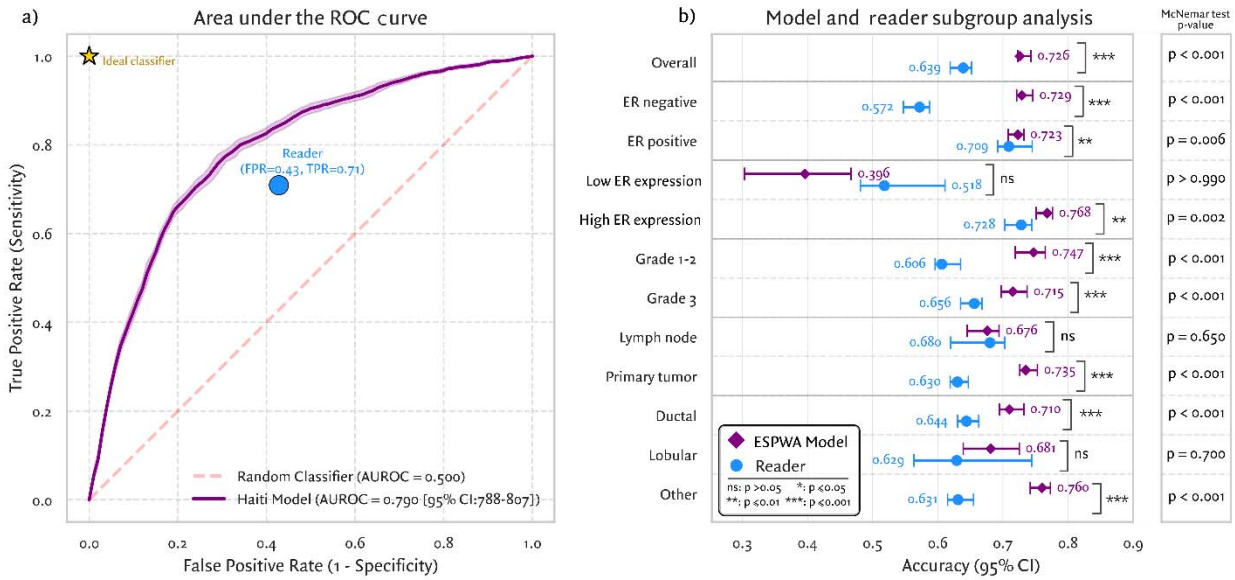
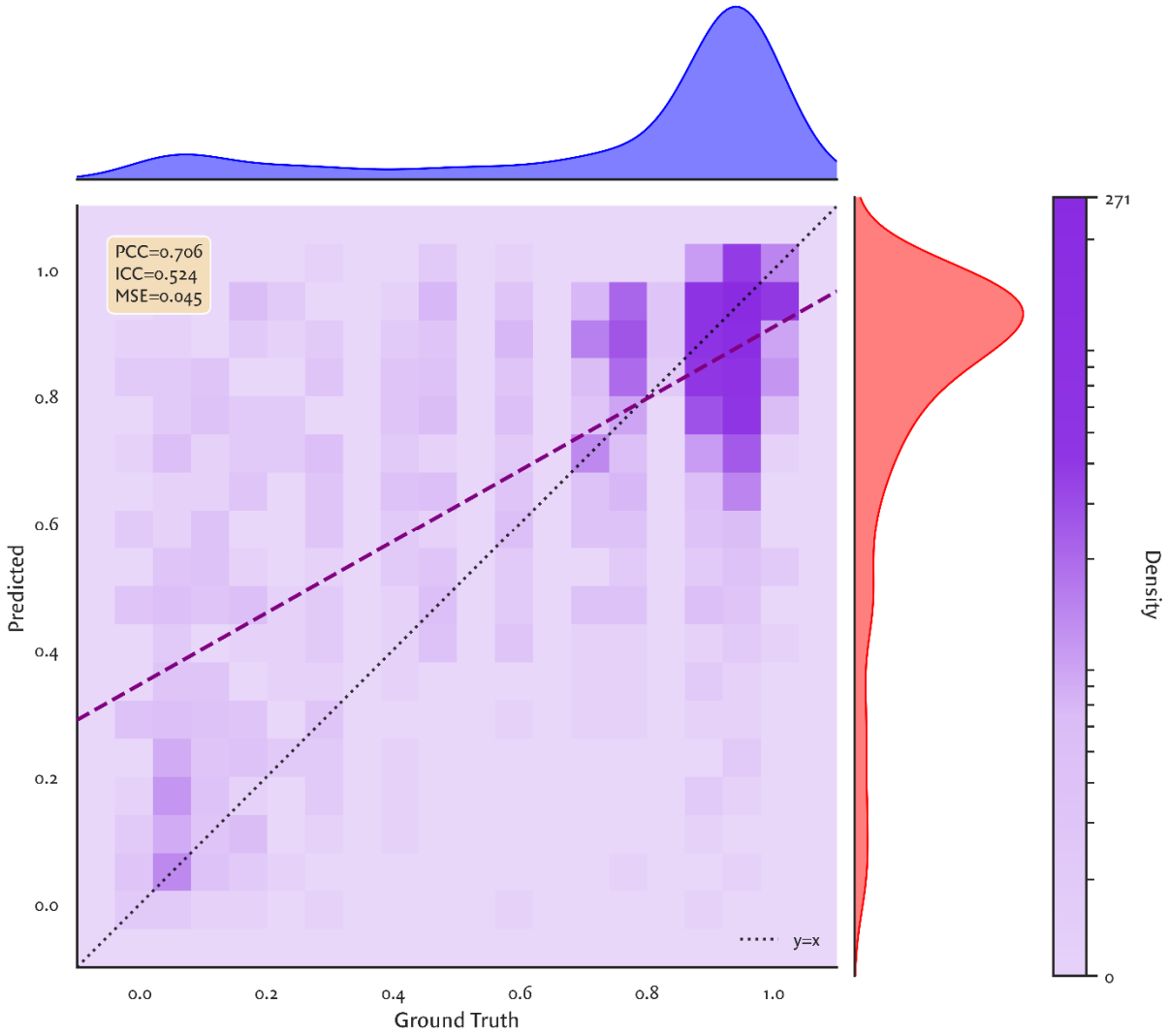


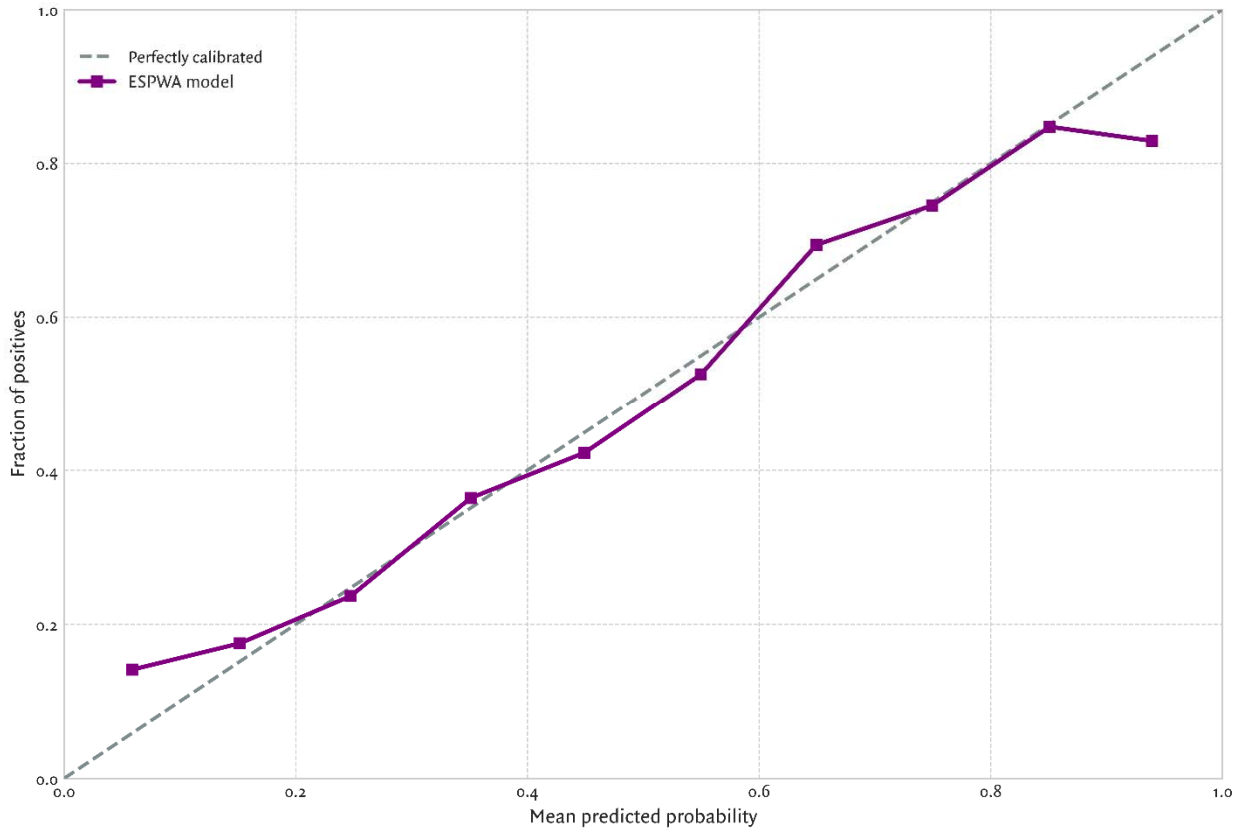
Figure 4: Comparison of ESPWA with a breast pathologist in ER status prediction from H&E WSI. A) Receiver operating characteristic (ROC) curve depicting performance of ESPWA (purple line) compared to a breast pathologist (Reader, blue dot). The ideal classifier is represented by the yellow star at the top left corner. B) Subgroup analysis comparing classification accuracy between ESPWA (purple) and the breast pathologist (blue circles) across clinical and pathological subgroups. Error bars represent 95% confidence intervals. Statistical comparisons by McNemar tests for all subgroups are provided.



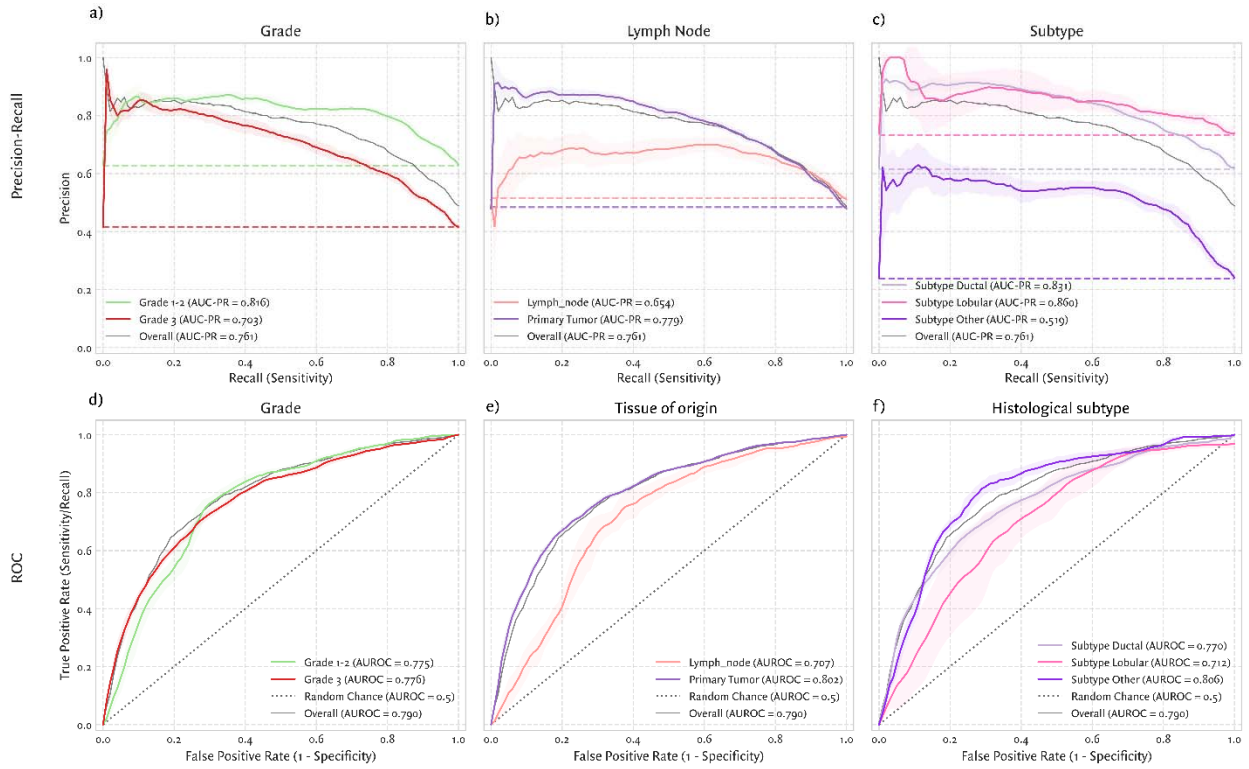
Supplemental Figure 1: Density plot of predicted vs. ground truth quantitative ER expression values for ESPWA. The heatmap intensity (purple) reflects number of WSIs (sample density), with marginal density distributions shown for the ground truth (top, blue) and predicted values (right, red). The purple dashed line indicates the linear regression fit, while the black dotted line represents the ideal y = x relationship. Performance metrics including PCC (0.706), ICC (0.524), and mean squared error (MSE = 0.045) are displayed, illustrating moderate agreement between ESPWA's predicted and actual ground truth ER expression levels.



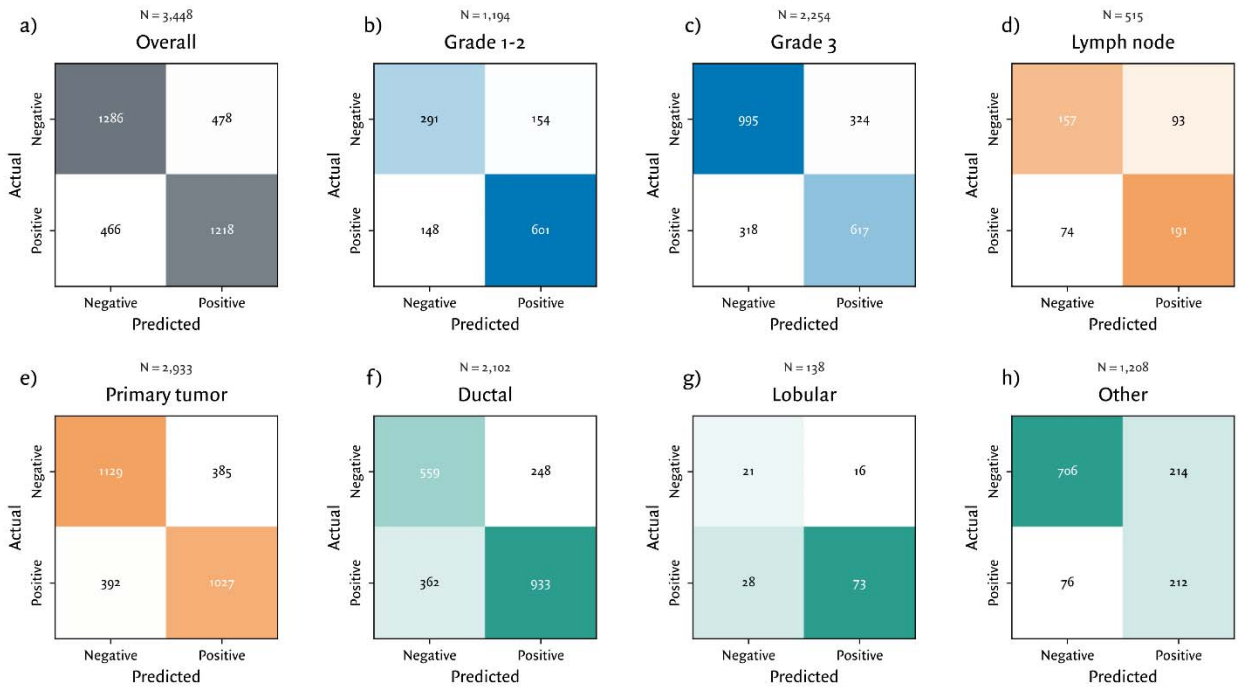
Supplemental Figure 2: Calibration plot for ESPWA for ER status prediction. Calibration performance of ESPWA (purple) is illustrated by plotting observed fraction of ER-positive cases against the mean predicted probability (model output). This plot illustrates that ESPWA is a well-calibrated model, as this model is in close alignment with a perfectly calibrated model (dashed line), thus enabling our decision curve analysis.



Supplemental Figure 3: Precision-recall (top row) and ROC curves (bottom row) evaluating model performance stratified by clinical and histologic subgroups. A & B) Precision-recall and AUROC curves for the tumor grade subgroups. C & D) Precision-recall and AUROC curves for the tissue of origin subgroups. E & F) Precision-recall and AUROC curves for histological subtype subgroups



Supplemental Figure 4: Confusion matrices for ESPWA stratified by clinical and histologic subgroups. (A) overall, (B-C) grade, (D-E) tissue type, (F-H) histologic subtype.



Cutpoint Strategy	Cutpoint	Sensitivity	Specificity	PPV	NPV	ER positives untreated (n = 1,684)	ER negatives untreated (n = 1,764)	Net benefit relative to "Treat All" (per 100 patients)
Sensitivity 10%	0.899	10.5% (9.3% - 11.8%)	98.0% (97.6% - 98.4%)	83.5% (79.6% - 87.3%)	53.4% (51.6% - 54.9%)	1,507 (1484 - 1525)	1,729 (1698 - 1767)	45.3 (44.0 - 46.3)
Sensitivity 20%	0.845	20.2% (19.1% - 21.2%)	96.7% (96.2% - 97.8%)	85.4% (82.7% - 89.6%)	55.9% (54.6% - 56.8%)	1,344 (1297 - 1403)	1,706 (1646 - 1758)	42.3 (40.1 - 43.7)
Default	0.5	72.3% (69.9% - 74.5%)	72.9% (71.3% - 74.9%)	71.8% (70.7% - 73.2%)	73.4% (71.5% - 74.6%)	466 (432 - 470)	1,286 (1253 - 1329)	23.8 (22.2 - 25.1)
Optimal (Youden's J)	0.482	74.5% (72.9% - 74.6%)	71.5% (70.4% - 72.7%)	71.4% (70.1% - 72.8%)	74.6% (73.1% - 74.7%)	430 (399 - 456)	1,262 (1227 - 1306)	23.2 (22.1 - 24.7)
Sensitivity 80%	0.415	80.0% (78.7% - 81.2%)	63.7% (62.6% - 65.6%)	67.8% (66.6% - 70.0%)	77.0% (75.9% - 78.5%)	336 (305 - 362)	1,123 (1076 - 1155)	18.8 (16.9 - 21.0)
Sensitivity 90%	0.252	90.0% (89.2% - 91.1%)	43.1% (41.4% - 44.8%)	60.2% (59.2% - 62.2%)	81.9% (80.0% - 83.7%)	168 (156 - 185)	761 (743-786)	7.6 (4.7 - 9.5)
Treat All (baseline)	0.000	100% (100% - 100%)	0% (0% - 0%)	48.8% (47.2% - 50.4%)	0% (0% - 0%)	0 (0 - 0)	0 (0 - 0)	0 (0 - 0)

This table presents ESPWA's performance on the validation cohort under various model thresholds. For each metric, we report the point estimate and 95% confidence interval (CI), calculated via a bootstrap procedure with 1,000 resamples.

The column "Cutpoint Strategy" describes the criterion for selecting the classification threshold, which may include the point of maximum accuracy (Youden's Index), thresholds that ensure predefined sensitivity levels (e.g., 10%, 20%, 80%, 90%), and a comparison to the "Treat All" standard of care, corresponding to the highest achievable sensitivity. The "Cutpoint" column indicates the numerical value of the model's probability threshold, with scores greater than or equal to this value classified as ER-positive.

"Sensitivity" measures the proportion of ER-positive cases correctly identified, calculated as True Positives divided by the sum of True Positives and False Negatives, while "Specificity" measures the proportion of ER-negative cases correctly identified, calculated as True Negatives divided by the sum of True Negatives and False Positives.

"Positive Predictive Value (PPV)" indicates the probability that a case classified as ER-positive is truly positive, calculated as True Positives divided by the sum of True Positives and False Positives, and "Negative Predictive Value (NPV)" indicates the probability that a case classified as ER-negative is truly negative, calculated as True Negatives divided by the sum of True Negatives and False Negatives.

The column "ER positives untreated" represents the number of ER-positive cases incorrectly classified as negative (False Negatives), corresponding to patients who would be missed for treatment; the cohort includes 1,684 ER-positive patients.

The column "ER negatives untreated" represents the number of ER-negative cases correctly classified as negative (True Negatives), corresponding to patients who would correctly avoid unnecessary treatment; the cohort includes 1,764 ER-negative patients. Finally, "Net Benefit" measures the clinical utility of the model by weighing the benefits of true positives against the harms of false positives, with higher values indicating greater usefulness compared to the default strategies of treating all or no patients.

Supplemental Table 1. Threshold-based performance and net benefit analysis of ESPWA

	Subgroup	N. samples	ER+	ER-	Accuracy	Precision	Recall	Sensitivity	Specificity
Overall									
	All Samples	3,448	1684	1764	0.639 (0.619-0.652)	0.611 (0.595-0.622)	0.709 (0.677-0.736)	0.709 (0.677-0.736)	0.572 (0.562-0.589)
ER Status									
	ER-positive	1,684	1684	0	0.709 (0.692-0.745)	—	—	—	—
	ER-negative	1,764	0	1764	0.572 (0.547-0.587)	—	—	—	—
ER quantitative expression									
	ER-Low expression (0-10%)	144	144	0	0.518 (0.481-0.611)	—	—	—	—
	ER-High expression (>10%)	1,479	1475	0	0.728 (0.703-0.744)	0.999 (0.996-1.000)	0.728 (0.703-0.745)	0.728 (0.703-0.745)	0.750 (0.113-1.000)
Tumor grade									
	Grade 1-2 (Well-differentiated)	1,194	749	445	0.606 (0.596-0.635)	0.655 (0.630-0.677)	0.842 (0.826-0.878)	0.842 (0.826-0.878)	0.158 (0.143-0.189)
	Grade 3 (Poorly - differentiated)	2,254	935	1319	0.656 (0.635-0.668)	0.567 (0.535-0.617)	0.597 (0.583-0.605)	0.597 (0.583-0.605)	0.695 (0.661-0.728)
Tissue type									
	Lymph node	515	265	250	0.680 (0.620-0.703)	0.696 (0.656-0.728)	0.693 (0.632-0.731)	0.693 (0.632-0.731)	0.665 (0.599-0.699)
	Primary tumor	2,933	1419	1514	0.630 (0.619-0.647)	0.595 (0.583-0.612)	0.713 (0.693-0.738)	0.713 (0.693-0.738)	0.554 (0.533-0.577)
Histological subtype									
	Ductal	2,102	1295	807	0.644 (0.630-0.663)	0.709 (0.692-0.737)	0.689 (0.669-0.699)	0.689 (0.669-0.699)	0.576 (0.553-0.633)
	Lobular	138	101	37	0.629 (0.563-0.744)	0.701 (0.631-0.840)	0.844 (0.766-0.914)	0.844 (0.766-0.914)	0.080 (0.034-0.184)
	Other	1,208	288	920	0.631 (0.615-0.655)	0.401 (0.358-0.443)	0.748 (0.717-0.777)	0.748 (0.717-0.777)	0.588 (0.563-0.622)

Supplemental Table 2. Pathologist Performance Metrics Across Clinical and Histologic Subgroups Across the ZL dataset

Acknowledgments: We specifically thank Scott Kilcoyne of DigitCells for the use of a KFBIO KF-PRO-005 scanner.

Data Sharing Statement: Any requests for raw and analyzed data will be reviewed by the Massachusetts General Hospital Institutional Review Board (IRB). Any data and materials (e.g. imaging or IHC data) that can be shared will need approval from the MGH IRB and a Material Transfer Agreement in place. All data shared will be deidentified.

Code Availability: Code from this study is available at: https://github.com/QTIM-Lab/ESPWA.

Competing Interests Statement: SW has consulted for Foundation Medicine, Veracyte, Hologic, Eli Lilly, Biovica, Pfizer/Arvinas, Puma Biotechnology, Novartis, AstraZeneca, Genentech, Regor Therapeutics, Stemline/Menarini, and Gilead, received institutional research funding to Genentech, Eli Lilly, Pfizer/Arvinas, Nuvation Bio, Regor Therapeutics, Sermonix, Puma Biotechnology, Stemline/Menarini, and Phoenix Molecular Designs, and received honoraria from Eli Lilly, Guardant Health, and 2ndMD.

REFERENCES

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin [Internet] 2024 [cited 2025 Jul 20];74(3):229–63. Available from: /doi/pdf/10.3322/caac.21834

2. Burstein HJ, Lacchetti C, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor–positive breast cancer: ASCO clinical practice guideline focused update. Journal of Clinical Oncology [Internet] 2019 [cited 2025 Jul 20];37(5):423–38. Available from: https://ascopubs.org/doi/pdf/10.1200/JCO.18.01160

3. Dangi RR, Sharma A, Vageriya V. Transforming Healthcare in Low-Resource Settings With Artificial Intelligence: Recent Developments and Outcomes. Public Health Nurs [Internet] 2025 [cited 2025 Jul 20];42(2):1017–30. Available from: /doi/pdf/10.1111/phn.13500

4. Fadelu T, Damuse R, Lormil J, et al. Patient Characteristics and Outcomes of Nonmetastatic Breast Cancer in Haiti: Results from a Retrospective Cohort. Oncologist [Internet] 2020 [cited 2025 Jul 20];25(9):e1372. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7485367/

5. Giaquinto AN, Sung H, Newman LA, et al. Breast cancer statistics 2024. CA Cancer J Clin [Internet] 2024 [cited 2025 Aug 16];74(6):477–95. Available from: /doi/pdf/10.3322/caac.21863

6. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nature Reviews Cancer 2021 22:2 [Internet] 2021 [cited 2023 Jul 12];22(2):114–26. Available from: https://www.nature.com/articles/s41568-021-00408-3

7. Lipkova J, Chen RJ, Chen B, et al. Artificial intelligence for multimodal data integration in oncology. Cancer Cell [Internet] 2022 [cited 2023 Jul 12];40(10):1095–110. Available from: https://pubmed.ncbi.nlm.nih.gov/36220072/

8. Pistilli B, Lohrisch C, Sheade J, Fleming GF. Personalizing Adjuvant Endocrine Therapy for Early-Stage Hormone Receptor–Positive Breast Cancer. American Society of Clinical Oncology Educational Book [Internet] 2022 [cited 2025 Aug 16];(42):60–72. Available from: https://ascopubs.org/doi/pdf/10.1200/EDBK_350358

9. Martei YM, Pace LE, Brock JE, Shulman LN. Breast Cancer in Low- and Middle-Income Countries: Why We Need Pathology Capability to Solve This Challenge. Clin Lab Med [Internet] 2017 [cited 2025 Jul 20];38(1):161. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6277976/

10. Franzoi MA, Agostinetto E, Perachino M, et al. Evidence-based approaches for the management of side-effects of adjuvant endocrine therapy in patients with breast cancer. Lancet Oncol [Internet] 2021 [cited 2025 Jul 21];22(7):e303–13. Available from: https://www.thelancet.com/action/showFullText?pii=S1470204520306665

11. Couture HD, Williams LA, Geradts J, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer [Internet] 2018 [cited 2025 Jul 20];4(1):1–8. Available from: https://www.nature.com/articles/s41523-018-0079-1

12. Shamai G, Schley R, Cretu A, et al. Clinical utility of receptor status prediction in breast cancer and misdiagnosis identification using deep learning on hematoxylin and eosin-stained slides. Communications Medicine 2025 4:1 [Internet] 2024 [cited 2025 Jul 20];4(1):1–16. Available from: https://www.nature.com/articles/s43856-024-00695-5

13. Naik N, Madani A, Esteva A, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. Nat Commun [Internet] 2020 [cited 2025 Jul 20];11(1):5727. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7670411/

14. Fleming KA, Naidoo M, Wilson M, et al. An Essential Pathology Package for Low- and Middle-Income Countries. Am J Clin Pathol [Internet] 2017 [cited 2025 Aug 16];147(1):15–32. Available from: https://dx.doi.org/10.1093/ajcp/aqw143

15. Wang X, Steensma JT, Bailey MH, Feng Q, Padda H, Johnson KJ. Characteristics of The Cancer Genome Atlas cases relative to U.S. general population cancer cases. Br J Cancer [Internet] 2018 [cited 2025 Jul 20];119(7):885. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6189215/

16. Hirko KA, Rocque G, Reasor E, et al. The impact of race and ethnicity in breast cancer—disparities and implications for precision oncology. BMC Med [Internet] 2022 [cited 2025 Jul 20];20(1):1–12. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-022-02260-0

17. Goel N, Kim DY, Guo JA, Zhao D, Mahal BA, Alshalalfa M. Racial Differences in Genomic Profiles of Breast Cancer. JAMA Netw Open [Internet] 2022 [cited 2025 Jul 20];5(3):e220573–e220573. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2789516

18. Woriax HE, Thomas SM, Plichta JK, et al. Racial/Ethnic Disparities in Pathologic Complete Response and Overall Survival in Patients with Triple-Negative Breast Cancer Treated with Neoadjuvant Chemotherapy. Journal of Clinical Oncology [Internet] 2024 [cited 2025 Jul 20];42(14):1635–45. Available from: https://ascopubs.org/doi/pdf/10.1200/JCO.23.01199

19. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering 2021 5:6 [Internet] 2021 [cited 2023 Sep 2];5(6):555–70. Available from: https://www.nature.com/articles/s41551-020-00682-w

20. Lu MY, Chen B, Williamson DFK, et al. A visual-language foundation model for computational pathology. Nature Medicine 2024 30:3 [Internet] 2024 [cited 2024 Apr 29];30(3):863–74. Available from: https://www.nature.com/articles/s41591-024-02856-4

21. Vickers AJ, Van Claster B, Wynants L, Steyerberg EW. Decision curve analysis: confidence intervals and hypothesis testing for net benefit. Diagnostic and Prognostic Research 2023 7:1 [Internet] 2023 [cited 2025 Jul 20];7(1):1–9. Available from: https://diagnprognres.biomedcentral.com/articles/10.1186/s41512-023-00148-y

22. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Medical Decision Making [Internet] 2006 [cited 2025 Jul 20];26(6):565–74. Available from: /doi/pdf/10.1177/0272989X06295361?download=true

23. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res [Internet] 2019 [cited 2025 Jul 20];3(1):1–8. Available from: https://diagnprognres.biomedcentral.com/articles/10.1186/s41512-019-0064-7

24. Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. Bioengineering [Internet] 2024 [cited 2025 Jul 20];11(4). Available from: https://pubmed.ncbi.nlm.nih.gov/38671759/

25. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare [Internet] 2020 [cited 2025 Jul 20];25. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7325854/

26. Onu CC, Lebensold J, Hamilton WL, Precup D. Neural Transfer Learning for Cry-based Diagnosis of Perinatal Asphyxia. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH [Internet] 2019 [cited 2025 Jul 20];2019-September:3053–7. Available from: https://arxiv.org/pdf/1906.10199

27. Martei YM, Pace LE, Brock JE, Shulman LN. Breast Cancer in Low- and Middle-Income Countries: Why We Need Pathology Capability to Solve This Challenge. Clin Lab Med [Internet] 2017 [cited 2025 Jan 25];38(1):161. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6277976/

28. Skjervold AH, Valla M, Bofin AM. Oestrogen receptor low positive breast cancer: associations with prognosis. Breast Cancer Res Treat [Internet] 2023 [cited 2025 Jul 20];201(3):535–45. Available from: https://pubmed-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/37462784/

29. Benefield HC, Allott EH, Reeder-Hayes KE, et al. Borderline estrogen receptor-positive breast cancers in black and white women. J Natl Cancer Inst [Internet] 2020 [cited 2025 Jul 20];112(7):728–36. Available from: https://pubmed-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/31742342/

30. Nelson AM, Milner DA, Rebbeck TR, Iliyasu Y. Oncologic Care and Pathology Resources in Africa: Survey and Recommendations. J Clin Oncol [Internet] 2016 [cited 2025 Jan 25];34(1):20–6. Available from: https://pubmed.ncbi.nlm.nih.gov/26578619/

31. Huang Z, Bianchi F, Yuksekgonul M, Montine TJ, Zou J. A visual–language foundation model for pathology image analysis using medical Twitter. Nature Medicine 2023 29:9 [Internet] 2023 [cited 2024 Jan 27];29(9):2307–16. Available from: https://www.nature.com/articles/s41591-023-02504-3